



King's Research Portal

Document Version

Publisher's PDF, also known as Version of record

[Link to publication record in King's Research Portal](#)

Citation for published version (APA):

Venturini, T., Jacomy, M., Bounegru, L., & Gray, J. (2019). Exploração visual de redes para jornalistas de dados. In *Métodos Digitais: teoria-prática-crítica Livros ICNOVA*. <https://www.icnova.fcsh.unl.pt/en/icnova-metodos-digitais/>

Citing this paper

Please note that where the full-text provided on King's Research Portal is the Author Accepted Manuscript or Post-Print version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version for pagination, volume/issue, and date of publication details. And where the final published version is provided on the Research Portal, if citing you are again advised to check the publisher's website for any subsequent corrections.

General rights

Copyright and moral rights for the publications made accessible in the Research Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the Research Portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the Research Portal

Take down policy

If you believe that this document breaches copyright please contact librarypure@kcl.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



MÉTODOS DIGITAIS TEORIA-PRÁTICA-CRÍTICA

EDITADO POR
JANNA JOCELI OMENA



COLEÇÃO
ICNOVA

icNOVA INSTITUTO
DE COMUNICAÇÃO
DA NOVA

FICHA TÉCNICA

TÍTULO

Métodos Digitais: teoria-prática-crítica

AUTOR

Janna Joceli Omena (Editora)

COLEÇÃO

Livros ICNOVA

EDIÇÃO

ICNOVA — Instituto de Comunicação da Nova
Faculdade de Ciências Sociais e Humanas/Universidade NOVA de Lisboa
Av. Berna, 26
1069-061 Lisboa — Portugal
www.icnova.fcsh.unl.pt icnova@fcsh.unl.pt

DIREÇÃO

Francisco Rui Cádima
Maria Lucília Marques
Cláudia Madeira

ISBN

978-972-9347-34-4 (Digital)

978-972-9347-33-7 (Impresso)

DESIGN E PAGINAÇÃO

José Domingues | UNDO

DATA DE PUBLICAÇÃO

Dezembro, 2019

APOIO

FCT Fundação
para a Ciência
e a Tecnologia



A edição deste livro é financiada por fundos nacionais através

da FCT — Fundação para a Ciência e a Tecnologia no âmbito do projeto Ref.º: UID/CCI/04667/2019

ÍNDICE

INTRODUÇÃO

| | |
|------------------------------------|---|
| O que são métodos digitais? | 5 |
| JANNA JOCELI OMENA | |

UM

FUNDAMENTOS

| | |
|---|----|
| Métodos Digitais: dos desafios à bildung | 19 |
| BERNHARD RIEDER, THEO RÖHLE | |

| | |
|--|----|
| O tecido social: rastros digitais e métodos quali-quantitativos | 37 |
| TOMMASO VENTURINI, BRUNO LATOUR | |

DOIS

CONCEITOS, ABORDAGENS E PRÁTICAS

| | |
|---------------------------------|----|
| A Plataformização da Web | 49 |
| ANNE HELMOND | |

| | |
|--|----|
| Engajados de outra maneira: as mídias sociais — das métricas de vaidade à análise crítica | 73 |
| RICHARD ROGERS | |

TRÊS

REDES DIGITAIS E AFFORDANCES VISUAIS

| | |
|--|----|
| Exploração visual de redes para jornalistas de dados | 99 |
| TOMMASO VENTURINI, MATHIEU JACOMY, LILIANA BOUNEGRU, JONATHAN GRAY | |

| | |
|---|-----|
| Sistema de leitura de redes digitais multiplataforma | 121 |
| JANNA JOCELI OMENA, INÊS AMARAL | |

QUATRO

O PENSAMENTO NATIVAMENTE DIGITAL

**Métodos digitais nos estudos em saúde
*mapeando usos e propondo sentidos*** 143
ELAINE TEIXEIRA RABELLO, FÁBIO CASTRO GOUVEIA

O papel do *designer* no contexto do *Data Sprint* 161
MICHELE MAURI, BEATRICE GOBBO, GABRIELE COLOMBO

CINCO

EXPLORAR PRATICAMENTE, NAVEGAR CRITICAMENTE

**Produção e consumo de vídeos em 360º — tendências para o jornalismo
brasileiro no YouTube** 183
ANA MARTA M. FLORES

**Circulação de imagens entre plataformas: desafios e experiências
com uma API de visão computacional** 203
ANDRÉ MINTZ, CARLOS D'ANDRÉA

Autoras & Autores 225



INTRODUÇÃO

O QUE SÃO MÉTODOS DIGITAIS?

JANNA JOCELI OMENA

Que métodos você adota?

Métodos digitais.

Ah, ok. Os que utilizo são mais qualitativos.

No contexto de pesquisa relacionada com a Internet, a dicotomia dos métodos de investigação como qualitativos ou quantitativos é uma premissa antiga e enraizada na nossa forma de ver o mundo. E a comunidade acadêmica não é exceção a esta regra. Ao questionar o que são métodos digitais, e tomando como exemplo o típico diálogo exposto acima, a resposta parece ser simples: métodos quantitativos. Torna-se trivial a imediata alusão ao uso de ferramentas computacionais, também inferidas como métodos que pouco dialogam com abordagens qualitativas. Por exemplo, extrair dados de uma plataforma digital e aplicar fórmulas estatísticas para mensurar interação ou impacto — sobretudo via a contabilização de publicações diárias, números de *likes* ou voltar as atenções para os conteúdos mais recomendados. Neste sentido, modelos reflexivos e interpretativos logicamente fogem ao escopo dos *supostos* métodos digitais, tidos apenas como instrumentos para coleta ou processamento de dados online.

Na mesma linha de pensamento, por um lado, as limitações impostas pelas plataformas web (instabilidade, mudanças contínuas ou restrições ao acesso a dados públicos via interfaces técnicas) são alvo de crítica e reflexão para os investigadores dos media digitais. Por outro lado, e ao longo do processo de pesquisa, estas mesmas problemáticas são desligadas da etapa que define a extração

e análise de dados; excluídas sobretudo do design interrogativo¹. Aqui a infraestrutura da plataforma, os seus métodos e técnicas não interferem no processo analítico, nem na concepção das perguntas de pesquisa. Dois mundos muito bem separados e definidos. Estaríamos todos bem esclarecidos se a proposta deste livro seguisse à risca o senso (ainda) comum e o enquadramento instrumental associados aos métodos digitais.

Talvez seja melhor começar novamente e esclarecer o que não são métodos digitais. Faz-se primeiro necessário entender que estes métodos propõem o deslocamento da discussão da teoria específica-do-meio (*medium-specific theory*) do ponto de vista ontológico (propriedades e características) para o ponto de vista epistemológico (método). Discussão teórico-prática fundamentada pelo precursor desta abordagem metodológica, Richard Rogers (2013, 2019), os métodos digitais, portanto, *não* correspondem a um campo para aplicação de métodos existentes, nem à migração do instrumentário das Ciências Sociais para o ambiente online (ver Rogers, 2010, 2015). Por exemplo, a aplicação de questionários ou inquéritos via e-mail, grupos de discussão online, plataformas de media sociais ou mesmo recorrer ao mercado de *crowdsourcing* (Amazon Mechanical Turk, Prolific, Lucid) *não* são exemplos de métodos digitais. Estes também *não* são a justaposição de análises estatísticas combinadas com observação etnográfica (Venturini & Latour, 2010), tão-pouco métodos ‘prontos’ (pré-fabricados) ou os chamados métodos ‘por encomenda’ (Marres, 2017). Métodos digitais *não* correspondem ao desenvolvimento e uso fortuito de software de extração ou análise de dados digitais. Pelo contrário, os métodos digitais nos convidam a observar, considerar e redirecionar os mecanismos inerentes às plataformas web e os seus objetos *nativos* digitais para investigação social. É um processo essencialmente de alfabetização digital e requer, portanto, uma nova cultura de conhecimento.

De forma objetiva, os métodos digitais podem ser definidos como uma prática de pesquisa *quali-quant*i que re-imagina a natureza, os mecanismos e os dados nativos às plataformas web e motores de busca para estudar a sociedade. Tem como ponto de partida e arena investigativa a Internet e o ambiente online. A partir daí, é feita a conceptualização da pesquisa que segue o meio (*medium*), captura a sua dinâmica, e fundamenta os seus resultados e visões sobre as mudanças culturais e sociais no online (*online groundedness*) (Rogers, 2013). Consequentemente, ao seguir a lógica fluida e não estática deste meio, estes métodos demandam contínuas mudanças e adaptações. Esta instabilidade tende a incomodar os estudiosos do digital pela ausência de controle sobre o processo metodológico. Contudo,

1 *Query design*.

lidar com o digital é também sinônimo de tentar compreender o transitório, passageiro, efêmero. Esta dinâmica ativa e nunca entediante de seguir a lógica do meio (*medium*) é inerente à abordagem dos métodos digitais que, por essa razão, reúne um conjunto particular de questionamentos. *Por onde começar? Como fazer? É possível? E se?* Mesmo que as soluções para estas questões possam mudar ao longo do tempo, perguntas como estas e as ações práticas por elas provadas são uma constante no realizar métodos digitais.

Na prática, e contrariando as expectativas de muitos investigadores, os métodos digitais implicam trabalho manual e um monitoramento ativo, por exemplo, contínua verificação dos dados coletados ou produção de visualizações exploratórias (ver Rogers, 2019). Pensar *com e através* dos dispositivos digitais reflete específicos tipos de estudo, como por exemplo: a reconfiguração do uso de motores de busca para pesquisa; mapeamento de redes na *web* através de análises de hiperlink; utilizar o Twitter como uma máquina de contar histórias; identificar e analisar conteúdos com alto nível de Reações no Facebook; ou transformar a Wikipédia em plataforma de referência para os estudos culturais, entre outras possibilidades (ver Rogers, 2019).

Neste entendimento, é natural que nos questionemos sobre como conduzir estudos através e sobre a Internet e as plataformas de media sociais sob a perspectiva dos métodos digitais. Como formular as perguntas de investigação? De que forma devemos pensar os objetos nativos digitais e as gramáticas dos media sociais (i.e. hashtags, URLs, likes, recomendação algorítmica, classificação de informação)? Que tipos de estudos podem ser realizados através destes métodos? Quais são as *affordances* e limitações dos métodos digitais? Para além de proporcionar possíveis respostas e reflexão crítica sobre estas questões, o objetivo principal deste livro é proporcionar uma introdução teórico-prática-técnica sobre métodos digitais. Ao mesmo tempo em que defende, em linha com Latour, Jensen, Venturini, Grauwin, & Boullier (2012), a convergência entre duas opostas e tradicionais formas de fazer investigação: as abordagens qualitativas e quantitativas. Preserva, portanto, o posicionamento de que a teoria ‘anda de mãos dadas’ com a investigação aplicada. Para concretizar este objetivo *Métodos Digitais: Teoria-Prática-Crítica* é organizado em cinco partes que refletem as fundações dos métodos digitais.

2 Na língua inglesa os tipos de questões são: ‘where to start’, ‘how to’, ‘can we’, ‘what if’. Questões que evidenciam o processo logístico e exploratório de análise de dados online e objetos nativos digitais.

FUNDAÇÕES DOS MÉTODOS DIGITAIS

A realização de pesquisa fundamentada nos métodos digitais envolve processos e questionamentos similares às práticas tradicionais de investigação. Contudo, quatro pontos determinantes separam estas últimas dos métodos digitais. No primeiro, os métodos digitais assumem uma *posição de interdependência no processo investigativo* e, assim, serem presentes desde a concepção da investigação até ao seu processo analítico. No segundo ponto está a *consideração da infraestrutura* das plataformas *web* e ou motores de busca. Esta deve ter um papel ativo no design interrogativo da pesquisa, pois os mecanismos das plataformas *web* interferem, moldam e organizam a forma como vemos e entendemos as questões sociais. Por mecanismos, entende-se o sistema de captura e organização de dados impostos pela plataforma, sistemas algorítmicos de personalização e recomendação, por exemplo. Neste ângulo, não é possível estudar a sociedade através de uma plataforma, sem estudar a plataforma em si.

No ponto três deparamos-nos com a exigência do (mínimo) *conhecimento técnico-prático* sobre a interligação de técnicas de extração-análise-visualização de dados online e a sua natureza relacional com o contexto, os objetivos e as perguntas de partida do objeto de estudo. Por fim, mas não menos importante, o pressuposto de que a proposta dos métodos digitais é simultaneamente uma *lógica interpretativa-quantificativa* e um *processo reflexivo*. A consciência prática sobre estes quatro pontos reflete-se no melhor entendimento da lógica funcional dos métodos digitais. Neste ponto de partida devem iniciar-se estudos baseados na perspectiva dos métodos.

O esquema apresentado na Figura 1 é uma tentativa de reunir o processo *espiral* de trabalho com métodos digitais, juntamente com os tipos de perguntas inerentes a sua lógica funcional (Por onde começar? Como fazer? É possível? E se?). Todas as etapas estão, de certa forma, interligadas tendo também um caráter relacional. Mesmo sem haver uma ordem direcionada de trabalho, normalmente, o seu fluxo tende a iniciar do design interrogativo³ e o seu entrelaçamento com a infraestrutura da plataforma à práxis de extração, análise e visualização exploratória de dados, indo até à definição das análises e visualizações finais. Esse processo é refinado e transformado quando a sua trajetória é repetida mas, desta vez, com uma aprendizagem dos erros e acertos. Na ordem prática, para além de seguir a perspectiva específica do meio (*medium-specificity*) e dos objetos nativos

³ A técnica de construção de listas de partida como as de palavras-chave, fontes especialistas, URLs ou hashtags. Estas são importantes pois estruturam o design interrogativo (*query design*).

digitais, testar diferentes formas e experimentar quantas vezes for necessário é um aspecto inerente aos métodos digitais. Por fim, podemos dizer que a lógica funcional apresentada no diagrama também nos informa que a pesquisa baseada em métodos digitais não lida necessariamente com *big data*, mas sobretudo com o aspecto da natureza relacional dos dados.

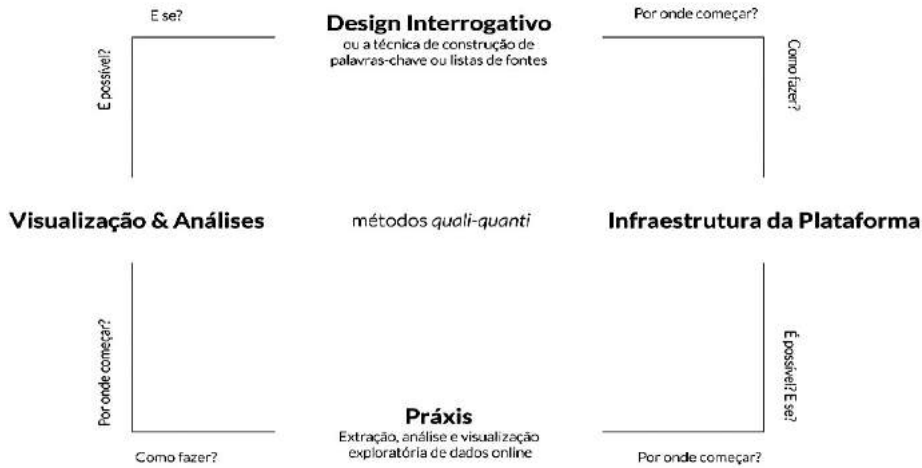


FIGURA 1 – A lógica funcional dos métodos digitais.

Esta breve discussão sobre as fundações dos métodos digitais teve como objetivo oferecer à leitora e ao leitor deste livro uma introdução ao seu propósito maior. O de apresentar os métodos digitais sob um olhar teórico-prático-crítico (e técnico!). Igualmente, *Métodos Digitais: Teoria-Prática-Crítica* oferece uma base referencial à comunidade acadêmica, organizações não-governamentais, estudantes, investigadores e demais grupos interessados nesta prática de pesquisa. Para tal, a estrutura do livro é dividida em cinco partes, a seguir apresentadas.

TEORIA-PRÁTICA-CRÍTICA

A parte UM “Fundamentos”, apresenta o desafio da educação digital e as novas possibilidades de abertura de pesquisa para ciências sociais através dos métodos digitais. Para Bernhard Rieder e Theo Röhle não se trata apenas de discutir sobre os desafios inerentes a estes métodos (i.e. o poder da evidência visual ou *black-boxing*), mas reconhecer os conhecimentos mobilizados pelas ferramentas digitais. Rieder

& Röhle nos convidam a questionar sobre o que pode significar “entender” uma técnica computacional. Através de três exemplos comumente relacionados com a metodologia digital – estatística, análise de redes, e milhares de imagens, os autores questionam de que maneira estes métodos e ferramentas constituem o conhecimento gerado, apresentado e legitimado nas humanidades e ciências sociais. Neste percurso, Rieder & Röhle defendem a necessidade de tornar realidade no domínio acadêmico “uma prática que oscila entre o trabalho técnico concreto e a reflexividade metodológica”, ao passo que desenvolvem a noção de digital *Bildung* de David Berry.

No capítulo seguinte, Tommaso Venturini & Bruno Latour refletem sobre o digital e como os rastros digitais impactam na forma de olhar e fazer ciências sociais. Ao comparar estas com as ciências naturais no século xv – que, depois da invenção da tinta de prensa móvel, levaram décadas para adquirir maturidade e causar impacto significativo no campo – Venturini e Latour afirmam que as ciências sociais se encontram hoje numa posição similar. “Ao invés de utilizar os dados digitais para revitalizar as suas práticas, tais ciências ainda tentam fazê-los se passar por novos terrenos, mas por meio de antigos métodos.” Em contrapartida, os autores argumentam que os dados digitais proporcionam novas possibilidades para os cientistas sociais ao permitir diferentes perspectivas de ver, analisar, teorizar e criticar o tecido social, das microinterações às macroestruturas. Esta visão oligóptica da sociedade torna-se possível através da promessa dos métodos digitais e a sua proposta *quali-quantitativa*.

DOIS destaca os “Conceitos, Abordagens e Práticas” relacionados com os métodos digitais. Reúne não apenas reflexões críticas, mas ensinamentos fundamentais sobre o pensar através da lógica do meio (*médium-specificity*). Anne Helmond desenvolve a noção de *plataformização da web* sob uma visão teórica-técnica-crítica introduzindo conceitos e abordagens essenciais para a realização de pesquisa em Ciências Sociais que dialogam, observam e estudam a sociedade através da Internet. Seguindo a definição computacional de Tim O’Reilly (2005), Helmond apresenta de forma precisa e densa a importância e o impacto da *programabilidade* no processo que envolve as dinâmicas de descentralização e recentralização de dados online e o seu fluxo interno e externo às plataformas (i.e. Facebook). Para isto, e através de uma abordagem técnico-materialista, Helmond dissecou o que O’Reilly (o criador da expressão Web 2.0) afirmou ser o máximo da vantagem intrínseca das plataformas *web*: as interfaces de programação de aplicações (ou APIs). Deste modo, decorre da ascensão das APIs de media sociais aos seus níveis e condições de programabilidade. A partir desta reflexão, Helmond não apenas elucida a lógica de *plataformização da web*, bem como justifica a mudança de conceitos de “sites de redes sociais” para “plataformas de media

sociais”. Apesar da terminologia “rede social” ser utilizada no dia a dia para referenciar empresas como Facebook, é essencial para o entendimento deste novo conceito a definição de programabilidade e o papel das APIs nas plataformas web.

Dando continuidade ao enquadramento proposto nesta parte, Richard Rogers introduz uma abordagem às chamadas alternativas *métricas de vaidade* — tal como o sistema Klout, que quantifica atividade online e suas intensidades com base nas métricas fornecidas pelas plataformas — frequentemente incorporadas como base para estudos sobre os media sociais. Em oposição a isto, e sob a lente da *análise crítica*, Rogers argumenta que as plataformas digitais deixam de ser espaço de vaidades direcionadas pela lógica do sucesso, conexões produtivas e consumo, para se tornarem numa rede de questões problemáticas (*issue networks*) e engajamento (*engagement*). Ao considerar tanto a cultura de uso da plataforma como a sua forma de funcionamento, Rogers propõe outras medidas de engajamento a serem consideradas no processo de investigação. Tais como voz dominante, interesse, compromisso, posicionamento e alinhamento. Podemos destacar como aspectos decisivos desta abordagem analítica a identificação da *linguagem* utilizada pelos atores, e a longevidade do seu uso; distinguir “especificamente para quem uma questão problemática é uma questão de interesse num dado momento”; e detectar quais são os atores que participam de um programa e de um antiprograma. Seguindo a perspectiva dos métodos digitais, a *análise crítica* oferece o redirecionamento de dados das plataformas para pesquisa social.

TRÊS reúne capítulos que exploram “Redes digitais e *affordances* visuais” através de um sistema de leituras de redes multiplataforma e da exploração dos recursos analíticos oferecidos pelas redes no contexto do jornalismo de dados. Este último, conta com a experiente contribuição de Tommaso Venturini e do Mathieu Jacomy em *análise visual de redes*, bem como de Liliana Bounegru e Jonathan Gray no campo do Jornalismo e Métodos Digitais. Os autores propõem uma abordagem inovadora para os jornalistas que desejam narrar e dar sentido a conjuntos de dados complexos através da exploração visual de redes digitais. De entre as variáveis gráficas de um grafo (constituído por nós e arestas, e dados relacionais), utilizam a posição, tamanho e cor dos nós para interpretar uma rede. A partir da rede de hiperlinks do *Décodex*, ferramenta online do jornal *Le Monde*, analisaram todos os websites contidos neste catálogo de fontes de informação. A análise proposta parte da identificação dos principais atores na rede às particularidades como idioma (através de *link domains*), detecção de sites confiáveis, não confiáveis e imprecisos, também os satíricos e conspiratórios. Sugerem também que o uso de cálculos e estatísticas simples de distribuição pode complementar a exploração visual de redes. A riqueza deste capítulo está no carácter iterativo da sua proposta, juntamente com o enten-

dimento de que, no processo de avaliar a confiabilidade dos sites pelo intermédio de redes de *hyperlinks*, a demarcação de “categorias e territórios são entrelaçadas e mutuamente constitutivas”. Com isto, os autores sinalizam novos cenários para o estudo e a prática do jornalismo digital.

Ainda no âmbito do estudo de redes digitais, Inês Amaral e eu própria somamos à proposta de Venturini e colegas, a noção de gramatização das plataformas, combinada com a prática de recolha e análise de dados como componentes essenciais no processo de leitura de redes digitais multiplataforma. Chamamos, portanto, a atenção para as camadas de conhecimento técnico inerentes ao processo analítico de redes digitais. A representação gráfica é também para nós um fator imprescindível, em particular o potencial narrativo fornecido pela espacialização do ForceAtlas2 (algoritmo de layout direcionado por força) e a natureza relacional dos dados. Desta forma, sugerimos que a leitura de redes digitais multiplataforma pode ser realizada a partir de *camadas fixas* de interpretação (centro, ponto intermédio, periferia e elementos isolados) que proporcionam *múltiplas formas* de leitura. Ilustramos a nossa proposta através de uma matriz de leitura de diferentes tipos de rede, i.e. redes de *page like* (Facebook), de coocorrências de hashtags (Instagram), de recomendações algorítmicas (YouTube), e de imagem-label (Vision APIs). Com base em estudos exploratórios prévios, apresentamos uma proposta de sistema de leitura de redes digitais multiplataforma. Argumentamos que este deriva de um triplo entendimento sobre a infraestrutura da plataforma, a sua representação gráfica e o seu valor hermenêutico.

QUATRO reflete “o pensamento nativamente digital” sob um olhar crítico e prático – pensamento concretizado apenas através do “aprendizado pelo fazer na prática”. O capítulo de Elaine Teixeira Rabello e do Fábio Castro Gouveia tem como ponto de partida o exercício de mapear de que modo os estudos de saúde vêm adotando “métodos digitais” no processo de investigação. Esta questão direcionou o estudo para a revisão na base de dados bibliográfica Scopus, onde se verificou considerável associação entre “métodos digitais” com os estudos de *big data* nas publicações deste campo. Somado a isto, os autores identificaram três acepções que justificam a adoção da expressão “métodos digitais” pelos investigadores em Saúde: i) digital como ferramenta ou recurso; ii) digital como arena ou fonte de dados; e iii) digital como objeto de estudo em si. Para além de discutir os resultados, a contribuição maior deste capítulo pode ser o provocativo desafio lançado aos investigadores em Saúde Pública e Coletiva. Isto é, a necessidade de as metodologias de pesquisa neste campo ultrapassarem a lógica arena-ferramenta-fonte, somando as plataformas digitais em si e as suas *affordances* como elementos-chave para pesquisa. Como resposta a este desafio, Rabello & Gouveia argumentam que

existem dois momentos indispensáveis para integração dos métodos digitais aos estudos de saúde. O primeiro está na formulação das perguntas do estudo para que a pesquisa seja, em si, *nativa* digital desde a sua concepção. O segundo, resume-se à capacidade de redirecionar a linguagem das plataformas digitais para o propósito da investigação em saúde. Rabello & Gouveia defendem o valor destes momentos mediante estudos desenvolvidos em *data sprints*.

A investigação científica com base nos métodos digitais é, de certa forma, incompleta sem a experiência prática, ativa e colaborativa proporcionada por um *data sprint*. Este pode ser definido como workshops intensos de pesquisa, codificação e aplicação de métodos digitais e visualização de dados (em caráter experimental, exploratório ou confirmatório), onde participantes provenientes de diferentes formações acadêmicas e áreas diversas se reúnem fisicamente para trabalhar juntos e responder coletivamente a uma série de questões de pesquisa⁴. Neste contexto, o capítulo de Michele Mauri, Beatrice Gobbo e Gabriele Colombo é imprescindível não apenas para os *designers* da informação e visualização de dados, mas para investigadores de áreas diversas que almejam participar ou desenvolver projetos via a abordagem dos *data sprints*. A riqueza deste capítulo está exatamente na sua capacidade de não apenas apresentar o papel do designer no contexto dos *data sprints*, mas no registro do que é um *sprint*, quais são as suas características, os seus diferentes tipos e formas de organização.

Para os autores, o designer, assim como um cartógrafo, assume um papel diferente em cada fase do *sprint*, devendo prioritariamente *informar*, *guiar* e *registrar*. A primeira fase (informar) está vinculada à elaboração de tutoriais ou o ensino passo a passo do processo de visualização de dados (que ao longo dos anos é aprimorado). Um exemplo é a construção de uma rede de imagens e relativas descrições (*image-label networks*), proporcionada por APIs de visão computacional e um conjunto de ferramentas e técnicas. A segunda fase (guiar) mostra que o designer deve estar envolvido no processo de discussão sobre as questões de pesquisa e objetivos, bem como é flexível às exigências que podem surgir ao longo do *data*

4 A Iniciativa dos Métodos Digitais, da Universidade de Amsterdão, é precursora no desenvolvimento deste tipo de *data sprints*, e referência no campo (ver *Summer School* e *Winter School* em <https://wiki.digitalmethods.net/Dmi/DmiAbout>). A Universidade Nova de Lisboa é também uma recente referência no contexto dos *data sprints*, através do laboratório de Media Digitais iNOVA Media Lab, parte do Instituto de Comunicação da Nova – ICNOVA (ver <https://smart.inovamedialab.org/>). Iniciativas como estas proporcionam: a criação e desenvolvimento de novas abordagens para investigação digital e dos media sociais; práticas e técnicas de visualização de dados; e também, possibilitam descrições (amplas ou específicas) e *insights* sobre o estudo em questão. O artigo “Data-Sprinting: a Public Approach to Digital Research”, da autoria de Tommaso Venturini, Anders Munk e Axel Meunier, é mais uma importante sugestão para os investigadores interessados em aprender sobre este novo formato para fazer investigação digital.

sprint. A visualização aqui resultado de um processo exploratório e intuitivo de tentativas e falhas. Importa que a visualização seja útil para a compreensão dos resultados ou perguntas de pesquisa. Por fim, cabe ao designer a função de resumir o processo investigativo com métodos digitais através de um protocolo visual (registrar).

CINCO, “Explorar praticamente, navegar criticamente”, engloba estudos alicerçados na visão dos métodos digitais que propõem contribuir para o campo do jornalismo e para pesquisa sobre circulação de imagens. Mesmo sendo um concentrado no YouTube e o outro com evidência nos estudos multiplataforma, os capítulos a seguir apresentados servem de modelo não apenas para os seus objetivos em particular, mas, sobretudo, para serem reproduzidos em diferentes contextos. Ana Marta M. Flores desenha um panorama de produção e consumo do vídeo 360° no YouTube e identifica sinais que pontuam tendências emergentes para o jornalismo imersivo. O capítulo entrelaça os estudos de tendências com os métodos digitais e tem como pressuposto o jornalismo brasileiro. Flores situa o estudo a partir dos conceitos de “tendência” e “jornalismo de inovação” e, igualmente, através das instâncias que sinalizam as mudanças no jornalismo contemporâneo. Por exemplo, conteúdo versus narrativa e tecnologia versus formato. Desta forma, sugerem o pensar a atividade jornalística para além das práticas recorrentes do campo. Neste contexto, Flores mapeia e analisa a produção de vídeos 360° de canais como Folha de S. Paulo, Valor Econômico, BBC Brasil, Diário de Pernambuco, entre outros. Em paralelo, interroga a cultura de classificação algorítmica da plataforma ao verificar o que os brasileiros encontram (e consomem) quando realizam buscas por “jornalismo” e “vídeo 360”.

Dois aspectos merecem destaque neste estudo: o processo de investigação desenhado a partir dos métodos digitais (e transformado por meio das idas e vindas entre extração-análise-visualização exploratória de dados) e os resultados obtidos. Estes revelam que o YouTube não recomenda vídeos 360° produzidos pelos canais jornalísticos brasileiros, mas, sim, aqueles de caráter acadêmico criados por estudantes de jornalismo. Um outro ponto destaca a natureza de consumo destes vídeos: as temáticas sobre natureza, turismo e desporto parecem ser mais atraentes. Para além dos ricos *insights* apresentados no capítulo, talvez o principal seja o resultado de que “a maioria dos vídeos produzidos pela mídia jornalística no Brasil, entre dezembro de 2015 e abril de 2018, não se apropria de estratégias narrativas específicas para imersão”. Flores, portanto, conclui que a apropriação do vídeo 360° é ainda experimental no jornalismo brasileiro.

O capítulo de André Mintz e Carlos d’Andréa propõe uma abordagem multiplataforma para o estudo de circulação de imagens. O modelo apresentado pelos

autores tem como base o Localizador Uniforme de Recursos – URL (como identificador único de uma determinada imagem) e a análise de imagens por intermédio de visão computacional (*Vision APIs*). Apesar de introduzirem o estudo como experimental, a proposição teórica-prática-crítica de Mintz & d'Andréa merece ser tomada como exemplo para os estudos de conteúdos visuais nas plataformas de media sociais. Os autores sistematizam a metodologia e um estudo de caso relatando de forma detalhada os desafios inerentes à visão computacional (condições e restrições na tarefa de recuperação de imagens baseada em conteúdo) e à coleta de dados (seguir e rastrear a circulação de imagens em tempo real). Do mesmo modo, discutem as possibilidades analíticas via decomposição das URLs e as suas implicações metodológicas. O Mundial de Futebol de 2018 serviu de cenário para o estudo de duas das imagens mais “retuitadas” durante a transmissão ao vivo do sorteio final dos grupos. O estudo é desenvolvido a partir da identificação de domínios (*link domains*) e a sua relação segundo ocorrências em diferentes plataformas, localidade geográfica e distribuição entre plataformas ao longo do tempo. Em resposta às limitações impostas pelo caráter circunstancial deste tipo de estudo, Mintz & d'Andréa propõem a noção de *horizonte de recuperabilidade* para que estas limitações sejam consideradas reflexivamente e como “indícios para uma compreensão das dinâmicas que são simultaneamente constituintes e constitutivas da realidade observada”.

REFERÊNCIAS

- Latour, B., Jensen, P., Venturini, T., Grauwin, S., & Boullier, D. (2012). “The whole is always smaller than its parts” – a digital test of Gabriel Tarde’s monads. *British Journal of Sociology*, 63(4), 590-615.
- Marres, N. (2017). *Digital sociology: the reinvention of social research*. Bristol: Polity Press.
- O’Reilly, T. (2005). What is Web 2.0: Design patterns and business models for the next generation of software. Recuperado de <https://www.oreilly.com/pub/a/web2/archive/what-is-web-20.html>
- Rogers, R. (2010). Internet Research: The Question of Method – A Keynote Address from the YouTube and the 2008 Election Cycle in the United States Conference. *Journal of Information Technology & Politics*, 7(2-3), 241-260.
- Rogers, R. (2013). *Digital Methods*. Cambridge, MA: MIT Press.
- Rogers, R. (2015). Digital Methods for Web Research. *Emerging Trends in the Behavioral and Social Sciences*, 1-22.
- Rogers, R. (2019). *Doing digital methods*. London: Sage.
- Venturini, T., & Latour, B. (2010). The Social Fabric: Digital Traces and Quali-quantitative Methods. In *Proceedings of Future En Seine, 2009* (pp. 87-101). Paris: Editions Future en Seine.



UM

FUNDAMENTOS

MÉTODOS DIGITAIS: DOS DESAFIOS À *BILDUNG*¹

BERNHARD RIEDER
THEO RÖHLE

Em artigo anterior (Rieder & Röhle, 2012), argumentamos que as tradições das humanidade e ciências sociais existentes, incluindo suas particularidades, interesses e métodos, recorrentemente encontram um objeto — o computador —, que é caracterizado por suas próprias lógicas, logísticas, estilos de raciocínio (Hacking, 1992), hábitos, (melhores) práticas, modos de valorização, redes ator-rede e instituições. O computador pode muito bem ser um objeto técnico comedido, mas sua história acumulada — seu conteúdo, portanto — está plena de elementos heterogêneos, os quais constituem um tipo de *a priori* que não pode ser facilmente ignorado. Agora que várias tentativas são feitas para construir versões ou extensões “digitais” de disciplinas por muito tempo estabelecidas, este encontro marca um momento de desestabilização e desterritorialização, um momento que implica contingência significativa e diferentes resultados possíveis. Apesar de parecer duvidoso que até mesmo a “ciência normal” de Kuhn (1962) já tenha sido de fato estabelecida, este é um momento que provoca e requer debate de longo alcance e investigação sobre a prática, o sentido e o propósito de nossas disciplinas acadêmicas.

O encontro entre as humanidades e a computação se desenrola de diferentes maneiras, em diferentes arenas, mas precisa ser abordado em princípio, bem como em relação a configurações particulares. O fato de, que, após 50 anos de experimentação, muitas das questões fundamentais permanecerem extremamente controversas, pode ser visto como um indicador do quanto essas questões se aproximam dos debates centrais a respeito dos sentidos e dos propósitos do

¹ O termo alemão *Bildung* pode ser traduzido como formação intelectual relacionada à educação e à cultura, no sentido de que ter *Bildung* é ser culto, ter cultura, e não ter *Bildung* é ser inculto, sem cultura.

trabalho acadêmico. Enquanto termos como “humanidades digitais”, “análises culturais” (*cultural analytics*), “métodos digitais” ou “*web science*” podem figurar como palavras da moda, a proliferação deles pode ser vista como indicador de uma “virada computacional” (Berry, 2011a), muito mais profunda do que a simples ascensão de modos de análise quantitativos ou “científicos”. Amplos e, muitas vezes, incomuns conjuntos de dados, técnicas de visualização avançadas e processamento impreciso conduziram, por muito, alguns daqueles que tinham em mente números, cálculos e computadores, a uma distância segura para prepararem as novas possibilidades computacionais. Assim, a nossa questão central era: se esses novos métodos são mais do que apenas outro conjunto de ferramentas em nosso arsenal, como lidamos com as transformações fundamentais que desafiam práticas e paradigmas epistemológicos estabelecidos?

O ponto de partida para a nossa investigação anterior era o conceito de “método”. Este é definido pelo Dicionário Oxford como “busca por conhecimento, modo de investigação”, assim, somos lembrados que essa busca é tanto sistemática quanto ordenada. Ademais, o método é direcionado e intencional: decisões específicas estão atadas a objetivos específicos. Assim como uma planta baixa na arquitetura ou uma receita, os métodos de pesquisa guiam o comportamento e, mesmo se algumas de nossas abordagens são somente explícitas de maneira moderada, suas afinidades ou semelhanças nos possibilitam compartilhar experiência e estabelecer pontos de referência que fornecem orientação — mesmo quando há pouco acordo a respeito da utilidade e da validade.

Embora sejamos cautelosos em relação à avaliação feita por Tom Scheinfeldt a respeito de nossa era, por ele considerada como uma “era pós-teórica” (Cohen, 2010), o diagnóstico que ele faz de um “momento metodológico” é, certamente, apropriado. Da tradição acadêmica alemã, desenvolvemos nossa perspectiva contra um cenário de décadas de *Methodenstreit* (“disputa de método”), começando com o endosso de Weber a respeito da sociologia, por ele considerada como uma disciplina de “entendimento” (*verstehend*) em vez de “explicação” (*erklärend*), que depois se tornou a famosa *Positivismusstreit* (“disputa do positivismo”) entre Adorno e Popper. Parte disso foi, por vezes, a profunda, paralisante e estéril oposição entre métodos de pesquisa quantitativos e qualitativos na ciência social empírica. Embora não tão análoga ao problema das “duas culturas” em Snow (1959), certamente há paralelos que apontam para diferentes modos de saber e pensar — estilos de raciocínio — alcançados em horizontes normativos maiores, como visto nas discussões entre tipos de pesquisa “crítica” e “administrativa”, resumida pelo confronto entre Adorno e Lazarsfeld.

Nossa recusa em ceder a simples oposições é construída sobre uma abordagem antiessencialista a muitos dos conceitos que aparecem em tais debates. Computação, quantificação, algoritmo, visualização, grafo, análise de dados, estatística, software, e assim por diante, são termos que apontam para conceitos — mas também para conjuntos de objetos, práticas e habilidades — que são considerados ter heterogeneidade e variação internas significativas. Isso não significa que esses conceitos não são apreendidos em configurações e constelações particulares, que são produtivas em formas muito específicas em termos de conhecimento e poder; mas isso quer dizer que os espaços de design e “apropriação” (Akrich, 1998) dos métodos computacionais oferecem margem de manobra considerável e não performam ou se traduzem em lógicas singulares. Mesmo se o “digital” se tornou um ponto de passagem obrigatório, ele funciona como um moedor de carne: o material triturado não sai como um único pedaço, mas como muitos.² Para conectar novamente o *Methodenstreit*, os métodos computacionais podem ser tanto dedutivos quanto indutivos (ver, por exemplo, o conceito de análise exploratória de dados de Tukey, 1962), tanto quantitativos quanto qualitativos, em perspectiva, tanto críticos quanto administrativos. Entretanto, esses espaços de movimento, de liberdade epistêmica, devem ser construídos e defendidos, algumas vezes, por alianças forjadas, e outras, por demarcação; certamente por meio de um melhor entendimento a respeito daquilo com o qual os computadores podem contribuir para a produção de conhecimento, e as maneiras como eles produzem esse “excedente” epistêmico.

Se a metodologia digital é configurada para mudar a maneira como os estudiosos trabalham com seu material, como eles o “veem” e com ele interagem, uma questão iminente é como tais métodos afetam o modo como geramos, apresentamos e legitimamos o conhecimento nas humanidades e ciências sociais. De que maneira as propriedades técnicas dessas ferramentas são constitutivas do conhecimento gerado? Quais são as habilidades técnicas e intelectuais que precisamos dominar? O que significa ser um pesquisador na era digital? De modo geral, as respostas a essas questões dependem de quão bem somos capazes de acessar criticamente as transformações metodológicas que continuamente testemunhamos.

Como sugerem uma gama crescente de investigações a respeito do status do *big data* (Gitelman, 2013; Elmer, Langlois & Redden, 2015; Amore & Piotukh, 2015), bem como as discussões em curso nas humanidades digitais (Gold, 2012;

² Para uma investigação detalhada a respeito de diferentes tipos de processamento digital, conferir: Winkler (2015) — a capa do livro apresenta o uso metafórico do moedor de carne.

Arthur & Bode, 2014; Svensson & Goldberg, 2015), há algo profundamente complicado acerca deste momento metodológico. Argumentos que, se algumas críticas que estão sendo formuladas para o campo mais amplo das humanidades digitais são, de fato, justificadas, isso não deveria ser visto como desencorajamento, mas como um desafio, no sentido mais cativante do termo.

Neste capítulo, gostaríamos de rapidamente sintetizar o que consideramos ser os cinco desafios principais antes de questionarmos o conceito de “*Bildung* digital” (Berry, 2011a) como um meio de enfrentar esses desafios. Nosso objetivo com essa discussão é, talvez de modo paradoxal, direcionar o holofote do “digital” e das habilidades associadas a ele — programação — para a plethora de conceitos e *conhecimentos* mobilizados pelas ferramentas digitais. Para isso, discutimos três exemplos que nos possibilitam concretizar e complicar o debate.

CINCO DESAFIOS

Em artigo anterior (Rieder & Röhle, 2012), apresentamos uma lista não exaustiva de grandes grupos de questões, que devem ser abordadas, se quisermos integrar os novos métodos, de maneira produtiva, sem renunciarmos ao controle sobre a infraestrutura conceitual de nosso trabalho. A nossa questão não era sobre como conduzir uma “boa” pesquisa digital em *stricto sensu*: não estávamos preocupados com armadilhas metodológicas específicas ou “erros” nas coleções de dados, ou com as escolhas e aplicações de ferramentas metodológicas, mas com as mais amplas ramificações da pesquisa digital no campo das humanidades e ciências sociais. Nesse sentido, queríamos encarar os desafios enfrentados até mesmo pelo “melhor” trabalho de campo.

Um primeiro desafio, que chamamos “O Fascínio da Objetividade”, suscitou a questão de por que as ferramentas computacionais provocaram tamanho interesse quando se trata de estudar questões sociais ou culturais. Uma possível explicação diz respeito à noção de que o computador é capaz de ir além das particularidades humanas e em direção ao domínio da objetividade. Discutimos o fascínio que o ideal de raciocínio imparcial e mecânico era capaz de historicamente induzir, e questionamos se esse fascínio poderia nos impedir de desnudar as muitas decisões explícitas e implícitas que adentraram em nossas ferramentas e instrumentos. Questões de preconceito e subjetividade, que o computador foi pensado para eliminar, novamente entraram em um plano menos tangível — por meio de decisões relativas aos modos de formalização e procedimentos algorítmicos, bem como por meio das várias formas de processamento de dados

que podem mascarar a parcialidade (cf. Baroca & Selbst, 2015). Isso se tornou, sobretudo, um problema central quando estudamos plataformas comerciais de mídias sociais. Ao considerarmos as “políticas de circulação” (Beer, 2013) nas quais essas plataformas estão incorporadas, bem como os resultantes ecossistemas elaborados de regulações de API (Bucher, 2013; Puschmann & Burgess, 2014; Rieder et al., 2015), as questões concernentes à pré-seleção constituem um grande dilema metodológico. O desafio é, então, aceitar o fato de que, em um nível epistemológico, os métodos computacionais frequentemente criam complicações em vez de resolvê-las.

Sob o título “O Poder da Evidência Visual”, discutimos o papel de produções visuais, como representações de topologias de rede, *timelines* ou cartografias sofisticadas. Uma vez que essas visualizações possuem extraordinárias qualidades estéticas — e então retóricas —, questionamos como o poder argumentativo das imagens poderia (e deveria) ser criticado. Ressaltamos a tradição de investigação crítica a respeito do uso de imagens das quais as humanidades se alimentaram ao longo dos anos, mas remarcamos que a situação atual, de fato, mudou, desde que as humanidades digitais produzem imagens e delas dependem como evidência e dispositivos heurísticos. Assim, o desafio é manter uma autorreflexão produtiva a respeito de nossas próprias práticas visuais, para reconhecermos como a análise e a cognição são — recorrentemente e historicamente (Halpern, 2015) — parciais e entrelaçadas com relações de poder, sem abandonarmos a promessa de adquirirmos *insights* via formas visuais (Ducker, 2014, pp. 130-137).

Utilizamos o termo “Encaixapretamento” (*black-boxing*) para nos referirmos à nossa habilidade de entender o método, de ver como ele funciona, em quais pressupostos ele é construído, para reproduzi-lo e criticá-lo. Apesar de a escrita de softwares nos forçar a explicitar todos os procedimentos, expondo-os como códigos de programação, a “legibilidade” não está, de forma alguma, garantida. Contudo, um processo aberto de escrutínio é um dos pilares do conhecimento e, ao final, de reivindicação de conhecimento para a legitimidade social. Argumentamos que esse problema se apresenta em ao menos três níveis:

- a) possibilidade prática para acessar a camada mais óbvia de especificação funcional, como um código-fonte de ferramenta,
- b) habilidade para entender o código e, ainda mais importante, a habilidade de compreender suas ramificações epistemológicas,
- c) os métodos que se tornaram opacos apesar de serem completamente explicitados, tais como técnicas provenientes do campo de aprendizado de máquina, em que as conexões feitas entre entradas e saídas (*inputs* e

outputs) não podem mais serem facilmente retraçadas por observadores humanos.

Este ponto realmente concerne à questão de como o excedente epistemológico que é oferecido pela computação pode ser especificado, controlado e retransmitido para outros, sem ser, ocasionalmente, vítima da simplicidade enganosa de interfaces gráficas de usuários e de visualizações brilhantes.

Nós identificamos as “Perturbações Institucionais” como um quarto conjunto de desafios. Vimos uma chance de, que, dada a crescente necessidade de conhecimento computacional, as humanidades poderem, cada vez mais, contratar pesquisadores adeptos à computação para suas disciplinas. Além disso, os métodos computacionais podem ter vantagens em situação em que até mesmo a pesquisa humana é, cada vez mais, financiada com base em projetos — o que implica uma pragmática muito particular, baseada em quadros de estruturação de prazos, expectativas planejadas e “resultados” que podem ser identificados e entregues. Defendemos que o desafio é desenvolver a sensibilidade para essas amplas repercussões de inovação metodológica. Em muitas áreas, há um argumento a ser construído, em defesa de métodos baseados em princípios, em vez de métodos baseados em “esforço contínuo” mecanizado (Wang, 1963, p. 93).

A quinta questão que sublinhamos foi “A Busca pelo Universalismo”. Argumentamos que o estabelecimento de conceitos e princípios difusos se tornou, cada vez mais, comum sempre que os computadores entram em jogo. Quando se percebe que a realidade adere a um sistema de regras específico, o computador aparenta ser a ferramenta por excelência para representar esse sistema e para calcular suas dinâmicas. Os deveres epistemológicos e a natureza redutora de modelos subjacentes são, muitas vezes, “esquecidos” quando se trata das explicações derivadas deles. Em vez disso, o escopo das explicações é estendido indefinidamente, reminescente das aspirações universalistas que atravessam os discursos históricos a respeito da computação. Conceitos advindos da ciência das redes (*network science*) são um caso em questão. O desafio é, portanto, chegar a uma demarcação mais adequada da explicação acerca do alcance de modelos formais, por exemplo, combinando diferentes configurações metodológicas, tanto digitais quanto não digitais.

Em termos de conclusão, continuamos a defender o envolvimento com os novos métodos. Por envolvimento, queremos dizer tanto a atual aplicação desses métodos quanto uma reflexão crítica a tais usos. Assim, defendemos uma transferência do conceito de “prática técnica crítica”, proposto por Agre (1997a), para o domínio acadêmico: uma prática que oscila entre o trabalho técnico concreto e

a reflexividade metodológica. Abordagens atuais que se baseiam no conceito de Agre são muito promissoras nesse sentido. Como argumentam Matt Ratto, Sara Ann Wylie e Kris Jalbert (2014), o engajamento atual com a materialidade – o que eles chamam de “construção crítica” – pode ser um complemento produtivo das tradicionais formas linguísticas de produção de conhecimento, também em campos como os STS (*Science and Technology Studies*) e os estudos de mídias (*media studies*). Em vez de desenvolver métodos com um objetivo claro em mente, o processo de design pode ser um meio para avançar uma atitude mais inquisitiva em relação aos nossos ambientes digitais – “trazendo aspectos inconscientes da experiência à consciência e, assim, tornando-os disponíveis para a escolha consciente”, como afirmam Sengers e outros (2005, p. 50) ao esboçarem a ideia de “design reflexivo”.

A seguir, focaremos especificamente no desafio do “encaixapretamento” e, de maneira geral, no papel das ferramentas digitais em constelações de pesquisas emergentes. Todos esses desafios, entretanto, conectam-se, mais ou menos diretamente, à pergunta acerca do que precisamos saber para tornar realidade essa prática crítica, reflexiva, inquisitiva e cheia de nuances. Assim, nós nos voltamos para a questão do conhecimento e da habilidade, que tem sido discutida com particular vigor nas humanidades digitais, muitas vezes com o foco na programação, considerada como a expertise de bacias hidrográficas que separa “quem está dentro e quem está fora” (Ramsay, 2011). Consideramos essa ênfase a fim de codificarmos uma compreensão um pouco redutora acerca da computação, e sugerimos uma apreciação mais profunda e conceitual do conhecimento técnico e prático, face a um arsenal cada vez maior de métodos digitais.

DOS DESAFIOS À BILDUNG

Nesta seção, abordamos a questão dos desafios para as humanidades (digitais) e para as ciências sociais à luz da noção de “*Bildung* digital” proposta por Berry (2011b, p. 20), que diz respeito a “uma arte liberal que é para ‘todos os humanos’”, embora iremos nos concentrar em humanistas digitais e cientistas sociais em vez de focar em um público geral.³ A nossa questão trata do que precisamos saber para nos tornarmos acadêmicos digitais capazes de “examinar, teorizar, criticar e imaginar” (Berry, 2011b, p. 169) uma metodologia de pesquisa – a busca sistemática e racional

³ A descrição de Berry (2011b, p. 22) a respeito do “*Bildung* digital” como “um processo dinâmico de pensamento reflexivo e de um repensar colaborativo” aparenta compartilhar várias características com tradições do design, que evocam a noção de “reflexão-em-ação” de Donald Schön (1983), como também o faz Agre (1997b, p. 10).

do conhecimento — que é alcançada pela computação. Fundamentalmente, acreditamos que esse debate permanece vago e superficial sem um conjunto concreto de referências. Por isso, iremos discutir três exemplos, que esperamos contribuir com uma discussão mais aprofundada sobre como os desafios por nós identificados podem ser relacionados com uma noção mais ampla de “*Bildung* digital”.

Uma questão central nessa discussão é se é possível (ou desejável) treinar humanistas “computacionalmente iluminados”, que irão, eles mesmos, escrever os métodos computacionais que aplicarão em suas análises. Sustentamos que essa noção é tentadora, mas, em última análise, irrealista e até potencialmente problemática: enquanto alguém pode aprender a escrever um código em poucos dias, a prática de programação ou desenvolvimento de software requer uma aculturação extensa e muitas, muitas horas de prática. Se considerarmos a disponibilidade de tempo como um recurso limitado, a prioridade dada à programação pode, na verdade, prejudicar outras habilidades técnicas e conceituais, que facilitam a compreensão crítica de procedimentos computacionais. O foco específico no código pode prejudicar o que realmente é codificado.

Isso porque, para qualquer programador experiente, o código pode muito bem ser o meio de expressão, mas, assim como um escritor tenta dizer algo por meio da linguagem, o sentido expresso pela programação é funcionalidade; e enquanto os dois não podem ser completamente separados, programadores e cientistas da computação geralmente raciocinam em um nível conceitual que certamente é circunscrito pelos requisitos da computação mecânica — o que um de nós chamou de “sombra da computação” (Rider, 2012) —, mas exprimível em várias formas, de um vocabulário e conversação sistematizados a fluxogramas e, mais frequentemente, à notação matemática. Enquanto a implementação certamente não é irrelevante, o cerne metodológico, a definição daquilo que a computação adiciona, reside naquilo que o programa faz. Este nível funcional pode ser de uma complexidade assustadora, mesmo que várias técnicas sofisticadas possam ser resumidas a um pequeno número de ideias. Considerar essas ideias sob a ampla noção de “o digital”, limita a análise a uma visão superficial, que corre o risco de esconder a dimensão metodológica fundamental ou racional do trabalho desempenhado pelos métodos que são realizados em software. Enfrentar o desafio destacado antes depende, ao menos em parte, de sermos capazes de chegar ao núcleo conceitual das técnicas computacionais que utilizamos. Só então poderemos acessar os potenciais, as limitações e os estilos de raciocínio adquiridos pelas ferramentas que integramos às nossas configurações de pesquisa.

Para aprofundar este argumento, voltamo-nos para três exemplos, que possibilitam uma abordagem diferenciada e destacam a dificuldade de definição

de princípios gerais. Em todos os exemplos, questionamos o que pode significar “entender” uma técnica computacional.

ESTATÍSTICAS

Uma vez que as ciências sociais, empíricas, têm usado as ferramentas digitais como parte integral de seu trabalho, por décadas, a estatística aplicada é um bom lugar para começar. Um dos pacotes de software mais utilizado nas ciências sociais é o SPSS (*Statistical Package for the Social Sciences*), e a significativa dependência dos pesquisadores a este programa coloca a questão de em que medida os pesquisadores são capazes de “entender” – ou mesmo buscar entender – as consideráveis escolhas metodológicas e epistemológicas feitas por várias técnicas analíticas oferecidas. Se considerarmos, por exemplo, a análise de regressão, uma técnica que é extremamente produtiva (literalmente nenhum aval implicado) em pesquisa acadêmica, bem como em setores empresariais e do governo, como meio de produzir um excedente epistêmico, como poderíamos entender, de maneira mais precisa, em que consiste a técnica e sua contribuição intelectual?

O código-fonte do SPSS não está disponível, mas o modo como o software calcula suas medidas analíticas é bem documentado em notação matemática, e se baseia em construtos estabelecidos e muito bem discutidos, como o coeficiente de Pearson para correlação (r) ou técnicas de regressão estabelecidas. Olhando para uma alternativa de código-aberto, o que nós realmente ganhamos ao ler o código-fonte em vez de simplesmente consultarmos a documentação e verificarmos os papéis de pesquisa aos quais ele se refere?

Enquanto uma crítica à padronização e à racionalização da pesquisa por meio de pacotes de software amplamente disponíveis é importante e suscita muitas preocupações⁴, isso não nos diz como a agência epistemológica pode ser novamente integrada às ferramentas que realizam procedimentos metodológicos extremamente complexos, disponíveis por meio de simples interfaces gráficas. Uma crítica às ferramentas digitais é incompleta sem uma crítica aos seus *usuários*, bem como às configurações mais amplas nas quais eles estão incorporados. Por mais banal que isso possa soar, o que é necessário para se entender e usar o SPSS reflexivamente – ou qualquer pacote de estatística para esse fim – é um conhecimento robusto de estatística e teoria da probabilidade, e não um curso intensivo em Java.

⁴ Para uma discussão aprofundada acerca do significado de SPSS para a sociologia, conferir: Uprichard et al. (2008).

O que se torna uma caixa preta no que tange a tal ferramenta não é meramente um conjunto de procedimentos de cálculo, que são, ao final, bem documentados, mas a estatística — um campo que não tem apenas o seu próprio conteúdo epistemológico, porém, conta com muitos debates internos, contradições e divergências. As “treze maneiras de olhar para o coeficiente de correlação”, identificadas por Rodgers e Nicewander (1988), bem como os debates acerca do teste de hipóteses nulas, ao qual Gigerenzer, Krauss e Vitouch (2004) se referem como “ritual nulo”, são apenas dois dos muitos exemplos para os desacordos bastante fundamentais na prática de estatística aplicada. Enquanto o software pode ser planejado de modo a destacar essas divergências, é pedir muito a um programa que carregue o peso de fornecer uma educação no campo em que é produzido. Isso suscita e complica a questão da incorporação educacional de ferramentas digitais. Se estudantes e pesquisadores são treinados para usar essas ferramentas sem atenção considerável aos espaços conceituais que eles mobilizam, os resultados podem ser bastante problemáticos. A *Bildung* digital (educação digital) requer, portanto, atenção não apenas à forma do software, mas também aos atuais conceitos e métodos expressos e operacionalizados por meio de procedimentos computacionais.

ANÁLISE DE REDES

Um argumento similar pode ser feito para o popular campo de visualização de rede. Novamente, é importante notar que as formas de ponto e linha vêm com suas próprias implicações e obrigações, e análises de grafos e ferramentas de visualização, como o *Gephi* (Bastian et al., 2009), também estruturam o processo de pesquisa. Mas aonde vamos a partir daí? Se considerarmos que a teoria dos grafos ainda oferece meios eficazes e interessantes para analisar conjuntos de dados, com o que se pareceria a prática analítica e crítica? Por exemplo, como podemos considerar os algoritmos de layout que transformam matrizes⁵ de adjacência com n dimensões em diagramas de redes bidimensionais? Esses artefatos se interpõem como mediadores porque cada algoritmo revela o grafo de maneira diferente, destacando aspectos específicos de sua estrutura, portanto, produzindo uma *interpretação* específica.

Há diferentes famílias de algoritmos — muitas abordagens se baseiam em simulações de força, mas outras estratégias, como a *simulated annealing* também

⁵ Uma matriz adjacente é um modo de representar um grafo como um tipo especial de tabela (uma matriz quadrada — mesmo número de linhas e colunas), que especifica quais nós estão conectados entre si.

existem —, apesar de o mesmo algoritmo, quando alimentado com diferentes parâmetros, poder produzir resultados completamente diferentes. Se aplicarmos o algoritmo *ForceAtlas2* (Jacomy et al., 2004) a um arquivo de grafo, devemos ir ao repositório fonte do *Gephi* no *Github* e pesquisar pelo arquivo *ForceAtlas2.java* e tentar dar sentido a isso? O que encontraríamos lá? Milhares de linhas de código Java que implementam uma simulação altamente interativa de forças de atração e repulsão, que fazem amplo uso da noção de *swinging* (em um sentido bem literal) para encontrar uma posição “ideal” para os nós na tela, sem que estes fiquem presos a essa posição.⁶ É muito ingênuo acreditar que alguém sem treinamento considerável em programação e modelagem de simulação possa dizer alguma coisa significativa sobre como o *ForceAtlas2* implementa o conceito de força direcionada de maneira distinta de seu antepassado histórico e conceitual, o trabalho de Fruchterman e Reingold; e muito menos como essas diferenças afetam a espacialização em circunstâncias concretas. Como as propriedades de nós e a estrutura topológica afetam as posições no mapa? Quais aspectos das estruturas latentes nos dados o diagrama revela?

Mesmo com o treinamento desejado, testar e rodar o algoritmo em diferentes conjuntos de dados, com diversos parâmetros, é necessário para começar a entender como os resultados se relacionam com instâncias computacionais, pois nenhum cérebro humano pode antecipar o espaço resultante de funções simples, repetidas milhares de vezes. Outra vez, o problema advém do fato de ferramentas como o *Gephi* terem tornado a análise de redes acessível a amplas audiências, que rapidamente produzem diagramas de rede sem terem adquirido conhecimento robusto dos conceitos e das técnicas mobilizados pelo software. Cada vez mais, isso leva a uma falta de consciência das camadas de mediação implicadas na análise de rede e, portanto, a leituras limitadas ou essencialistas das saídas (*outputs*) que foram produzidas, que perdem seu caráter artificial e *analítico*. Uma visualização de rede está mais próxima de um coeficiente de correlação do que um mapa geográfico, e precisa ser tratada adequadamente.

Novamente argumentamos que o domínio crítico do conteúdo metodológico introduzido pelo software teria sido melhor aproveitado se o estudo do material na teoria dos grafos, na espacialização gráfica e, em particular, na literatura acerca de aplicações analíticas concretas, tivesse sido feito. Olhando para a história e o estado da arte da sociometria e da ciência das redes, seria útil adquirir

⁶ Considere um problema análogo: um simples algoritmo para escaladas, que consista em “sempre subir”, irá terminar no topo de uma montanha (um local ideal), mas não necessariamente no mais elevado (o local global ideal). O *swinging* neutraliza um problema similar: ficar “preso” em um ótimo local.

“alfabetização em grafos”. Para ser mais específico, um bom começo seria se aprofundar no estudo de Freeman, *The Development of Social Network Analysis* (2004). Inevitavelmente, é crucial gastar tempo considerável testando diferentes algoritmos em conjuntos de dados distintos para adquirir conhecimento a respeito de como eles especificamente interpretam um grafo. A prática reflexiva requer muito mais do que uma atitude crítica: requer um envolvimento *profundo* com espaços de conhecimento associados para dar sentido às possibilidades e limitações.

MILHARES DE IMAGENS

Os dois exemplos mencionados certamente não são totalmente representativos das ferramentas utilizadas no campo, mas o argumento pode ser estendido para além dos pacotes de software mais complexos que acabamos de abordar. O trabalho feito por Lev Manovich (2012) sob o rótulo de “Análise Cultural” (Cultural Analytics) pode servir como exemplo: para ordenar imagens de mangá, em preto e branco, em um gráfico de dispersão, Manovich utilizou “entropia calculada sobre os valores em escala de cinza de todos os pixels em uma página” para o eixo y e, após apontar para a história do conceito de entropia, ele explica que essa medida expressa em termos das imagens em questão: “Se uma imagem consiste em algumas áreas monocromáticas, sua entropia será baixa. Em contraste, se uma imagem tem várias texturas e detalhes, e suas cores [...] variam significativamente de um lugar a outro, sua entropia será alta” (Manovich, 2012, p. 266).

Independentemente do que pensamos sobre o que Manovich está fazendo com essas imagens, em termos intelectuais (História da arte? Ciência da imagem?), é o seu treino considerável e sua experiência em trabalhar com imagens digitais o que o possibilita relacionar, de modo confiante, uma medida matemática com as propriedades visuais atuais das imagens em questão. Não estamos qualificados para dizer se os resultados alcançados por Manovich dessa operação realmente são úteis para seus objetivos analíticos, mas essa não é a questão aqui. O que importa é a habilidade aplicada nesse exemplo, isto é, a capacidade de raciocinar sobre imagens, em termos matemáticos, conectar esses termos a propriedades visuais da imagem tal como ela é percebida pelos seres humanos, e derivar um excedente epistêmico a partir de toda essa operação. O que ganharíamos ao olhar para o código-fonte do script desenvolvido por Manovich? Talvez encontrássemos um erro. Talvez pudéssemos chegar a uma implementação mais eficiente. Embora Manovich não forneça a métrica usada na notação matemática (por que não?), a

referência feita a Claude Shannon é uma boa razão para acreditar que a medida de entropia em questão é algo como $-\sum (p^* \log_2(p))$, em que p contém o histograma da imagem em 256 células se a imagem estiver codificada em 8-bit.

Agora, assim como os famosos quatro conjuntos de dados de Anscombe (1973), que são bem diferentes na estrutura, mas têm as mesmas propriedades estatísticas, uma medida muito sintética como a entropia, que expressa algo acerca de um complexo objeto como uma imagem em um único número, pode rotular um amplo número de diferentes imagens com valores iguais. Assim, Manovich não apenas teve de se comprometer com a medida de entropia como tal, mas também com a medida de entropia tal como ela reage ao conjunto de dados em questão. Pelo que entendemos, uma escala de cinzas teria um alto valor de entropia desde que o histograma não tivesse nenhuma informação a respeito de como as cores são espacialmente distribuídas; é uma simples contagem de ocorrência para todas as cores. Será que um certo estilo de coloração em uma mangá “quebraria” a medida? Para certos conjuntos de dados – Barnett Newman ou Piet Mondrian talvez? –, a medida poderia ser completamente inútil porque o elemento saliente seria o arranjo de superfícies em vez da probabilidade de distribuição das cores.

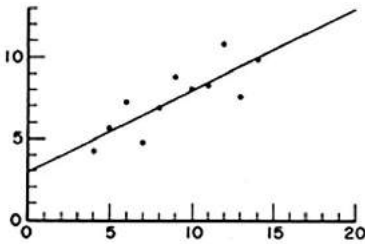


Figure 1

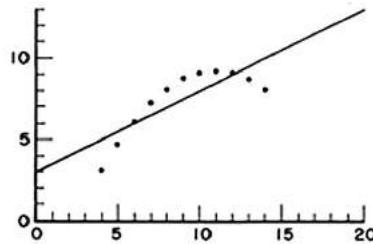


Figure 2

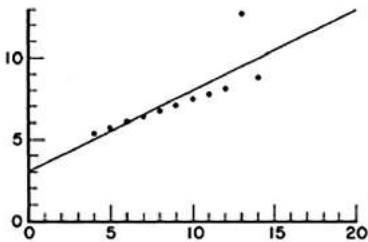


Figure 3

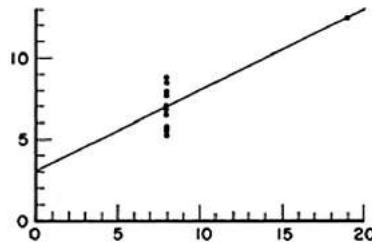


Figure 4

FIGURA 1 – Os quatro diagramas de dispersão de Anscombe (1973). Eles têm valores idênticos por número de observações, média de x , média de y , coeficiente de regressão de y em x , equação de linha de regressão, soma de células de x , soma de regressão de células, soma residual de células

de y , erro padrão estimado de b_i e r -quadrado múltiplo. Ascombe utiliza tais valores como argumento para a utilidade da visualização na estatística.

Não há dúvidas de que as habilidades de programação são úteis nesse contexto. Todavia, a entropia não é um conceito de “programação”; ela é, como muitas medidas estatísticas, um meio para sintetizar dados, um meio para falar sobre os dados de um ponto de vista muito particular. Ela é redutora, certamente, mas redutora em um *modo específico*, e aí reside seu caráter epistêmico. Como conceito, a entropia se liga às complexas histórias da teoria da informação e da estatística,⁷ e o uso reflexivo terá de atentar para essas conexões.

Esse é o trabalho que humanistas e cientistas sociais têm de fazer. Eles não podem simplesmente delegar essa tarefa a colaboradores das ciências da computação ou contratarem programadores para realizá-la. Percebam que essa é uma discussão técnica muito complexa, que não possui nenhuma questão acerca de programação. Qualquer programador, um pouco capaz, poderia produzir um script a partir da especificação “calcular a entropia do histograma em escala de cinzas” e, em ambientes como o MATLAB, existem até mesmo funções predefinidas que fazem todo o trabalho para nós. Primeiro e acima de tudo, os verdadeiros “conteúdo” e compromisso metodológicos não são apenas uma questão de software. Certamente que nós só podemos fazer isso, em primeiro lugar, porque há software, e as interfaces escondem e consolidam nossos compromissos, mas o conhecimento requerido para julgar o método em questão está somente, em uma pequena parcela, relacionado com a questão do código; em vez disso, abrange um espaço da teoria da informação para a história da arte e os estudos visuais de uma maneira que certamente envolve abstração, mas um tipo distinto daquele que a programação exige.

CONCLUSÕES

Ao passo que nossos três exemplos poderiam ser considerados muito específicos, pensamos que argumentos similares poderiam ser feitos para uma ampla variedade de casos em que o software performa um *método*. Embora os conceitos metodológicos e técnicos entrem em negociação com a implementação, o “conteúdo” do software é um procedimento *expresso* em código, e não simplesmente

⁷ Para um relato dessas histórias, acessíveis e interessantes a humanistas, conferir a contextualização feita por Christian Kassung (2001, pp. 132-260) do livro *Man Without Qualities*, escrito por Robert Musil, no contexto da física moderna.

código. Certamente podem encontrar casos em que a dimensão matemática de uma ferramenta é completamente trivial, mas argumentaríamos que, em muitas das ferramentas utilizadas por pesquisadores, um trabalho metodológico significativo é performed por técnicas, que têm sua origem no domínio conceitual de disciplinas como estatística, ciência da informação, sociometria, ciência da computação e – muitas vezes – matemática.

Esse é o ponto crucial aqui. Apesar de concordamos totalmente com Berry (2012) que a *Bildung* digital – em particular para o humanista digital, mas também para além disso – se beneficiaria da “iteração [...] amplamente definida como competência na leitura, escrita e execução de código de computador”, o foco na programação como “código de escrita”, em vez de “implementar uma técnica”, corre o risco de perder este nível mais conceitual que é, a nosso ver, epistemologicamente mais relevante para bolsas de estudo que contam com ferramentas digitais e, em muitos casos, mais acessíveis em termos de habilidades a serem adquiridas.

Embora nossa evidência seja apenas anedótica, em muitas das humanidades, notamos um desejo de explicar a tecnologia tão rápido quanto possível por meio de alguma outra coisa. Racionalidade instrumental, utopias cibernéticas, neoliberalismo, fantasia de controle perfeito, positivismo, maximização de receita e assim por diante. Essas avaliações podem, em última instância, ser esclarecedoras e significativas em um nível de análise muito amplo. Todavia, se quisermos enfrentar os desafios dos métodos computacionais, temos de encontrar a tecnologia como tecnologia por pelo menos alguns instantes. Paradoxalmente, o foco unilateral no aspecto “digital” dos métodos computacionais e, em conjunto, na programação, como a *Via Regia* para a iluminação digital, implica um reducionismo que, novamente, serve para manter a tecnologia “pequena”. Não há dúvidas de que habilidades de programação e “iteração” (repetição de uma ou mais ações) são extremamente valiosas, bem como uma maneira de facilitar algumas das complexidades mais exigentes envolvidas em métodos computacionais. Contudo, esperamos ter demonstrado, por meio de nossos exemplos, que as ferramentas por nós utilizadas mobilizam amplas matrizes de conhecimento, as quais devemos, apenas de má vontade, condensar na suposta coerente categoria de “o digital”. O problema de “encaixapretar” (black boxing) não começa com a opacidade do código de computador, mas com o desejo de banir a tecnologia do “mundo de significação” (Simondon, 1958, p. 10).⁸ Atrás dos louváveis esforços

⁸ “A cultura está desequilibrada porque reconhece certos objetos, tais como o objeto estético, e concede-lhes o direito de residência no mundo do significado, enquanto relega outros objetos, e, em particular, objetos técnicos, ao mundo não estruturado pelas coisas que não têm significado, apenas um uso.” (Simondon, 1958, p. 10)

para aumentar os níveis de capacidade técnica reside o perigoso fantasma de que as epistemologias tecnológicas são basicamente “finas” e, uma vez que a habilidade de programação tenha sido adquirida, o domínio e o controle retornam.

Acreditamos, ao contrário, que qualquer ferramenta de software, não trivial, pressupõe camadas espessas de mediação que se conectam à computação, como tal, certamente, mas, na maioria dos casos, também demanda conceitos, métodos e estilos de raciocínio adaptados provenientes de vários outros domínios. Podemos criticar a padronização de pesquisas por meio de software o quanto quisermos, mas, para sermos francos, não há prática crítica da estatística sem conhecimento considerável de estatística, independentemente da questão referente a quais ferramentas são usadas. O problema da *Bildung* não pode ser reduzido à aquisição de um conjunto de habilidades. O que Simondon (1958) chama de “cultura técnica” (aculturação técnica) não deveria estar limitada ao treinamento técnico, mas precisa começar com o reconhecimento de que a tecnologia constitui uma maneira fundamental de se relacionar com o mundo e a diversidade humana, o que anda de mãos dadas com o pluralismo tecnológico (Simondon, 1958, p. 218).

Temos de ser capazes de pensar *com* e *na* tecnologia como um meio de expressar uma vontade e um meio de conhecimento. Isso não é apenas necessário para decidirmos quando aplicar técnicas e interpretar resultados que elas produzem; também é necessário decidir em que o computacional é supérfluo, enganador ou simplesmente bajulado por uma ideia de pesquisa “inovadora” de alguma agência de financiamento. Os métodos digitais vieram para ficar e para ir além de reflexos simplistas de entusiasmo e rejeição – precisamos de engajar-nos na prática crítica, isto é, de modo consciente, refletir a quantidade absurda de conhecimento que temos a atulhar as nossas ferramentas.

REFERÊNCIAS

- Agre, P. (1997a). Toward a Critical Technical Practice: Lessons Learned in Trying to Reform AI. *Social Science, Technical Systems, and Cooperative Work: Beyond the Great Divide*, ed. G.C. Bowker et al., 131-157. Mahwah, NJ: Lawrence Erlbaum Associates.
- . (1997b). *Computation and Human Experience*. Cambridge; New York: Cambridge University Press.
- Akrich, M. (1998). *Les utilisateurs, acteurs de l'innovation*. *Education Permanente* 134: 79-90.
- Amoore, L. & Volha Piotukh (2015). *Algorithmic Life. Calculative Devices in the Age of Big Data*. London: Routledge.
- Anscombe, F. John (1973). Graphs in Statistical Analysis. *The American Statistician* 27 (1): 17-23.

- Arthur, Paul L. and Katherine Bode, eds. 2014. *Advancing Digital Humanities. Research, Methods, Theories*. Houndsmills, New York: Palgrave Macmillan.
- Barocas, S. & Andrew D. Selbst. 2016. Big Data's Disparate Impact. *California Law Review* 104. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2477899.
- Bastian, M., Heymann, S. & Jacomy, M. (2009). Gephi: an open source software for exploring and manipulating networks. *ICWSM* 8: 361-362.
- Beer, D. (2013). *Popular Culture and New Media. The Politics of Circulation*. New York: Palgrave Macmillan.
- Berry, D. M. (2011a). "The computational turn: Thinking about the digital humanities." *Culture Machine* 12 (2).
- . (2011b). *Philosophy of Software. Code and Mediation in the Digital Age*. New York: Palgrave Macmillan.
- . (2012). *Iteracy: Reading, Writing and Running Code*. <http://stunlaw.blogspot.nl/2011/09/iteracy-reading-writing-and-running.html>.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioural Sciences*. 2nd edition. Hillsdale: Lawrence Erlbaum Associates.
- Cohen, P. (2010). *Digital keys for unlocking the humanities' riches*. New York Times, November 17, <http://www.nytimes.com/2010/11/17/arts/17digital.html>.
- Drucker, J. (2014). *Graphesis. Visual Forms of Knowledge Production*. Cambridge, MA: Harvard University Press.
- Freeman, L. (2004). *The Development of Social Network Analysis: A Study in the Sociology of Science*. Vancouver: Empirical Press.
- Gigerenzer, G., Krauss S. & Vitouch O. (2004). The Null Ritual. What You Always Wanted to Know About Significance Testing but Were Afraid to Ask. In *The Sage Handbook of Quantitative Methodology for the Social Sciences*, ed. David Kaplan, 391-408. Thousand Oaks, CA: Sage.
- Hacking, I. (1992). Statistical language, statistical truth and statistical reason: The self-authentication of a style of scientific reasoning. In *The social dimensions of science*, ed. E. McMullin, 130-157. Notre Dame: University of Notre Dame Press.
- Halpern, O. (2014). *Beautiful Data. A History of Vision and Reason since 1945*. Durham; London: Duke University Press.
- Jacomy, M., Venturini T., Heymann S. & Bastian M. (2014). ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software. *PLoS one* 9 (6): e98679.
- Kassung, C. (2001). *EntropieGeschichten*. Robert Musils "Der Mann ohne Eigenschaften" im *Diskurs der modernen Physik*. Munich/Paderborn: Fink. Kuhn, Thomas S. 1962. *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.
- Manovich, L. (2012). How to Compare One Million Images. In *Understanding Digital Humanities*, ed. David M. Berry, 248-278. New York: Palgrave Macmillan.
- Ramsay, S. (2011). Who's In and Who's Out. *Stephen Ramsay* (blog), 8 January. <http://stephenramsay.us/text/2011/01/08/whos-in-and-whos-out/>.
- Ratto, M., Wylie, S. A. & Jalbert, K. (2014). Introduction to the Special Forum on Critical Making as Research Program. *The Information Society* 30 (2): 85-95.
- Rieder, B. (2012). What is in PageRank? A Historical and Conceptual Investigation of a Recursive Status Index. *Computational Culture* 2, http://computationalculture.net/article/what_is_in_pagerank.
- Rieder, B. & R.hle, T. (2012). Digital Methods: Five Challenges. In *Understanding Digital Humanities*, ed. David M. Berry, 67-85. New York: Palgrave Macmillan.
- Rieder, B., Abdulla, R., Poell, T., Woltering, R. & Zack, L. (2015). Data critique and analytical opportunities for very large Facebook Pages: Lessons learned from exploring 'We are all Khaled Said'. *Big Data & Society* 2 (2): 2053951715614980.

- Rodgers, J. L. & Nicewander, W. A. (1988). Thirteen Ways to Look at the Correlation Coefficient. *The American Statistician* 42 (1): 59-66.
- Schön, D. A. (1983). *The Reflective Practitioner. How Professionals Think in Action*. New York: Basic Books.
- Simondon, G. (1958). *Du mode d'existence des objets techniques*. Paris: Aubier.
- Snow, C.P. (1959). *The Two Cultures*. Cambridge, MA: Cambridge University Press.
- Svensson, P. & Goldberg, D. T., eds. (2015). *Between Humanities and the Digital*. Cambridge, Mass.: MIT Press.
- Tukey, J. W. (1962). The Future of Data Analysis. *The Annals of Mathematical Statistics* 33: 11-67.
- Uprichard, E., Burrows, R. & Byrne D. (2008). SPSS as an 'inscription device': from causality to description? *The Sociological Review* 56 (4): 606-622.
- Wang, H. (1963). Toward Mechanical Mathematics. In *The modelling of Mind*, ed. K.M. Sayre & F.J. Crosson. South Bend, IN: Notre Dame University Press.
- Winkler, H. (2015). Prozessieren. Die dritte, vernachlässigte Medienfunktion. Munich/Paderborn: Fink.

*Tradução por Tiago Salgado.

O TECIDO SOCIAL: RASTROS DIGITAIS E MÉTODOS QUALI-QUANTITATIVOS

TOMMASO VENTURINI
BRUNO LATOUR

De acordo com Tito Lívio, em 493 a.C., os plebeus romanos se refugiaram no Monte Sagrado para protestar contra o Senado. Os patrícios enviaram Menenius Agrippa para negociar com os rebeldes. Agrippa os convenceu a retornar à cidade de Roma ao discursar longamente, contando-lhes uma fábula, reescrita depois por William Shakespeare (Coriolano, Ato 1, Cena 1):

Contra o estômago os membros se insurgiram certo dia, acusando-o de no meio do corpo colocar-se, preguiçoso sempre e inativo, e, como sorvedouro, absorver, insaciável, a comida, sem nunca contribuir com sua parte para o comum trabalho, enquanto os outros órgãos viam, andavam, refletiam, sentiam e falavam, contribuindo cada um, assim, com sua parte, para proverem às comuns necessidades e apetites do corpo. Respondeu-lhes o estômago...

“É certo, meus amigos incorporados”, disse, “que eu recebo, antes de outro qualquer, todo o alimento de que viveis, e é justo que assim seja, por ser eu o depósito e celeiro de todo o corpo. Mas se estais lembrados, pelos canais de vosso sangue tudo de novo mando à corte, ao coração, à alta sede do cérebro, e assim, pelos sinuosos passos da oficina humana, os nervos mais potentes e as menores arteríolas de mim recebem tudo de quanto necessitam para a vida. E muito embora todos vós, a um tempo, meus bons amigos...” isso disse o estômago, “observai bem”.

“E muito embora todos vós, a um tempo, não vejais o que eu dou em separado para cada um, muito fácil é provar-vos por um cálculo certo e rigoroso, que recebeis de mim toda a farinha, sobrando-me de tudo só o farelo”¹

Quando Menenius Agrippa recitou sua fábula, as artes medicinais ainda estavam bem longe de desenvolverem a técnica da cintilografia. Tal técnica de diagnóstico possibilita seguir os movimentos e a acumulação de isótopos radioativos administrados a um paciente. Com o auxílio da cintilografia, os membros do corpo, personificados na fábula, puderam saber não apenas como o estômago distribuiu os nutrientes, por quais canais e em quais proporções, mas, também, como cada membro coopera com os outros, no corpo humano. Se eles soubessem disso tudo, a história de Agrippa poderia ter terminado bem diferente.

Frequentemente, é dito que as tecnologias digitais estão revolucionando as ciências sociais, tanto quanto elas revolucionaram as ciências naturais, há algumas décadas. Este não é o caso. A situação nas ciências sociais é muito mais parecida com a das ciências naturais no século xv, logo após a introdução da prensa de tipos móveis. Em “A Imprensa como agente de mudança” (*The Printing Press as an Agent of Change*), Elizabeth Eisenstein (1979) minuciosamente descreve como a introdução da imprensa desencadeou a Revolução Científica do século xvi. No entanto, como Eisenstein deixa bem claro, a revolução não aconteceu da noite para o dia. Várias décadas depois da invenção de Gutenberg, os mesmos erros que os naturalistas costumavam reproduzir à mão ainda eram impressos. Foi apenas depois de a imprensa adquirir um certo nível de maturidade e circulação que ela passou a ter um impacto significativo nas ciências naturais.

As ciências sociais estão em uma posição similar hoje. Ao invés de utilizar os dados digitais para revitalizar suas práticas, tais ciências ainda tentam fazê-los se passar por novos terrenos, por meio de antigos métodos. Cibercultura, comunidades virtuais, identidades online, comunicação mediada por computador: todas essas noções foram elaboradas para colocar em quarentena a novidade da mídia digital. Todavia, a velocidade com a qual as tecnologias digitais infiltram na modernidade torna essa resistência cada vez mais insustentável. Não é mais possível circunscrever as interações digitais a um setor específico, uma vez que

¹ Versão original, em latim, em *Ad Urbe Condita*, II, 32: *Tempore quo in homine non ut nunc omnia in unum consentiant, sed singulis membris suum cuique consilium, suus sermo fuerit, indignatas reliquas partes sua cura, suo labore ac ministerio uentri omnia quaeri, uentrem in medio quietum nihil aliud quam datis uoluptatibus frui; conspirasse inde ne manus ad os cibum ferrent, nec os acciperet datum, nec dentes quae acciperent conficerent. Hac ira, dum uentrem fame domare uellent, ipsa uma membra totumque corpus ad extremam tabem uenisse. Inde apparuisse uentris quoque haud segne ministerium esse, nec magis ali quam alere eum, reddentem in omnes corporis partes hunc quo uiuimus uigemusque, diuisum pariter in uenas maturum confecto cibo sanguinem.*

os fios dessas interações estão, agora, ubiquamente entrelaçados no tecido da existência coletiva. As mídias digitais oferecem muito mais do que apenas um outro campo para a aplicação de métodos existentes: elas possibilitam a reestruturação do estudo da existência social (Rogers, 2004).

Até agora, o acesso aos fenômenos coletivos tem sido sempre incompleto e dispendioso. Em comparação aos seus colegas das ciências naturais, os cientistas sociais sempre foram insuficientemente equipados. Enquanto os físicos podiam seguir bilhões de átomos em seus aceleradores, e os biólogos podiam fazer crescer milhões de micróbios em seus microscópios; os cientistas sociais não podiam manter, simultaneamente, a amplitude e a profundidade em suas observações. Os métodos utilizados davam-lhes uma visão bipartida da existência social, visto que podiam focar em interações específicas ou tocar a superfície de estruturas globais.

Entre outras coisas, isso explica porque, em ciências sociais, a noção de “emergência” adquiriu um significado distinto das ciências naturais. Em física, biologia e medicina, tal concepção é usada para descrever fenômenos de auto-organização, devidos à coordenação espontânea de muitos agentes. Nas ciências sociais, fenômenos emergenciais se tornam estruturas, quadros explicativos que recebem e influenciam interações. Essa visão estruturalista e dualista advém, em grande parte, do fato de as ciências sociais nunca terem métodos para reconectar o micro e o macro, e mostrar como fenômenos globais são construídos pelo conjunto de interações locais. As tecnologias digitais prometem revolucionar essa situação, ao prover às ciências sociais a possibilidade de seguir cada fio da interação e mostrar como a vida social é tecida coletivamente.

O PROBLEMA DA VIDA COLETIVA

O problema da vida social é que toda situação que implica em mais de duas pessoas pode rapidamente se tornar inacreditavelmente complicada. Tome a complexidade de um ser humano e a multiplique pelo número de pessoas envolvidas na interação e, então, pelo número de interações necessárias para organizar o mais simples dos fenômenos sociais. Adicione o fato de as interações poderem afetar e serem afetadas por qualquer tipo de agente não humano (tecnologias, elementos da natureza, obras de arte etc.), e o fato de esses agentes poderem carregar as consequências de outras interações que se estendem no tempo e no espaço. Se pudermos fazer alguma coisa com essa complexidade, isso se deve ao amplo número de instrumentos simplificadores com os quais nossas sociedades podem contar: instituições, normas, padrões, procedimentos, entre outros. Ape-

sar do que podemos pensar, as chamadas “estruturas sociais emergentes” são, de fato, mais simples do que as interações das quais são feitas.

Emergência é uma noção muito útil, uma ideia que beneficia tanto os atores sociais quanto os cientistas sociais: ela possibilita aos atores assumirem a estabilidade dos fenômenos coletivos, e aos pesquisadores presumirem que esses fenômenos fazem sentido (Johnson, 2001). Contudo, a noção de emergência oculta sentidos completamente diferentes nesses dois casos. Para os atores sociais, não há outro modo de acessar propriedades emergentes senão as construindo. Todos aqueles que tentaram formar uma banda de rock, organizar um torneio de xadrez ou iniciar uma roda de tricô sabiam bem disso: os fenômenos sociais não existem por eles mesmos. Os fenômenos coletivos emergem por meio do trabalho dos participantes, e se estendem tanto quanto os atores os levam adiante, e duram tanto quanto eles o aguentam.

Sem dúvida, os fatos sociais podem ser materializados: decisões podem ser transcritas, fronteiras reforçadas por barreiras, relacionamentos selados com símbolos. As estruturas sociais podem ser estabilizadas por infraestruturas materiais (Latour, 2005), mas é apenas pelo trabalho coordenado de atores que os fenômenos coletivos podem emergir e durar (Weisman, 2007). Esse trabalho, todavia, raramente é reconhecido pelas ciências sociais, que, desde a sua origem, nunca tiveram os recursos necessários para explorar a construção de estruturas sociais. Tais ciências podem seguir poucas interações específicas, mas desmembrar os fatos sociais em milhares de interações que os constituem sempre esteve além do seu alcance. Assim, elas preferiam evitar a questão.

Durante o século XIX, uma sociologia recente encontrou tudo o que precisava na estatística. Ao combinar duas tradições matemáticas distintas, referentes ao cálculo de probabilidades e a estimativa de erros, a estatística forneceu às ciências sociais um caminho mais curto e conveniente (Desrosières, 1998). Ao invés de traçar as maneiras pelas quais estruturas emergentes eram construídas, pela simplificação de milhares de trocas locais, os pesquisadores poderiam ganhar acesso aos fenômenos sociais por meio de um “salto matemático”. Onde quer que os atores sociais se *reunissem*, os pesquisadores poderiam *agregar*. Por exemplo, ao invés de estudar como organizações criminosas selecionavam e socializavam com seus recrutas, os pesquisadores poderiam calcular o perfil da “média criminal”. Ao invés de estudar como coligações criminosas expandem seus investimentos por meio da coordenação de uma multiplicidade de indivíduos criminosos, os pesquisadores poderiam estimar a *incidência média* de cada crime. O ganho em eficácia era enorme. As ciências sociais finalmente poderiam não apenas representar as estruturas sociais, mas também contribuir para a sua solidificação.

Certamente que a adoção da estatística veio com um custo. Ao contornar o trabalho de construção coletiva via atalho estatístico, os pesquisadores introduziram uma distinção fictícia entre microinterações e macroestruturas. Apesar de essas distinções nunca serem alguma coisa a mais do que uma ilusão de ótica que reflete o salto da agregação estatística, os cientistas sociais sempre as consideraram muito convenientes. Por meio da distinção micro/macro, os pesquisadores que se interessavam pelas interações locais poderiam observá-las, sem ter de interferir no emaranhado de influências que cercavam cada evento social. Ademais, aqueles que queriam observar o quadro completo da vida social, poderiam fazer isso sem se preocupar com cada pincelada. Ao separar micro e macro, os cientistas sociais validavam, simultaneamente, métodos qualitativos e quantitativos.

Entretanto, a abordagem dualista baseada na distinção micro/macro tem uma desvantagem significativa: ela apenas funciona *após* os atores terem finalizado seu trabalho de simplificação e *até* eles conseguirem mantê-lo. Não é possível cruzar amostragens se não houver limites claros entre classes, grupos ou gêneros. Não é possível confiar em médias ou curvas regulares se os atores não estiverem em conformidade com as normas ou com a normalidade. Não se pode estimar desvios se comportamentos desviantes estão dissuadidos ou sancionados. As análises estatísticas podem aproximar fatos coletivos apenas na medida em que foram definidos, normalizados e estabilizados pelos atores sociais. A abordagem dualista pode muito bem descrever fenômenos que já *emergiram*, mas é inútil para descrever os fenômenos *emergentes*.

APRENDER COM AS CONTROVÉRSIAS

A fraqueza da abordagem dualista se torna especificamente clara se, ao invés de nos concentrarmos em áreas consensuais, voltarmos para o estudo de controvérsias, de modo a observar a vida social, onde o acordo não é mais (ou não é ainda) alcançado. No estudo de disputas, não podemos depender nem de métodos quantitativos (porque as posições locais são tão variadas que não há nada a ser extraído reunindo-os em uma média geral) nem de métodos qualitativos (porque mesmo a controvérsia mais específica é sempre parte de uma rede global de influências). Onde as disputas irrompem, onde os atores não podem chegar ao acordo, onde os laços sociais são desfeitos, ou simplesmente onde a mudança se desenrola tão rapidamente que velhas instituições se dissolvem antes que outras sejam criadas, os métodos quantitativos e qualitativos chegam ao limite.

Esta é a razão pela qual as ciências sociais sempre estiveram desconfortáveis com as controvérsias: os seus métodos demandam que elas os reduzam a fenômenos residuais, ou melhor, a estágios de transição entre períodos de estabilidade. Isso é sobretudo evidente na sociologia de ciência e tecnologia. Durante anos, entre os acadêmicos, os argumentos foram considerados eventos insignificantes na estrada para o avanço científico. Mesmo autores como Thomas Kuhn (1962), que levou as controvérsias científicas a sério, terminaram por defini-las como fases de passagem entre dois períodos de “ciência normal”.

Muito do interesse pelos Estudos em Ciência e Tecnologia (STS), tal como desenvolvidos pelas escolas de Edimburgo (BLOOR, 1976) e de Bath (Collins & Pinch, 1993), reside na demonstração de que, longe de serem fenômenos residuais ou de transição, as controvérsias são o verdadeiro mecanismo da ciência. Em ciência e tecnologia (Bijker & Law, 1992), os desacordos e os conflitos são a norma, e o consenso só pode ser alcançado por meio de esforços árduos de coordenação e estabilização (Callon, Lascoumes & Barthe, 2009). Os paradigmas científicos não são sistemas coerentes que abrem o caminho para os acadêmicos. Ao invés disso, são produtos frágeis do trabalho deles. Os paradigmas requerem negociações intermináveis para serem mantidos, e a sua existência é constantemente ameaçada pelo risco de tais negociações falharem (Latour, 1988).

Ao trabalhar com estudantes no Instituto de Estudos Políticos em Paris, em um curso intitulado Cartografia de Controvérsias,² viemos a entender que, longe de estar limitada às ciências, essa situação é comum a todos os fenômenos coletivos. As controvérsias nos ensinaram que nada em nossas vidas sociais dura sem esforços. A vida em comunidade lembra muito mais um mar agitado que um longo e tranquilo rio, com milhares de correntes, que nenhum contrato social poderia definitivamente acalmar. Assim, instituições sociais não são estruturas que se impõem aos indivíduos, mas uma espécie de “Lei de Guerra” sempre sujeita à negociação. Certamente que essas instituições existem e têm efeitos, mas tais efeitos são o resultado de interações, em vez de serem a premissa delas.

A combinação de estabilidade e mudança, que caracteriza as controvérsias, não pode ser desemaranhada ao se utilizar de uma abordagem dualista (Venturini, 2010a). Os métodos qualitativos só podem ser endereçados a poucos elementos, que criam consenso entre os atores (a parte menos interessante da disputa), posto que nos impedem de encerrar disputas sociais em uma rede de oposições múltiplas que as caracterizam. Os métodos qualitativos inibem qualquer entendimento de como o ponto de equilíbrio pode ser temporariamente negociado

² Disponível em: <<http://controverses.sciences-po.fr>>.

enquanto a controvérsia ainda está aberta, visto que nos impedem de mostrar como toda querela está atada a uma extensa rede de oposição. Estudar controvérsias requer um novo conjunto de métodos “quali-quantitativos”, que nos permitam rastrear fenômenos sociais ao longo dos processos de construção, desconstrução e reconstrução que os constituem.

A PROMESSA DE MÉTODOS COMPUTADORIZADOS

Os métodos quali-quantitativos não são meramente a justaposição de análises estatísticas com observação etnográfica. Apagar a fronteira micro/macro implica estar preparado para coletar e processar uma quantidade de dados sem precedentes. Se interessar pela construção de fenômenos sociais requer rastrear cada um dos atores envolvidos e cada uma das interações entre eles. Este era um objetivo impossível há apenas algumas décadas, mas que se torna cada vez mais realista à medida que as tecnologias digitais se espalham.

O interesse pelas mídias digitais está no fato de cada interação que passa por elas deixar rastros facilmente gravados, massivamente arquivados e recuperados a baixo custo. A cada dia, pesquisadores descobrem novas fontes de dados digitais: arquivos públicos e privados são sugados para dentro da memória de computadores; transações econômicas são feitas online; redes sociais online se consolidam na web. A mediação digital se espalha como um grande rolo de papel carbono, de modo a oferecer às ciências sociais mais dados do que elas jamais imaginaram.

Graças à rastreabilidade digital, os pesquisadores não precisam mais escolher entre precisão e alcance em suas observações: agora é possível seguir uma multiplicidade de interações e, simultaneamente, distinguir a contribuição específica que cada uma delas tem para a construção de fenômenos sociais. Concebidas em uma era de escassez, as ciências sociais entram em uma época de abundância. Em virtude da riqueza desses novos dados, nada justifica manter as velhas distinções. Providas de uma quantidade de dados comparável às ciências naturais, as ciências sociais finalmente podem corrigir sua vista cansada e, simultaneamente, manter o foco e o alcance de suas observações.

O potencial de tal abordagem quali-quantitativa foi iluminado pelo projeto MACOSPOL (*MApping COntroversies on Science for POLitics* – Mapeamento de controvérsias em ciência para a política). Fundado pela União Europeia, esse projeto coordenou oito laboratórios europeus no esforço de inventar e testar ferramentas digitais disponíveis para o mapeamento de controvérsias (Venturini, 2010b).

Embora a quantidade de recursos identificados fosse animadora, os resultados do projeto também revelaram a relativa imaturidade dos métodos digitais. Hoje, mais e mais ferramentas estão disponíveis para extrair e analisar dados, e muitos pesquisadores começaram a utilizá-las (Lazer et al., 2009). Ainda assim, o êxito dessas iniciativas permanece limitado por sua natureza intermitente. Em muitos casos, os acadêmicos se restringiram a digitalizar uma parte ou outra de suas pesquisas sem reconsiderar todo o processo de investigação. Os resultados de tal abordagem certamente são interessantes e têm a vantagem de serem facilmente comparáveis com aqueles dos métodos clássicos. No entanto, uma verdadeira transição para os métodos quali-quantitativos não será possível até que a pesquisa em ciência social seja completamente digitalizada. Para que os novos métodos alcancem seu potencial de inovação, é necessário que cada passo na pesquisa seja repensado de maneira coerente:

1. A identificação dos conjuntos de dados (*data set*) deve privilegiar arquivos digitais, a web, a mídia online e as redes sociais online, a documentação digital e a literatura, e de modo geral toda sorte de rastros digitais;
2. A extração de dados deve estar baseada em navegação assistida ou em diferentes técnicas de coleta de dados (rastreamento, consulta, raspagem, análise etc.);
3. Os dados de diversas fontes devem estar integrados em uma base de dados relacional comum (*data mashing*);
4. A análise e a modelagem de dados devem se apoiar na web, no digital, e quando possível em ferramentas com código aberto;
5. Os resultados de pesquisa devem ser publicados na web, preferencialmente em arquivos abertos e em um formato padrão para facilitar a reutilização deles.

É fácil ver que o processo que descrevemos é, de fato, uma espiral. Disponíveis na web, os resultados de uma investigação digital podem se tornar os dados para uma nova análise. Ainda mais importante: os dados digitais, bem como as ferramentas para extrair e analisá-los, podem ser publicados juntamente com os resultados, tornando acessível aos públicos não apenas as conclusões do projeto de pesquisa, mas também todos os passos para se chegar a eles. Este é o verdadeiro apelo dos métodos digitais: em vez de apenas descrever os procedimentos experimentais, é possível, agora, conceder acesso direto a eles. Em uma sequência de pesquisas digitais, nenhum passo está bloqueado, nenhuma transformação é irreversível, nenhum resultado está dado. Cada elemento na sequência é dis-

ponibilizado aos públicos e está aberto ao exame minucioso. Em um projeto de pesquisa que se dedica a investigar o digital, navegar pelos dados e navegar pela cadeia de processamento de dados é, no fim das contas, a mesma coisa.

O OLIGÓPTICO QUALI-QUANTITATIVO

A promessa dos métodos digitais está na possibilidade de abertura de pesquisas em ciências sociais. Contudo, para manter essa promessa, todas as pesquisas devem ser nativamente digitais. Enquanto estiverem confinados a uma parte do processo de pesquisa, os métodos digitais não irão implantar o seu potencial inovador e correrão o risco de produzir todo tipo de efeitos incompatíveis.

A questão da representatividade, por exemplo, ilustra perfeitamente a dificuldade de conjugar métodos digitais e tradicionais. Uma crítica que constantemente tem sido endereçada à análise de rastros digitais é que esses rastros dizem respeito a uma amostra não representativa da sociedade. Do ponto de vista das ciências sociais tradicionais, isso certamente é verdade. É bem sabido que a alfabetização digital não é uniformemente difundida na sociedade (homens, jovens e aqueles com alto grau de escolaridade geralmente são representados mais enfaticamente em amostras digitais). Ainda assim, essa desproporção é um problema apenas enquanto insistirmos em tratar os dados digitais como se eles fossem dados de pesquisa (*survey*). A vantagem dos novos métodos é a possibilidade de rastrear o conjunto de fenômenos coletivos em vez de obtê-los por meio da agregação estatística. A questão da representatividade é, então, colocada de um jeito completamente diferente.

Os dados digitais são representativos e interessantes apenas se toda a sequência de processamento deles (identificação, extração, integração, análise e publicação) permanecer próxima ao trabalho dos atores sociais. De fato, não dizemos que os métodos quali-quantitativos irão nos possibilitar suavizar toda a complexidade da vida coletiva. Pelo contrário, a vantagem desses métodos é ser flexíveis o bastante para seguir *algum* fenômeno social ao longo de cada um de seus desdobramentos.

Nenhum método oferece uma visão panóptica da existência coletiva, e os métodos quali-quantitativos não são exceção. Os métodos digitais só podem oferecer uma visão oligóptica da sociedade (Latour & Hermant, 1988), exatamente como os métodos tradicionais. Todavia, pela primeira vez na história das ciências sociais, essa visão será ao menos contínua, desde a mais ínfima microinteração até a maior macroestrutura.

Por essa razão, abrimos este capítulo com a história de Meneus Agrippa e a tecnologia da cintilografia. A vantagem da cintilografia não é possibilitar observar qualquer fenômeno biológico, mas seguir algum desses fenômenos de maneira contínua, sem saltos ou falhas. A existência social não está dividida em dois níveis, como nos fizeram acreditar os métodos tradicionais. Microinterações e macroestruturas são apenas dois diferentes modos de olhar para o mesmo quadro coletivo, assim como a urdidura e a trama do tecido social. Na unidade gerada pela multiplicação de diferenças, na estabilidade produzida pela acumulação de mutações, na harmonia advinda das controvérsias, no equilíbrio dependente de milhares de fraturas está a maravilha da existência comunal. Os métodos qualitativos e quantitativos, por muito tempo, nos esconderam desse espetáculo. Os métodos digitais abrirão os nossos olhos.

REFERÊNCIAS

- Bijker, W. E., and Law, J. (eds.) (1992), *Shaping Technology, Building Society: Studies in Socio-technical Change*. Cambridge, Mass.: MIT Press.
- Bloor, D. (1976), *Knowledge and Social Imagery*. Chicago: University of Chicago Press.
- Callon, M., Lascoumes, P., and Barthe, Y. (2009), *Acting in an Uncertain World: An Essay on Technical Democracy*. Cambridge, Mass.: MIT Press.
- Collins, H., and Pinch, T. (1993), *The Golem: What Everyone Should Know about Science*. Cambridge, U.K.: Cambridge University Press.
- Desrosières, A. (1998), *The Politics of Large Numbers: A History of Statistical Reasoning*. Cambridge, Mass.: Harvard University Press.
- Eisenstein, E. (1979), *The Printing Press as an Agent of Change: Communications and Cultural Transformations in Early Modern Europe*. Cambridge, U.K.: Cambridge University Press.
- Johnson, S. (2001), *Emergence: The Connected Lives of Ants, Brains, Cities, and Software*. New York: Touchstone.
- Kuhn, T. (1962), *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.
- Latour, B. (1988), *Science in Action: How To Follow Scientists and Engineers through Society*. Boston: Harvard University Press.
- Latour, B., and Hermant, E. (1988), *Paris ville invisible*. Paris: La Découverte.
- Latour, B. (2005), *Reassembling the Social: An Introduction to Actor-Network Theory*. Oxford: Oxford University Press.
- Lazer, D. et al. (2009), “Computational Social Science,” in *Science*, Vol. 323: 721-723.
- Rogers, R. (2004), *Information Politics on the Web*. Cambridge, Mass.: MIT Press.
- Venturini, T. (2010a), “Diving in Magma: How To Explore Controversies with Actor-Network Theory,” in *Public Understanding of Science*, Vol. 19, No. 3, 258-273.
- Venturini, T. (2010b), “Building on Faults: How To Represent Controversies with Digital Methods,” in *Public Understanding of Science* (under review).
- Weisman, A. (2007), *The World Without Us*. New York: St. Martin's Press.

*Tradução por Tiago Salgado.



DOIS

**CONCEITOS,
ABORDAGENS
E PRÁTICAS**

A PLATAFORMIZAÇÃO DA WEB

ANNE HELMOND

Em 15 de agosto de 2006, o Facebook introduziu sua plataforma para desenvolvedores, ao conceder, a terceiros, acesso aos perfis de usuários e amigos, às fotos e aos eventos, a fim de expandir a “experiência Facebook” para aplicativos externos (Fatterman, 2006) – desse modo, tornou o Facebook um ambiente para desenvolvedores. Um ano depois, na primeira Conferência de Desenvolvimento F8, o Facebook lançou a plataforma Facebook, de modo a marcar oficialmente o avanço do Facebook como plataforma. A plataforma Facebook fornece aos desenvolvedores um conjunto de ferramentas para envio e recuperação de dados do Facebook e de fontes externas para o próprio Facebook, bem como uma profunda integração com o grafo social do Facebook (*Social Graph*) – um mapeamento das conexões entre pessoas e objetos, para a construção de aplicativos (Geminder, 2007; Hicks, 2010).

Neste capítulo eu investigo o desenvolvimento do Facebook como uma plataforma ao situá-lo no contexto de transformação dos sites de redes sociais em *plataformas* de mídias sociais. Situamos essa “plataformização”, ou o advento da plataforma, como o modelo dominante de infraestrutura e economia da web, bem como as consequências disso, no contexto histórico desse processo. A plataforma implica a extensão das plataformas de mídias sociais ao restante da web, bem como o movimento de tais plataformas para tornarem os dados da web, que lhes são externos, prontos para configurarem plataformas. A arquitetura tecnológica específica e a distinção ontológica das plataformas serão examinadas ao levarmos em conta um aspecto da “especificidade do meio” (*medium-specificity*) que elas apresentam (Rogers, 2013) e sua programabilidade. Ao fazer isso, seguimos à convocatória de Langois e outros (2009) por uma “perspectiva baseada em plataforma”, a qual, segundo Fenwick Mckelvey (2011), deveria criticamente

investigar a programabilidade das plataformas. A análise da descentralização das características da plataforma na web e a recentralização dos dados “prontos para plataforma” é uma maneira de examinarmos as consequências da programabilidade das plataformas de mídias sociais na web.

O novo modelo arquitetônico da plataforma explicitamente torna os sites na web acessíveis ao possibilitar a sua programabilidade com as interfaces técnicas – ou Interface de Programação de Aplicativos (a sigla API, em inglês). Para compreender esse acesso programático, recorro à noção de “fluxos de dados” (*data pours*), proposta por Alan Liu (2004), para conceituarmos as plataformas como sistemas de fluxos de dados, que configuram canais de dados, os quais possibilitam fluxos de dados com terceiros. Esses fluxos de dados não apenas configuram canais para que os dados fluam entre plataformas de mídias sociais e terceiros, mas também funcionam como canais de dados que preparam os dados externos para a configuração de plataformas.

PERSPECTIVA TÉCNICO-MATERIAL SOBRE PLATAFORMAS DE MÍDIAS SOCIAIS

O termo “plataforma” se tornou o conceito dominante tanto para o posicionamento de empresas de mídias sociais no mercado quanto para o modo como elas se dirigem aos usuários. Essa noção tem sido amplamente aceita pelos consumidores e pela imprensa (Gillespie, 2010). Em novos estudos sobre mídias, o conceito de plataforma ganhou destaque ao chamar a atenção para o “trabalho discursivo” que tais estudos empreendem (Gillespie, 2010, p. 348) e para o papel do software, que fortalece as mídias sociais na formação da participação e da socialidade (Bucher, 2012a; Hands, 2013; Langlois et al., 2009; Van Dijck, 2013).

Em uma das principais discussões sobre plataformas, Tartelon Gillespie (2010) apresenta uma variedade de plataformas com foco nas diferentes conotações do termo. No sentido computacional, o autor define plataforma como uma infraestrutura para a construção de aplicativos. Todavia, Gillespie (2010) afirma que as empresas de web 2.0 introduziram um sentido mais amplo da noção de “plataforma”, que ultrapassa o seu sentido computacional:

Esse uso mais conceitual de “plataforma” se baseia em todas as conotações do termo: computacional – algo no qual se pode construir e a partir do qual se pode inovar; política – um lugar a partir do qual se pode falar e ser

ouvido; figurativa — a oportunidade é tanto uma promessa abstrata quanto prática; e arquitetônica — em que o YouTube é elaborado como um meio aberto e igualitário que facilita a expressão, e não um *gatekeeper* elitista com restrições técnicas e normativas (Gillespie, 2010, p. 352).

Gillespie (2010) argumenta que este uso conceitual possibilita que as plataformas tragam, junto delas, vários atores. O sentido computacional de plataforma se dirige para os desenvolvedores, enquanto as outras conotações se endereçam aos atores, tais como usuários, anunciantes e clientes (Gillespie, 2010). O autor descreve o que em termos econômicos é chamado “modelo de negócios multilateral” (*multi-sided market*), no qual uma plataforma possibilita interações entre duas ou mais partes distintas (Gillespie, 2010, p. 990). Um exemplo de plataforma multilateral é o Facebook, que conecta usuários, anunciantes, desenvolvedores, e experimenta efeitos de rede em que o valor aumenta para todas as partes à medida que mais pessoas o utilizam (Hagiu, 2014).

Na literatura sobre economia e administração, Annabelle Gawer argumenta que as plataformas têm sido teorizadas a partir de duas perspectivas distintas: “a teoria econômica conceitua as plataformas como mercados” (Rochet; Tirole, 2003), ao passo que “o design de engenharia as teoriza como ‘arquiteturas tecnológicas modulares’ (Baldwin; Woodard, 2009)” (Gawer, 2014, p. 1240). Bernhard Rieder e Guillaume Sire (2014, p. 197) fazem uma convocatória importante para juntarmos essas duas perspectivas: estudar as plataformas como multilaterais, eles argumentam, “pode ampliar as análises de configurações concretas de poder e identificar pontos de controle, dinâmicas estruturais e recursos importantes de argumentação” (Rieder; Sire, 2013, p. 208). Ao seguir essas perspectivas tecno-econômicas sobre plataformas, examinamos, neste capítulo, como a arquitetura técnica e modular de plataformas de mídias sociais se conecta com o modelo de negócios delas.

Em seu trabalho sobre as plataformas, Gillespie (2010) enfatiza os aspectos econômicos e de participação das plataformas em detrimento da dimensão computacional delas ao dizer que “plataformas são ‘plataformas’ não necessariamente porque possibilitam que códigos sejam escritos ou rodados, mas porque elas proveem uma oportunidade para comunicar, interagir ou vender” (Gillespie, 2010, p. 351). Outros autores, tais como Ian Bogost e Nick Montfort (2009), sugerem um foco mais estreito em relação às plataformas ao trazerem o aspecto computacional para primeiro plano. A seguir, o interesse deste trabalho está em desenvolver um pouco mais essa dimensão computacional das plataformas para analisar o “trabalho que as plataformas fazem”, não em um sentido retórico

(cf. Gillespie, 2010), mas a partir de uma perspectiva técnico-material. Bogost e Montford (2009) refutam a ideia de que “nos dias de hoje, tudo é uma plataforma”, e nos alertam para efetivamente tratarmos as plataformas como infraestruturas computacionais. Estes dois autores enxergam as plataformas, em seu sentido computacional, como uma camada das novas mídias que é pouco estudada (Bogos & Montfort, 2009). Para solucionar esse ponto cego, ambos introduzem o termo “estudo de plataforma” (*platform studies*), ou seja, uma chamada para “o rigor técnico e a investigação aprofundada de como as tecnologias computacionais funcionam” para analisar “a conexão entre especificidades técnicas e a cultura” (Bogost & Montfort, 2009, p. vii).

Essas conexões foram exploradas por vários autores engajados em uma perspectiva política sobre as plataformas para analisarem “as *affordances* tecnológicas das plataformas em relação aos seus interesses políticos, econômicos e sociais”, considerando-as como locais importantes em que as “políticas de plataforma” se desenrolam (Hands, 2013; Langlois & Elmer, 2013).¹ As abordagens a respeito de políticas de plataformas incluem:

- a) o questionamento crítico sobre o conceito de plataforma (Gillespie, 2010; Mckelvey, 2011),
- b) a análise das “lógicas tecnoculturais” das plataformas (Gerlitz; Helmond, 2013; Langlois; Elmer; Mckelvey; Devereaux, 2009; Langlois et al., 2009),
- c) a análise do papel da arquitetura de plataforma na configuração da socialidade em rede (Bucher, 2012a; Van Dijck, 2013),
- d) e a análise das políticas de APIs (Bucher, 2013) e dos dados de plataforma (Puschmann & Burgess, 2013) (cf. Renzi, 2011).

Neste trabalho, tenho interesse pelo modo como as plataformas reformatam a web conforme a lógica das mídias sociais. Minha abordagem se baseia naquilo que Langlois e outros (2009) nomeiam “desagregação”, e examina criticamente as plataformas de mídias sociais ao desmontá-las e investigar seus componentes específicos (Langlois & Mckelvey et al., 2009). Essa contribuição para os estudos de plataforma e de mídias sociais se encontra numa detalhada perspectiva técnico-material sobre o desenvolvimento e a emergência do que compreendemos

¹ “Políticas de Plataforma” é o título de uma conferência realizada na *Anglia Ruskin University*, em Cambridge, no Reino Unido, nos dias 12 e 13 de maio de 2011. A conferência foi organizada por Josh Hands e Jussi Parikka, e reuniu vários estudiosos de políticas de plataforma. Após a conferência, a *Culture Machine* publicou uma edição especial com o título “Políticas de Plataformas” (2013).

hoje como plataformas de mídias sociais. Este argumento será desenvolvido adiante, ao focarmos no Facebook, uma das maiores e mais acessada plataforma de mídias sociais.²

FACEBOOK: SITE DE REDE SOCIAL OU PLATAFORMA?

Antes de o conceito de plataforma ganhar relevância, plataformas de mídias sociais, como o Facebook, eram muitas vezes conceituadas como sites de redes sociais, definidos por boyd e Ellison (2008, p. 211) como serviços web nos quais usuários podem criar um perfil, construir e disponibilizar uma lista de conexões com outros usuários na rede. Contudo, o Facebook sempre, cuidadosamente, absteve-se de chamar a si mesmo de rede social (Arrington, 2008; Locke, 2007). Em vez disso, ao longo do tempo, um dos fundadores do Facebook, Mark Zuckerberg, enquadrava a plataforma como um “diretório social” (Facebook Newsroom, 2006), um “utilitário social” (Facebook Newsroom, 2006), e uma “plataforma” (Facebook Newsroom, 2007). David Kirkpatrick (2010, pp. 215-217) descreve, em seu livro “O Efeito Facebook” (*The Facebook Effect*), como Zuckerberg sempre imaginou o Facebook como uma plataforma computacional para que outros aplicativos pudessem ser processados, desde o seu início, em 2004, como Thefacebook;

Ele [Zuckerberg] quis fazer com a web o que Gates fez com o computador pessoal: criar uma infraestrutura de software padrão que tornasse fácil a construção de aplicativos — daquela vez, aplicativos que tivessem um componente social. “Queremos transformar o Facebook em algo como um sistema operacional, para que você possa executar aplicativos”, ele [Zuckerberg] explicou. (Kirkpatrick, 2010, p. 217)

No outono de 2004, Zuckerberg trabalhava em outro projeto de software junto com o Thefacebook, o Wirehog, “um serviço de compartilhamento de conteúdo P2P (*peer-to-peer*)” (Kirkpatrick, 2010, p. 44). O aplicativo Wirehog foi integrado ao Thefacebook para utilizar as conexões de amizade dele a fim de compartilhar conteúdos no Thefacebook. Zuckerberg viu o Wirehog como “o primeiro exemplo que considerava o Facebook como uma plataforma para outros tipos de aplica-

² Conforme o Facebook, a plataforma tinha mais de 936 milhões de usuários ativos, em média, em março de 2015 (“Company Info,” s.d.). O Facebook.com é o segundo colocado na classificação dos “Top 500 sites da web”, feita pelo site Alexa, cuja ordem é “calcula usando uma combinação da média de visitantes diários e visualizações de páginas nos último mês” (“The top 500 sites on the web”, 2015).

tivos” (Kirkpatrick, 2010, pp. 99-100). Então, em vez de uma rede social, Mark Zuckerberg viu e elaborou o Facebook como uma plataforma, desde o seu início. O desenvolvimento do Facebook como uma plataforma deve ser percebido no vasto contexto da web 2.0 – “a web como plataforma” (O’Reilly, 2005) –, no qual a web foi situada como plataforma de desenvolvimento.

WEB 2.0: A WEB COMO PLATAFORMA

Os sites de redes sociais são tipicamente classificados como um tipo específico de aplicativo da web 2.0 (Beer & Burrows, 2007) ou um tipo dentre as mídias sociais (Van Dijck, 2013, p. 8). O termo “sites de redes sociais” se tornou popular na primeira conferência sobre web 2.0, em 2004, quando Tim O’Reilly definiu a web 2.0 ao considerar “a web como plataforma” – frase utilizada para situar a web como uma “robusta plataforma de desenvolvimento” na qual “websites se tornam componentes de softwares” (O’Reilly & Battelle, 2004). A web 2.0 ou “a web participativa” é agora compreendida como um grande conjunto de serviços que promovem a colaboração e a participação (Madden & Fox, 2006).³

O’Reilly colocou o sentido computacional do termo “plataforma” no centro do conceito de “web como plataforma”. Com o advento da web 2.0, O’Reilly (2005) não mais via a web apenas como um meio para a publicação de informações – que ele retrospectivamente rotulava de web 1.0 –, mas como uma infraestrutura para a construção de aplicativos, um sistema operacional de distribuição, que poderia fornecer serviços de software. Por esse motivo, Matthew Allen (2013) argumenta que deveríamos ver a web 2.0 como uma “tecnologia retórica” na qual “a indústria de computação tentou mudar a maneira como pensamos a internet” (Allen, 2013, p. 264): de um canal de publicação para uma plataforma de desenvolvimento de software.

Todavia e depois da conferência mencionada, essa definição mais computacional da web 2.0, a “web como plataforma”, não teve continuidade, tal como argumenta Robert Gehl (2010). Em vez disso, esse autor afirma que a web 2.0 era vista como um renascimento da indústria após o abalo do modelo “ponto com”, e ainda mais no âmbito de debates públicos e acadêmicos, como uma revolução que reformularia o panorama midiático (Gehl, 2010, pp. 26-37). As tecnologias da web 2.0 eram vistas como aquilo que borrava as fronteiras entre produção e

³ Conferir a crítica feita por Michael Stevenson (2014) a respeito da suposta “virada participativa” da web.

consumo (Bruns, 2008), e faziam emergir novas formas de participação de usuários como parte de uma “cultura participativa” (Jenkins, 2006) online. Desse modo, embora a definição original de web 2.0 implicasse considerar a web como plataforma computacional, o termo seria incorporado em um sentido mais metafórico (cf. Gillespie, 2010), como plataforma de participação, associada à retórica de “empoderamento” e “democratização” (Beer, 2009, p. 986).

DE SITES DE REDES SOCIAIS PARA PLATAFORMAS DE MÍDIAS SOCIAIS

Para alternar o foco dessa definição conceitual mais ampla de plataformas de volta para uma compreensão computacional mais específica, e a fim de desenvolver uma crítica a respeito de considerar o Facebook como plataforma, procuramos explorar o desenvolvimento tecnológico de plataformas de software na web e, em particular, plataformas de mídias sociais. Fazemos isso ao recorrermos a outra definição computacional de plataforma, que foi proposta por Marc Andreessen (2007a), fundador da Netscape, em uma publicação que discutia o lançamento da plataforma Facebook:

Definitivamente, uma “plataforma” é um sistema que pode ser reprogramado, portanto, customizado por outros desenvolvedores externos — usuários — e, dessa forma, adaptado a incontáveis necessidades e nichos, até então não pensados ou sequer ajustados pelos desenvolvedores originais.

Para Andreessen (2007b), o termo principal nessa definição de plataforma é “programável”, que elimina os usos mais conceituais do termo: “Se você pode programá-la, então ela é uma plataforma. Se não pode, então ela não é uma plataforma.”

A programabilidade de plataformas web 2.0, argumenta McKelvey (2011), oferece uma nova linha de crítica nos estudos de plataforma que iniciam com o questionamento a respeito de como uma plataforma performa a sua programabilidade. A noção de programabilidade tem sido fundamental para a compreensão da lógica das novas mídias (Chun, 2011; Manovich, 2001)⁴ e, por extensão, figura central na análise da lógica subjacente das plataformas de mídias sociais.

⁴ Em seu livro seminal, intitulado “A Linguagem das Novas Mídias”, Lev Manovich (2001) argumenta que todos os objetos das novas mídias são representações numéricas, e é isso o que torna as mídias

A fim de investigar as pré-condições da programabilidade de plataformas de mídias sociais, baseamo-nos na definição de plataforma de software proposta por Evans, Hagiu e Schmalensee (2006, p. vii): “um programa de software que torna serviços disponíveis para outros programas de software por meio de Interfaces de Programação de Aplicativos (APIs)”. Decorre dessa definição que, para se tornar uma plataforma, um programa de software – ou um website – precisa oferecer uma interface que possibilita a sua reprogramação: uma API.

Uma API é uma interface proporcionada por um aplicativo, que possibilita aos usuários interagirem com – ou em resposta a – dados ou solicitações de outros programas, outros aplicativos ou websites. As APIs facilitam a troca de dados entre os aplicativos, possibilitam a criação de novos aplicativos, e constroem a base para o conceito de “web como plataforma”. (Murugesan, 2007, p. 36)

De volta à concepção de O’Reilly, que considera a web como plataforma de desenvolvimento para novos serviços, apreendemos que não apenas a web como um todo, mas também os websites são transformados em plataformas ao tornarem disponível uma API.⁵ Por exemplo, o Facebook é uma plataforma porque ele disponibiliza uma API, que pode ser usada por desenvolvedores e proprietários de sites para construir novos serviços no Facebook e integrar novos websites e aplicativos aos dados e às funcionalidades do Facebook.⁶ O aplicativo de encontros *Tinder* é um exemplo de um aplicativo que foi feito sobre a plataforma do Facebook: ele requer que os usuários façam o *login* com o Facebook, e se utiliza de dados do próprio Facebook, tais como curtidas e amigos em comum, para encontrar potenciais parceiros. Outra maneira de integração com o Facebook é demonstrada pelo ato de implementar funcionalidades específicas dessa plataforma, como o botão curtir (*like*), em outras páginas/sites.

Na “web como plataforma”, os websites podem ter duas interfaces diferentes: uma interface de usuário, para o consumo humano (ex.: Facebook.com), e uma interface de software, para o consumo de máquinas (ex.: Facebook Graph API). Essa interface de software, a API, torna programável um website ao oferecer acesso estruturado aos seus dados e funcionalidades, e também o torna

programáveis. Essa programabilidade é fundamental para os princípios subjacentes às novas mídias: representação numérica, modularidade, automação, variabilidade e transcodificação cultural.

⁵ Isso chama nossa atenção para o uso da plataforma como um conceito recursivo, tal como apresentado pelo desenvolvedor de software Dave Winer (1995), quem desde cedo viu a internet como uma meta-plataforma ou “plataforma máquina” que pode ser usada para construir novas plataformas.

⁶ Mais de 30 milhões de aplicativos e websites foram integrados com a plataforma do Facebook (Liu, 2015).

uma plataforma na qual outros podem construir. Para ampliar essa ideia, as APIs são colocadas no centro da mudança de *sites* de redes sociais para *plataformas* de mídias sociais. No momento que os sites de redes sociais disponibilizam APIs, eles se tornam plataformas de mídias sociais ao disponibilizarem sua programabilidade. A API, então, torna-se um lócus fundamental para a investigação das consequências dessa programabilidade.

A ASCENSÃO DAS APIS DE MÍDIAS SOCIAIS

No campo dos estudos de mídias, as APIs de mídias sociais têm sido compreendidas como a “cola” tecnológica da web social ao conectarem serviços e possibilitarem o compartilhamento de conteúdos (Bodle, 2011; Bucher, 2013; Langlois; Mckelvey et al., 2009), como objetos protocolares (Bucher, 2013), como instrumentos regulatórios que governam relações entre a plataforma e terceiros (Puschmann & Burgess, 2013), como o modelo de negócios da web social⁷ (Bodle, 2011; Bucher, 2013), e como ferramentas que constroem dados para o mercado de dados (Vis, 2013). Predominantemente, as APIs têm sido usadas e discutidas como “um método para a coleta de dados em plataformas de mídias sociais” (Lomborg & Bechmann, 2014). Contudo, menos atenção tem sido dedicada à história das APIs de mídias sociais,⁸ isto é, a emergência delas na web como parte da infraestrutura material das plataformas de mídias sociais e as consequências delas para a adaptação de um modelo de plataforma. Um dos relatos mais completos até agora tem sido o do blogueiro de tecnologia Kin Lane (s. d.), que se nomeia “evangelista da API” e que estuda “os negócios e as políticas das APIs” desde 2010.

Lane (2012) traça o aparecimento histórico das APIs da web, voltadas para desenvolvedores externos, ao início dos anos 2000, quando a *Salesforce*, em 1999, que virou *eBay* em 2001, e a *Amazon*, em 2002, começaram a oferecer as APIs como soluções para o comércio eletrônico entre empresas. Essa primeira geração de APIs da web, sobretudo disponibilizada por empresas de comércio eletrônico, focou na troca de dados entre diferentes aplicativos de negócios para possibilitar transações e gerenciamento de vendas (Lane, 2012). A plataforma de serviços web

⁷ A indústria se refere a esse modo como “Economia de API” ou “os efeitos econômicos emergentes possibilitados por empresas, governos, organizações sem fins lucrativos e indivíduos que usam APIs para fornecerem acesso programável direto aos seus sistemas e processos” (Willmott & Balas, 2013).

⁸ Uma exceção dentro dos estudos de mídias é o trabalho de Taina Bucher (2013) acerca de “Objetos de sentimento intenso: o caso da API do Twitter” no qual a autora apresenta um contexto histórico do papel das APIs na engenharia de software e discute brevemente as primeiras APIs públicas da web.

da *Amazon*, por exemplo, possibilitou que websites terceirizados pesquisassem em seu catálogo, disponibilizassem produtos da própria *Amazon*, e ganhassem comissões por recomendações de seus próprios sites. Ao fazer isso, a API da *Amazon* ampliou seus serviços de comércio eletrônico a outros websites.

Em meados dos anos 2000, uma nova geração de APIs da web, ofertadas por sites de redes sociais, alterou o foco das transações de vendas para o acesso de conteúdos gerados por usuários, informações de usuários e suas conexões (Lane, 2012).

Em 2003, o site de marcações sociais (*social bookmarking*) *del.icio.us* começou a oferecer acesso programático ao seu site, seguido pelo Flickr, em 2004, pelo YouTube, em 2005, pelo Last.fm, Facebook e Twitter, em 2006, e depois por vários outros sites de redes sociais, que passaram a anunciar as suas APIs (Du Vander, 2012; Lane, 2012). Robert Bodle (2011, p. 325) descreve como esses sites disponibilizaram seus conteúdos e funcionalidades como parte de uma estratégia de negócios na qual terceiros podem adicionar valor a uma plataforma, por meio da construção de novos serviços tendo por base tal plataforma. O autor explica como Tim O'Reilly apoiou diferentes negócios a buscarem uma estratégia de plataforma, por meio da abertura de seus valiosos dados, a fim de tocarem no bloqueio de plataforma (Bodle, 2011, p. 325). Em seu manifesto pela web 2.0, O'Reilly (2005) encorajou ainda mais a reutilização de dados ao recomendar que eles fossem “projetados para a ‘hackeabilidade’ e remixabilidade”, por meio do acesso a tais dados e serviços por parte de terceiros. O'Reilly (2005) considerou os dados como os “blocos de construção” da web 2.0. Esse acesso fez emergir a típica prática da web 2.0, a criação de *mashups*, isto é, a construção de novos aplicativos por meio da remixagem de dados e funcionalidade proveniente de fontes existentes, que se utilizam de APIs (Benslimane, Dustdar & Sheth, 2008). Por isso, a web 2.0 também se tornou conhecida como “a web programável” (Anderson, 2012; O'Reilly, 2005). A seguir, analisamos os diferentes tipos de programabilidade que as plataformas de mídias sociais oferecem por meio de suas APIs, com a finalidade de formularmos uma crítica que considere o Facebook como plataforma ao enfatizar suas condições distintas de programação e suas consequências.

NÍVEIS E CONDIÇÕES DE PROGRAMABILIDADE

Em uma postagem sobre a nova plataforma do Facebook, Marc Andreessen (2007b) explicou como a programabilidade das plataformas da internet pode ser facilitada em diferentes níveis, de modo a produzir o que ele considera como três

tipos de plataformas de internet. Esses níveis também podem servir como uma maneira de investigação crítica do papel da arquitetura de plataforma.

De acordo com Andreessen (2007b), a maioria das plataformas de mídias sociais fornecem o tão chamado Nível 1 ou “API de acesso” (*Access API*). Neste nível, desenvolvedores externos podem acessar dados de uma plataforma e sua funcionalidade por meio de solicitações, feitas via API, que representam operações específicas para a realização de uma tarefa, por exemplo, ler, escrever ou apagar dados (Andreessen, 2007b). A API é acessada “do lado de fora do sistema principal”, ou seja, “o código dos aplicativos dos desenvolvedores permanece fora da plataforma” (Andreessen, 2007b). O serviço de compartilhamento de fotos, Flickr, é um exemplo de uma API de acesso, em que um desenvolvedor pode construir um aplicativo terceirizado, tal como um visualizador de slides para mostrar fotos marcadas como “pôr do sol”, ao usar a API do Flickr para requisitar esses dados. Nesse cenário, o código do aplicativo está localizado em um servidor externo, e o aplicativo está hospedado fora do Flickr. A programabilidade de uma plataforma de Nível 1 é caracterizada pelo simples acesso aos dados e à funcionalidade. Os desenvolvedores podem construir novos aplicativos com base na plataforma e integrarem dados e funcionalidades a seus sites e aplicativos externos, mas não podem reprogramar a própria plataforma. Em outras palavras, a programabilidade de plataformas de Nível 1 é um modo de as plataformas se expandirem para além de seus limites.

A “API Plug-In”, de Nível 2, possibilita aos desenvolvedores “construírem novas funções que podem ser introduzidas ou plugadas ao sistema principal e à sua interface para usuários” (Andreessen, 2007b). Andreessen utiliza o Facebook como exemplo de API Plug-In, visto que essa plataforma não apenas possibilita aos desenvolvedores acessarem dados e funcionalidades do Facebook para construir novos aplicativos (API de acesso Nível 1), mas também possibilita aos desenvolvedores carregarem e usarem seus aplicativos no ambiente⁹ do Facebook, por meio do *Canvas Frame*. O *Canvas* é “um painel gráfico no qual se possa colocar seu aplicativo ou jogo diretamente no Facebook.com” a fim de “integrá-los completamente à experiência principal do Facebook” (Facebook Developers, s. d.-a).¹⁰

⁹ Em 24 de março de 2015, o Facebook lançou a plataforma Messenger, que “possibilita aos desenvolvedores facilmente construírem aplicativos que se integram com o Messenger”, sendo este o aplicativo de troca de mensagem do Facebook. Por meio da plataforma Messenger, os desenvolvedores podem plugar seus aplicativos ao Messenger, a nova plataforma de nível 2 do Facebook.

¹⁰ Em sua documentação para desenvolvedores o Facebook explica como isso funciona: “Aplicativos no Facebook são páginas carregadas no *Canvas frame*, que é um *canvas* em branco no Facebook no qual se pode rodar seu aplicativo. Você preenche o *Canvas frame* ao fornecer uma *Canvas URL* e uma *Secure Canvas URL* que contém o HTML, o JavaScript e o CSS que compõem o seu aplicativo. Estes serão

Enquanto o aplicativo roda no Facebook, o código desse aplicativo está localizado fora da plataforma do Facebook (Andreessen, 2007b).¹¹ Aplicativos de tela (*canvas apps*) possibilitam que usuários customizem suas experiências no Facebook, aspecto que chama atenção à reconsideração feita por McKelvey (2011) a respeito da ideia de “programar como um ato de composição”, elaborada por John van Neumann.

No Nível 3, referente à API “Ambiente de tempo de execução” (*Runtime Environment API*), aplicativos terceirizados rodam com o ambiente de tempo de execução da própria plataforma (Andreessen, 2007b). Andreessen explica que essa abordagem é mais similar às “tradicionais” plataformas de computação, como o sistema operacional Windows, no qual desenvolvedores constroem aplicativos para serem executados nesse próprio sistema (Andreessen, 2007b). A plataforma como “ambiente de tempo de execução” é a abordagem menos comum na web, uma vez que ela requer um sistema técnico mais complexo tanto para desenvolvedores quanto para bancos de dados e gerenciamento de armazenamento (Andreessen, 2007b).¹² Como consequência, a programabilidade de plataformas de mídias sociais é tipicamente habilitada por meio da API de Acesso e da API Plug-In ou por ambas. De modo mais específico, nos termos de Andreessen (2007b), o tipo mais comum de plataforma de mídia social é a de Nível 1, a API de Acesso (Twitter, Facebook, YouTube, Tumblr e Instagram), seguido pela API Plug-In, de Nível 2 (Facebook).¹³

Ao diferenciar dois tipos de plataformas, Andreessen (2007b) oferece um sistema com o qual podemos avaliar plataformas individuais baseadas em suas condições de programabilidade. Similarmente, ao recorrer à tipologia de interfaces proposta por Florian Cramer e Matthew Fuller (2008), McKelvey (2011)

utilizados pelos usuários que navegam no Facebook em HTTP e HTTPS, respectivamente. Quando um usuário carrega seu aplicativo *Canvas* no Facebook, nós carregamos o URL da tela dentro de um *iframe* nessa página. Isso resulta na exibição de seu aplicativo dentro do cromo padrão do Facebook.” (Facebook Developers, s. d. -a)

11 A Canvas URL aponta para o *host* externo em que o aplicativo é localizado, que então é carregado dentro de um *iframe* no Facebook.

12 Exemplos de plataformas de ambiente de tempo de execução de nível 3 mencionados por Andreessen incluem o Salesforce, que possibilita aos usuários inserirem o próprio código, e a plataforma Ning, do próprio Andreessen (2007b), que serve “para a criação e a execução de aplicativos de redes sociais”. Apesar de Andreessen alegar que todas as “plataformas são boas, ponto final”, ele afirma que chama “esses modelos de plataformas de internet de ‘níveis’ porque, como você vai do nível 1 para o nível 2, e então, para o nível 3, como será explicado, cada tipo de plataforma é mais difícil de construir, mas muito melhor para quem a desenvolve”. Nesse sentido, ele promove plataformas de nível 3, incluindo a sua própria, a *Ning*, como sendo esse tipo de plataforma a melhor para desenvolvedores.

13 As plataformas “ambiente de tempo de execução” de nível 3 estão localizadas principalmente no domínio *business-to-business*, como a Salesforce ou a Amazon.

argumenta que “posto que plataformas têm interfaces diferentes, a linha crítica possibilita comparar como as plataformas facilitam a programabilidade”. De modo mais claro, podemos comparar APIs de mídias sociais para analisarmos que tipo de programabilidade essas plataformas vislumbram, o que elas possibilitam e restringem, que tipo de funcionalidades e dados são disponibilizados aos usuários, e por quem eles são disponibilizados.

PLATAFORMIZAÇÃO DA WEB

O termo “plataformização” é utilizado para se referir à emergência da plataforma como modelo econômico e infraestrutural dominante da web social, bem como às consequências da expansão das plataformas de mídias sociais em outros espaços online. É central a esse processo a oferta de APIs, que tornam os sites de redes sociais em plataformas de mídias sociais. Para compreender esses efeitos, analisamos como as distintas condições de programabilidade de plataformas de mídias sociais possibilitam que tais plataformas se estendam pela web e formatem dados da web que lhe são externos. De modo mais claro, as plataformas promovem sua programabilidade para descentralizar a produção e recentralizar a coleta de dados (Gerlitz & Helmond, 2013).

Historicamente, os websites tornaram viável a sua programabilidade por meio da troca de dados, conteúdos e funcionalidades com terceiros de três maneiras, de modo a oferecerem as condições prévias de plataformização da web: a) separação entre conteúdo e apresentação, b) modularização de conteúdo e funcionalidades, e c) interface com base de dados.

Separação entre conteúdo e apresentação

A maioria dos websites são criados com o uso da Linguagem de Marcação de Hipertexto (*HyperText Markup Language* – HTML), que descreve o conteúdo e a apresentação de um documento da web. Uma vez que a HTML é uma tecnologia de apresentação, projetada para o consumo humano, e muitos sites HTML são mal formatados, é difícil para uma máquina extrair e processar informação estruturada proveniente de um website (MYLLYMAKI, 2002, p. 635). A Linguagem de Marcação Estendida (*Extensible Markup Language* – XML) aborda essas questões ao separar conteúdo, estrutura e apresentação em um formato baseado

em texto, para consumo de máquina (W3C, s. d.).¹⁴ Esses formatos, legíveis para máquinas e pessoas, possibilitam o compartilhamento de informação estruturada entre sistemas, até então, incompatíveis (Myllymaki, 2002, p. 635; W3C, s. d.). A XML tem sido uma evolução extremamente importante para a web, pois possibilita que os dados de websites sejam lidos por máquinas e intercambiáveis entre sistemas diferentes. Essa linguagem possibilita a formatação estruturada de informação para fins de transmissão, de modo a formar a base para vários mecanismos de troca de dados, que possibilitam o fluxo de dados dentro e fora de outros websites¹⁵. Richardson e Ruby (2008) argumentam que a linguagem XML é fundamental para tecnologias como RSS, XML-RPC e SOAP, as quais “formaram uma web programável, que estendeu a web humana para a conveniência dos programas de software” (Richardson & Ruby, 2008, p. xviii).¹⁶

De acordo com Liu (2004), a separação entre conteúdo e apresentação, por meio da linguagem XML, explicita o conhecimento tecnológico subjacente da “transmissão de informação, pós-industrial”, a qual requer que o conteúdo seja “transformável”, “autonomamente móvel” e “automatizado” (LIU, 2004, pp. 57-58). Essa separação, conforme Liu (2004, p. 59), torna o conteúdo “transcendental”, a fim de que ele possa ser despejado de um recipiente para outro, movendo-se do banco de dados para o “banco de dados na web”. Liu (2004) também descreve como a linguagem XML sinaliza uma mudança da primeira geração de sites HTML independentes para um novo tipo de website, que é preenchido com dados provenientes de bancos de dados externos (LIU, 2004, p. 57). Essas novas páginas web

14 Estrutura de um documento XML:

```
<book category="Fiction">
<title lang="en">Emma</title>
<author>Jane Austen</author>
<year>1916</year>
</book>
```

15 A linguagem XML está no centro de vários mecanismos importantes de troca de dados na web, incluindo XML-RPC e SOAP. O protocolo XML-RPC é baseado na ideia de procedimento remoto, chamada (RPC), para “fornecer transferência de controle e dados uma rede de comunicação” (Birrell & Nelson, 1984, p. 39). Foi desenvolvido, em 1998, por Dave Winer da Userland e Microsoft, para fazer pedidos para um computador remoto e trocar dados na web (Laurent; Johnston; Dumbill & Winer, 2001, p. x). O SOAP (*Simple Object Access Protocol*) é um “protocolo leve, usado para troca de informação codificada em XML” (Laurent et al., 2001, p. 172). Os serviços da web baseados em XML-RPC e SOAP possibilitam a troca de dados estruturados entre diferentes máquinas na web, que se comunicam por meio do protocolo de transmissão HTTP. Recentemente, o JSON se tornou o formato preferido para transmitir dados, em detrimento do XML, considerado um formato mais leve. Além disso, o estilo arquitetônico REST (*Representational State Transfer*), ganhou destaque para a construção de serviços da web. Por exemplo, a plataforma de mídias sociais Twitter oferece uma API REST que retorna dados em JSON.

16 Conferir a nota anterior. O RSS é um formato de distribuição da web para sites e blogs, para publicar um *feed* de seu conteúdo mais recente. Ele se baseia em XML, conferir: <<http://cyber.law.harvard.edu/rss/rss.html>>.

empregam, para extraírem e exibirem conteúdo dinâmico de terceiros, o que Liu (2004) nomeia “fluxo de dados” (*data pours*). O fluxo de dados é um código incorporado em uma página web, que demarca um espaço nessa página, o qual transfere dados a partir de bancos de dados externos e para tais bancos (Liu, 2004, p. 59).

Publicada nos primeiros dias da web 2.0, a ideia de “fluxo de dados” de Liu (2008) pode ser lida como uma reflexão antecipada a respeito da modularidade crescente da web, que foi, depois, atualizada por ele:

Minhas presentes observações a respeito dos “fluxos de dados” se aplicam, com ainda mais força, à web 2.0, onde conteúdos produzidos por usuários fluem tanto para dentro quanto para fora de bancos de dados (*back-end databases*), por meio de páginas web “modelo” (*template*), que muitas vezes são elegantes e projetadas de modo minimalista, isto é, construídas em volta de uma abertura cega de código parametrizado – como um buraco negro reversível – que soca todo o conteúdo para dentro e o arremessa para fora novamente (Liu, 2008, p. 320).

Os “fluxo de dados”, agora comuns à web 2.0, estabelecem canais de dados, para os fluxos de dados, entre websites e banco de dados externos.

Modularização de conteúdos e recursos

Ao separar conteúdo de apresentação, a linguagem XML compartimentaliza o conteúdo web, por meio da descrição estruturada de cada elemento em uma página web, e da transformação desses pequenos módulos de dados, que podem ser reutilizados. A compartimentalização de conteúdo torna os conteúdos existentes na web disponíveis para o consumo de máquina, e possibilita a circulação de conteúdo por meio de elementos modulares. A modularização é um aspecto fundamental no moderno design de software, que possibilita o gerenciamento de sistemas complexos ao dividi-los em módulos menores e encorajar a reutilização desses módulos (Baldwin & Clark, 2000; Gehl, 2012; Mckelvey, 2011). Ulrich e outros (2008) argumentam que, na web 2.0, “serviços muitas vezes disseminam sua funcionalidade ao acoplarem componentes modulares, nomeados *widgets*”, quer dizer, “uma arquitetura de plataforma exhibe um tipo especial de modularidade, no qual um produto ou sistema é dividido em um conjunto de componentes” (Baldwin & Woodard, 2009, p. 25). Tais *widgets* possibilitam a integração de

um conteúdo e funcionalidade de determinado serviço com outro website, o qual conta com algumas linhas de código, que criam um fluxo de dados. Os *widgets* se tornaram centrais, pois são objetos específicos para plataformas de mídias sociais distribuírem seus conteúdos em diferentes espaços da web e se estenderem pela web.

Um importante desenvolvimento referente a essa extensão veio do site de compartilhamento de vídeos YouTube. Em 7 de julho de 2005, o YouTube (2005) anunciou um novo recurso, que possibilita aos usuários inserirem uma lista de seus vídeos no YouTube em seus próprios websites, ao copiarem e colarem o código HTML gerado. Este código incorporou um *widget* do YouTube e mostra uma lista de vídeos e miniaturas de imagens (*thumbnails*) que se vinculam aos vídeos no YouTube. Um mês depois, o YouTube anunciou um novo *widget*, que incorporou o *player* de vídeos, de modo que os vídeos do YouTube podem ser agora diretamente iniciados a partir de qualquer website (YouTube, 2005). O *widget* possibilitou a distribuição e a visualização de vídeos do YouTube fora desse mesmo site. Esse recurso de incorporação de vídeos é comumente visto como um fator importante do sucesso do YouTube, pois possibilitou ao YouTube fazer circular vídeos em redes sociais online, blogs e outras partes da web, por meio da modularização e descentralização dos recursos de sua plataforma (Cheng, Dale & Liu, 2008).

Enquanto o YouTube criou seus próprios *widgets* para distribuir conteúdos fora de seu próprio website, a rede social online MySpace teve um importante papel na popularização de *widgets* terceirizados, que compartilham conteúdos dentro dessa rede. Em contraste com outras redes sociais online populares em 2005 e 2006 — como o Friendster —, o MySpace possibilitou aos usuários inserirem códigos em suas páginas de perfil e adicionar botões de inicialização de músicas, álbuns de fotos e vídeos. Esta foi a primeira rede social online que apresentou tal arquitetura aberta e, com isso, fez surgir uma cultura de customização e assessoria de perfis (Boyd, 2007).

Com a habilidade de inserir códigos que podem ser incorporados em páginas de perfis, desenvolvedores terceirizados começaram a criar *widgets* a aparência e a funcionalidade do MySpace. Em novembro de 2005, a RockYou lançou o seu primeiro *widget* para o MySpace, em *flash*, para a criação e disponibilização de *slideshow*s de fotos. Um aspecto importante desses *widgets* iniciais é que, diferentemente dos *widgets* de compartilhamento do YouTube, eles não interagem diretamente com o banco de dados do MySpace. Os usuários não podiam carregar suas fotos diretamente do MySpace para o *widget* porque esse site não oferecia acesso estruturado (por meio de uma API ou outro recurso) a essas fotos. Em vez disso, os usuários tinham que, primeiramente, carregar

suas fotos para o site externo de hospedagem, ImageShack, dentro do *widget* RockYou (Tokuda, 2009).

Esta falta de uma interface direta com o banco de dados do MySpace diz respeito àquilo que Gehl (2012) se refere como a “falha de abstratação” do MySpace para extrair e monetizar conteúdo a partir de sua própria rede (Gehl, 2012, pp. 111-112). Ao passo que os *widgets* do MySpace estavam mais orientados a integrarem e distribuírem conteúdo em sua própria rede, os *widgets* do YouTube estavam orientados para a distribuição de conteúdo e funcionalidade fora de sua própria rede. Como muitos sites 2.0 começaram a oferecer códigos de incorporação e *widgets* para distribuírem seus conteúdos na web, a descentralização de recursos de plataformas se tornou central. Uma segunda distinção importante, distinta dos *widgets* do MySpace, diz respeito à interação direta dos *widgets* do YouTube com o banco de dados desse site. Todavia, os *widgets* voltados para o banco de dados do YouTube eram baseados em *fluxo de dados unidirecional* (*one-way data streams*), nas dinâmicas de descentralização, em que o conteúdo era recuperado a partir do banco de dados e disposto em um website externo. A nova geração de *widgets* seria baseada na interação direta com bancos de dados, a fim de possibilitar o *fluxo de dados bidirecional* (*two-way data streams*), nas dinâmicas de descentralização e recentralização, não apenas para ler os dados de bancos de dados, mas também para escrever novos dados neles.

Interação com bancos de dados

Os *plug-ins* sociais do Facebook são um conjunto de ferramentas, ou *widgets*, incluindo o ubíquo botão “curtir” (*like*), “que te possibilita compartilhar a sua experiência com o Facebook com seus amigos e outros no Facebook” (Facebook Help, s. d.). Os *plug-ins* funcionam como módulos que ampliam as funcionalidades das plataformas para websites externos (cf. Bodle, 2011). Ao mesmo tempo, Tania Bucher (2012b) argumenta: eles funcionam como “dispositivos criadores de conexões”, coletando dados criados pelas conexões ou “arestas” fora do Facebook.com, e os enviando de volta ao banco de dados da plataforma (Bucher, 2012b, p. 6). Os *plug-ins* sociais são uma parte importante da arquitetura de plataforma do Facebook, habilitando a descentralização da funcionalidade da plataforma e de dados nela produzidos, e possibilitando a recentralização de dados produzidos fora dela (Gerlitz & Helmond, 2013). Ao incorporarem um *plug-in* em seu website, donos e criadores de sites configuram canais bidirecionais de dados, fluxos de dados, nos quais os dados fluem entre o website e o banco de dados do Facebook.

Tecnicamente, um *plug-in* social funciona como uma chamada ou ligação à API (Helmond, 2013) e envia requisições específicas para a plataforma do Facebook. Por exemplo, obtém o número total de pessoas que gostaram de tal publicação ou publicaram um novo “curtir” (*like*) depois de clicar no botão “curtir”.

PREPARAÇÃO DE DADOS PARA AS PLATAFORMAS

Antes que esses *plug-ins* possam interagir com o banco de dados do Facebook a partir de um website externo, donos e criadores de sites precisam tornar seus websites compatíveis com a infraestrutura da plataforma do Facebook. Para fazer isso, os donos e criadores de sites precisam incorporar uma parte do código *JavaScript* em seus websites, que estabelece um canal de comunicação de dados com a plataforma do Facebook. Esse código inicializa o Facebook *Software Development Kit* (SDK) para a utilização de *plug-ins* sociais, Facebook *login*, e fazer chamadas à API, para os bancos de dados (Facebook Developers, s. d.-e). Ao fazerem isso, os donos e criadores de sites estão preparando suas páginas, como plataformas, para a comunicação de dados com o Facebook. Essa noção de preparar websites externos e dados web de plataformas amplia a ideia proposta por Gillespie (2014) referente a como os dados são “preparados como algoritmos” (Gillespie, 2014, p. 168) para realçarem o papel da infraestrutura de plataforma em reconfigurarem dados externos para se adequarem à agenda da plataforma.

Uma outra parte importante da infraestrutura do Facebook é o Facebook *OpenGraph*, que é explicitamente voltado para tornar os dados externos, prontos para plataforma. O *OpenGraph* “te possibilita integrar aplicativos diretamente à experiência do Facebook, o que aumenta o engajamento, a distribuição e o crescimento” (Facebook Developers, s. d.-d). Para integrar um aplicativo, os desenvolvedores precisam utilizar o Facebook SDK e o Facebook *Login* para configurarem relações entre o aplicativo, o Facebook e o usuário (Facebook Developers, s. d.-d). Essa integração possibilita aos aplicativos contarem “histórias” no Facebook tal como “Maria correu seis milhas com o *MyRunningApp* (Facebook Developers, s. d.-d). Os aplicativos submetem essas histórias ao *OpenGraph* de uma maneira muito estruturada, organizada em volta de quatro elementos, por exemplo: João (ator) está lendo *A Odisseia* (o objeto) no *Goodreads* (aplicativo). Há um número de ações pré-definidas tais como: “curtir”, “visualizar” e “ler”, mas os desenvolvedores também podem criar suas próprias ações. Bucher (2012b, p. 5) descreve esses esforços do Facebook “como um caminho para a construção de um mapa semântico da internet”. As integrações de aplicativos possibilitam que o Facebook colete

dados externos de aplicativos e atividades de uma maneira bastante estruturada, envie-os de volta ao banco de dados e os conecte a um usuário ou a outro dado. Isso amplia ainda mais as técnicas de coleta de dados no Facebook para aplicativos externos, e formata tais dados de acordo com a lógica da plataforma, para que, então, os dados sejam colocados em novas relações com a plataforma.

Donos e criadores de sites também podem preparar suas plataformas de websites ao marcarem esses sites com as *tags OpenGraph* (Facebook Developers, n.d.-b). Essas meta *tags* alimentam o *crawler* do Facebook com “informações estruturadas a respeito de uma página, tais como: título, descrição, imagem prévia etc.”, e controlam como o conteúdo aparece no Facebook a fim de “aprimorar a distribuição e o engajamento” (Facebook Developers, s. d.-c). De modo similar à prática de Otimização de Mecanismo de Busca (*search engine optimization* – SEO), realizada por donos e criadores de sites, preparar websites para o Facebook pode ser vista como uma forma de otimização de mídias sociais.

O *OpenGraph* mostra como o Facebook rigorosamente estrutura o fluxo de dados de aplicativos e websites externos para a plataforma a fim de prepará-lo. Enquanto as plataformas se posicionam como intermediárias neutras ou utilidades (Gillespie, 2010; Van Dijck, 2013), elas (pré)formatam a passagem de dados, por meio de sua infraestrutura, de acordo com a lógica de suas infraestruturas subjacentes.

DUPLA LÓGICA DE PLATAFORMIZAÇÃO

Os exemplos anteriores mostraram como o Facebook emprega a sua plataforma como um modelo de infraestrutura para ampliar a si mesmo para espaços online externos, e como ele emprega essas extensões para formatar dados para que a sua plataforma se encaixe em seus interesses econômicos, por meio da mercantilização das atividades de usuários e conteúdos da web e de aplicativos. Essa plataforma, eu argumento, está baseada na lógica dual de expansão das plataformas de mídias sociais para o resto da web e, simultaneamente, no movimento delas para prepararem dados externos da web e de aplicativos para fins de plataforma.

Como modelos de infraestrutura, as plataformas de mídias sociais ofertam um sistema tecnológico no qual outros possam construir, adaptam-se à conexão e alcançam resultados em outros websites, aplicativos, bem como em relação aos dados deles. Ao mesmo tempo, a preparação de dados externos para os seus próprios bancos de dados é fundamental para o modelo econômico das

plataformas de mídias sociais. Esse duplo processo – descentralização de recursos de plataforma e recentralização de dados prontos para plataforma – caracteriza o que chamamos dupla lógica de plataforma. Essa dupla lógica é operacionalizada por meio de objetos oriundos da própria plataforma, tais como: APIs, *plug-ins* sociais e o *OpenGraph* – que conecta o modelo de infraestrutura da plataforma aos seus objetivos econômicos. Esses elementos servem como dispositivos primordiais para que plataformas de mídias sociais se estendam pela web e para a criação de canais de dados – fluxos de dados – a fim de coletarem e formatarem dados da web externos às plataformas, para que possam se encaixar na lógica subjacente da plataforma.

Ao propormos uma perspectiva técnico-material a respeito das plataformas, mostramos o “trabalho realizado pelas plataformas” não em um sentido retórico (cf. Gillespie, 2010), mas em um sentido computacional. A noção de plataforma foi introduzida como meio de criticar as consequências da programabilidade das plataformas. Essa foi uma primeira exploração nessa área, ao mostrar como as plataformas de redes sociais colocam em ação a sua programabilidade para redesenhar a web para redes sociais.

REFERÊNCIAS

- Allen, M. (2013). What was Web 2.0? Versions as the dominant mode of internet history. *New Media & Society*, 15, 260-275. doi:10.1177/1461444812451567
- Amazon. (2002, July 16). *Amazon.com launches web services* [Press Release]. Recuperado de <http://phx.corporate-ir.net/phoenix.zhtml?c=176060&p=irol-newsArticle&ID=503034>
- Anderson, P. (2012). *Web 2.0 and beyond: Principles and technologies*. Boca Raton, FL: CRC Press.
- Andreessen, M. (2007a, June 12). *Analyzing the Facebook Platform, three weeks in* [Blog post]. Recuperado de https://web.archive.org/web/20071002070223/http://blog.pmarca.com/2007/06/analyzing_the_f.html
- Andreessen, M. (2007b, September 16). *The three kinds of platforms you meet on the Internet* [Blog post]. Recuperado de <https://web.archive.org/web/20071002031605/http://blog.pmarca.com/2007/09/the-three-kinds.html>
- Arrington, M. (2008, September 15). *Facebook isn't a social network: And stop trying to make new friends there* [Blog post]. Recuperado de <http://techcrunch.com/2008/09/15/facebookisnt-a-social-network-and-dont-try-to-make-new-friends-there/>
- Baldwin, C. Y., & Clark, K. B. (2000). *Design rules: The power of modularity*. Cambridge, MA: MIT Press.
- Baldwin, C. Y., & Woodard, C. J. (2009). The architecture of platforms: A unified view. In A. Gawer (Ed.), *Platforms, markets and innovation* (pp. 19-44). Cheltenham, UK: Edward Elgar Publishing Limited.
- Beer, D. (2009). Power through the algorithm? Participatory web cultures and the technological unconscious. *New Media & Society*, 11, 985-1002. doi:10.1177/1461444809336551

- Beer, D., & Burrows, R. (2007). Sociology and, of and in Web 2.0: Some initial considerations. *Sociological Research Online*, 12(5). Recuperado de <http://dx.doi.org/10.5153/sro.1560>
- Benslimane, D., Dustdar, S., & Sheth, A. (2008). Services mashups: The new generation of web applications. *IEEE Internet Computing*, 12(5), 13-15.
- Birrell, A. D., & Nelson, B. J. (1984). Implementing remote procedure calls. *ACM Transactions on Computer Systems*, 2, 39-59. doi:10.1145/2080.357392
- Bodle, R. (2011). Regimes of sharing. *Information, Communication & Society*, 14, 320-337. doi:10.1080/1369118X.2010.542825
- Bogost, I., & Montfort, N. (2009). Platform studies: Frequently questioned answers. *Digital Arts and Culture*. Recuperado de <http://escholarship.org/uc/item/01rok9br.pdf>
- boyd, d. (2007, June 24). *Viewing American class divisions through Facebook and MySpace* [Blog post]. Recuperado de <http://www.danah.org/papers/essays/ClassDivisions.html>
- boyd, d., & Ellison, N. (2008). Social network sites: Definition, history, and scholarship. *Journal of Computer-Mediated Communication*, 13, 210-230. doi:10.1111/j.1083-6101.2007.00393.x
- Bruns, A. (2008). *Blogs, Wikipedia, second life, and beyond: From production to produsage*. New York, NY: Peter Lang.
- Bucher, T. (2012a). *Programmed sociality: A software studies perspective on social networking sites* (Doctoral dissertation). University of Oslo, Oslo, Norway.
- Bucher, T. (2012b). A technicity of attention: How software “makes sense.” *Culture Machine*, 13, 1-13. Recuperado de <http://www.culturemachine.net/index.php/cm/article/viewArticle/470>
- Bucher, T. (2013). Objects of intense feeling: The case of the Twitter API. *Computational Culture*, 3. Recuperado de <http://computationalculture.net/article/objects-of-intense-feeling-the-case-of-the-twitter-api>
- Cheng, X., Dale, C., & Liu, J. (2008). Statistics and social network of YouTube videos. In *16th International Workshop on Quality of Service (IWQoS)* (pp. 229-238). New York, NY: IEEE. Recuperado de http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4539688
- Chun, W. H. K. (2011). *Programmed visions: Software and memory*. Cambridge, MA: MIT Press. Company Info. (n.d.). Recuperado de <http://newsroom.fb.com/company-info/>
- Cramer, F., & Fuller, M. (2008). Interface. In M. Fuller (Ed.), *Software studies: A lexicon* (pp. 149-152). Cambridge, MA: MIT Press.
- DuVander, A. (2012, January 4). *Over 2,000 APIs added in 2011: Social, telephony, open government* [Blog post]. Recuperado de <http://www.programmableweb.com/news/over-2000-apis-added-2011-social-telephony-open-government/2012/01/04>
- Evans, D. S., Hagi, A., & Schmalensee, R. (2006). *Invisible engines: How software platforms drive innovation and transform industries*. Cambridge, MA: MIT Press.
- Facebook Developers. (n.d.-a). *Canvas*. Recuperado de <https://developers.facebook.com/docs/games/canvas/>
- Facebook Developers. (n.d.-b). *Facebook content sharing best practices*. Recuperado de <https://developers.facebook.com/docs/sharing/best-practices>
- Facebook Developers. (n.d.-c). *A guide to sharing for webmasters*. Recuperado de <https://developers.facebook.com/docs/sharing/Webmasters>
- Facebook Developers. (n.d.-d). *Open Graph overview*. Recuperado de <https://developers.facebook.com/docs/opengraph/overview/>
- Facebook Developers. (n.d.-e). *Quickstart: Facebook SDK for JavaScript*. Recuperado de <https://developers.facebook.com/docs/javascript/quickstart/v2.2>
- Facebook Help. (n.d.). *What are social plugins?* Recuperado de <https://www.facebook.com/help/103828869708800>
- Facebook Newsroom. (2006, May 3). *Facebook expands to include work networks*. Recuperado de <https://newsroom.fb.com/news/2006/05/facebook-expands-to-include-work-networks-2/>

- Facebook Newsroom. (2007, May 24). *Facebook unveils platform for developers of social applications*. Recuperado de <https://newsroom.fb.com/news/2007/05/facebook-unveils-platform-for-developers-of-social-applications/>
- Fetterman, D. (2006, August 15). *Facebook development platform launches*. Recuperado de <https://www.facebook.com/notes/2207512130>
- Gawer, A. (2014). Bridging differing perspectives on technological platforms: Toward an integrative framework. *Research Policy*, 43, 1239-1249. doi:10.1016/j.respol.2014.03.006
- Gehl, R. (2010). *A cultural and political economy of Web 2.0* (Doctoral dissertation). George Mason University, Fairfax, VA.
- Gehl, R. (2012). Real (software) abstractions on the rise of Facebook and the fall of MySpace. *Social Text*, 30, 99-119.
- Geminder, K. (2007, June 2). Platform is here. Recuperado de <https://www.facebook.com/notes/facebook/platform-is-here/2437282130>
- Gerlitz, C., & Helmond, A. (2013). The Like economy: Social buttons and the data-intensive web. *New Media & Society*, 15, 1348-1365. doi:10.1177/1461444812472322
- Gillespie, T. (2010). The politics of “platforms.” *New Media & Society*, 12, 347-364. doi:10.1177/1461444809342738
- Gillespie, T. (2014). The relevance of algorithms. In P. Boczkowski, K. Foot, & T. Gillespie (Eds.), *Media technologies* (pp. 167-194). Cambridge, MA: MIT Press.
- Hagiu, A. (2014). Strategic decisions for multisided platforms. *MIT Sloan Management Review*, 55, 71-80.
- Hands, J. (2013). Introduction: Politics, power and “platformivity.” *Culture Machine*, 14. Retrieved from <http://www.culturemachine.net/index.php/cm/article/viewArticle/504>
- Helmond, A. (2013). The algorithmization of the hyperlink. *Computational Culture*, issue 3. Recuperado de <http://computationalculture.net/article/the-algorithmization-of-the-hyperlink>
- Hicks, M. (2010, April 21). *Building the social web together*. Recuperado de <https://www.facebook.com/notes/facebook/building-the-social-web-together/383404517130>
- Jenkins, H. (2006). *Convergence culture: Where old and new media collide*. New York: New York University press.
- Kirkpatrick, D. (2010). *The Facebook effect: The real inside story of Mark Zuckerberg and the world's fastest growing company*. New York, NY: Random House.
- Lane, K. (2012, December 20). History of APIs. Recuperado de <http://history.apievangelist.com/>
- Lane, K. (n.d.). *About Kin Lane*. Recuperado de <http://kinlane.com/about/>
- Langlois, G., & Elmer, G. (2013). The research politics of social media platforms. *Culture Machine*, 14, 1-17. Recuperado de <http://www.culturemachine.net/index.php/cm/article/viewArticle/505>
- Langlois, G., Elmer, G., McKelvey, F., & Devereaux, Z. (2009). Networked publics: The double articulation of code and politics on Facebook. *Canadian Journal of Communication*, 34(3). Recuperado de <http://www.cjc-online.ca/index.php/journal/article/viewArticle/2114>
- Langlois, G., McKelvey, F., Elmer, G., & Werbin, K. (2009). Mapping commercial Web 2.0 worlds: Towards a new critical ontogenesis. *Fibreculture*, issue 14. Recuperado de <http://fourteen.fibreculturejournal.org/fcj-095-mapping-commercialweb-2-0-worlds-towards-a-new-critical-ontogenesis/>
- Laurent, S. S., Johnston, J., Dumbill, E., & Winer, D. (2001). *Programming web services with XML-RPC*. Sebastopol, CA: O'Reilly Media, Inc.
- Liu, A. (2004). Transcendental data: Toward a cultural history and aesthetics of the new encoded discourse. *Critical Inquiry*, 31, 49-84. doi:10.1086/427302/

- Liu, A. (2008). *Local transcendence: Essays on postmodern historicism and the database*. Chicago, IL: University of Chicago Press.
- Liu, D. (2015, March 25). *F8 2015: New ways to connect with the Facebook family of apps*. Recuperado de <https://newsroom.fb.com/news/2015/03/f8-day-one-2015/>
- Locke, L. (2007). *The future of Facebook*. Recuperado de <http://content.time.com/time/business/article/0,8599,1644040,00.Html>
- Lomborg, S., & Bechmann, A. (2014). Using APIs for data collection on social media. *The Information Society*, 30, 256-265. doi: 10.1080/01972243.2014.915276
- Madden, M., & Fox, S. (2006). Riding the waves of “Web 2.0.” *Pew Internet and American Life Project*, 5. Recuperado de http://www.pewinternet.org/files/old-media/Files/Reports/2006/PIP_Web_2.0.pdf.pdf
- Manovich, L. (2001). *The language of new media*. Cambridge, MA: MIT Press.
- McKelvey, F. (2011). A programmable platform? Drupal, modularity, and the future of the web. *Fibreculture*, 18. Recuperado de <http://eighteen.fibreculturejournal.org/2011/10/09/fcj-128-programmable-platform-drupal-modularity-and-the-future-of-the-web/>
- Montfort, N., & Bogost, I. (2009). *Racing the beam: The Atari video computer system*. Cambridge, MA: MIT Press. Understanding Web 2.0. *IT Professional*, 9(4), 34-41. doi:10.1109/MITP.2007.78.
- Myllymaki, J. (2002). Effective web data extraction with standard XML technologies. *Computer Networks*, 39, 635-644.
- O'Reilly, T. (2005). *What is Web 2.0: Design patterns and business models for the next generation of software*. Recuperado de <http://oreilly.com/web2/archive/what-is-web-2.0.html>
- O'Reilly, T., & Battelle, J. (2004). *Opening welcome: The state of the internet industry*. Presented at the Web 2.0 Conference, Hotel Nikko, San Francisco, CA. Retrieved from http://web2con.com/presentations/web2con/intro_tim_john.ppt
- Puschmann, C., & Burgess, J. (2013). *The politics of Twitter data* (HIIG Discussion Paper Series No. 2013-01). Recuperado de <http://papers.ssrn.com/abstract=2206225>
- Renzi, A. (2011). What is the politics of platform politics? *Television & New Media*, 12, 483-485.
- Richardson, L., & Ruby, S. (2008). *RESTful web services*. Sebastopol, CA: O'Reilly Media, Inc.
- Rieder, B., & Sire, G. (2014). Conflicts of interest and incentives to bias: A microeconomic critique of Google's tangled position on the Web. *New Media & Society*, 16, 195-211. doi:10.1177/1461444813481195
- Rochet, J.-C., & Tirole, J. (2003). Platform competition in twosided markets. *Journal of the European Economic Association*, 1, 990-1029. doi:10.1162/154247603322493212
- Rogers, R. (2013). *Digital methods*. Cambridge, MA: MIT Press.
- Stevenson, M. (2014). Rethinking the participatory web: A history of HotWired's “new publishing paradigm,” 1994-1997. *New Media & Society*. Advance online publication. doi:10.1177/1461444814555950
- Tokuda, L. (2009, June). *Social applications: Growth, use, and monetization*. Presented at the Global ICT Summit 2009, Tokyo, Japan. Recuperado de http://www.ict-summit.jp/2009/pdf/report6_tokuda.pdf
- The top 500 sites on the web. (2015, June 14). Recuperado de <http://www.alexa.com/topsites>
- Ullrich, C., Borau, K., Luo, H., Tan, X., Shen, L., & Shen, R. (2008). Why web 2.0 is good for learning and for research: Principles and prototypes. In *Proceedings of the 17th International Conference on World Wide Web* (pp. 705-714). New York, NY: ACM. Recuperado de <http://dx.doi.org/10.1145/1367497.1367593>
- Van Dijck, J. (2013). *The culture of connectivity: A critical history of social media*. New York, NY: Oxford University Press.

- Vis, F. (2013). A critical reflection on Big Data: Considering APIs, researchers and tools as data makers. *First Monday*, 18(10). Recuperado de <http://firstmonday.org/ojs/index.php/fm/article/view/4878>
- W3C. (n.d.). *XML Essentials*. Recuperado de <http://www.w3.org/standards/xml/core>
- Willmott, S., & Balas, G. (2013) Winning in the API economy. Recuperado de <http://www.3scale.net/wp-content/uploads/2013/10/Winning-in-the-API-Economy-eBook-3scale.pdf>
- Winer, D. (1995, August 22). *What is a platform?* Recuperado de <http://scripting.com/dave-net/1995/08/22/whatisaplatfrom.html>
- YouTube. (2005, August 21). *August 2005*. Recuperado de http://youtube-global.blogspot.nl/2005_08_01_archive.html

*Tradução por Tiago Salgado.

ENGAJADOS DE OUTRA MANEIRA; AS MÍDIAS SOCIAIS — DAS MÉTRICAS DE VAIDADE À ANÁLISE CRÍTICA¹

RICHARD ROGERS

Este capítulo tem dois objetivos gerais: criticar as métricas contemporâneas das mídias sociais, sobretudo aquelas descritas como métricas de vaidade, e desenvolver um conjunto alternativo de métricas (análise crítica), que desloquem o foco da mensuração do *self* online e da vaidade em mídias sociais para a rede de questões problemáticas (*issue networks*) e para o engajamento (Van Es & Schaefer, 2017). A justificativa para essa mudança de foco diz respeito ao fato de as mídias sociais não serem apenas espaços para a performance de si e para uma produtiva rede social de contatos (*networking*), mas um local para a mobilização de públicos em torno de questões e causas sociais.

“Métricas de vaidade” é um termo crítico, dos estudos de negócios, que adverte os analistas quanto à dependência da contagem bruta de visualizações de páginas e curtidas (*likes*) como indicadores de sucesso na economia do *like* e do êxito (Gerlitz & Helmond, 2013; Ries, 2009). Todavia, como eu relato, o termo também abarca, muito bem, a atração de métricas que indicam quão bem alguém está indo online e o vaidoso ato de exibi-lo. Com base nas altimétricas, que são o estudo (e aplicação) do impacto de métricas em pesquisadores acadêmicos de mídias sociais, proponho um projeto alternativo de métricas para mídias sociais (Priem, Taraborelli, Groth & Neylon, 2010). Em vez de me concentrar nas redes

¹ Este projeto recebeu financiamento do programa de pesquisa e inovação *Horizon 2020*, da União Europeia, sob o registro 732942.

sociais presentes nas mídias sociais, apresento a rede de questões problemáticas (*issue networks*), ainda em relação às mídias sociais, como local de engajamento para questões profissionais, acadêmicas, ativistas, artísticas, jornalísticas, e para pesquisadores que trabalham com questões sociais amplamente definidas. Trata-se de uma rede (ou comunidade de interesse, como outros a chamaram) que reside entre as celebridades, na superfície, e aquelas rigorosamente situadas em nichos, localizados externamente. Trata-se de um meio termo em que os atores estão “altamente ligados localmente, mas dificilmente ligados globalmente” (Venturini, 2016, p. 105). No projeto, eu introduzo as métricas para as redes de questões como alternativa às métricas de vaidade das redes sociais. Eu as chamo “análise crítica”, e destaco, assim, a crítica das métricas de vaidade, mas também aponto métricas alternativas para mensurar outras formas de engajamento com as mídias sociais.

Eu discuto o tipo de métricas que têm sido desenvolvidas ao longo dos últimos anos para argumentar que essas medidas, tal como as pontuações do Klout, quantificam nossos modos de engajamento com base em suposições específicas sobre o uso das mídias sociais (Beer, 2016; Gerlitz & Lury, 2014; Van Doorn, 2014). Para que servem as mídias sociais? Por que mensurar a atividade delas? Como podemos explicar a emergência de uma métrica de engajamento tal como a pontuação do Klout?

O Klout é um sistema de pontuação para mensurar a influência com impacto. Ele advém da atividade de alguém em plataformas como Facebook, Twitter, Instagram, Google Plus, YouTube, FourSquare e Wikipédia, bem como o Klout (Figura 1). Os rankings são fundamentados online e em outras bases, incluindo o Google Trends e “rankings do mundo real”, tal como a comparação da lista de mulheres mais poderosas produzida pela revista *Forbes* com as mesmas mulheres avaliadas pela pontuação do Klout. Os resultados das duas listas influenciam a classificação nos rankings?

Os cientistas de dados do Klout, em artigo seminal de 2015, acompanhado de uma apresentação no Slideshare (a versão mais popular do artigo), introduzem o Klout como uma medida de influência nas mídias sociais, juntamente com sua utilidade impactante e comercial, no seguinte caso:

Um usuário pode postar uma mensagem no Facebook acerca de sua experiência em um restaurante, com um link para o site do estabelecimento. Um usuário que lê a mensagem original pode escolher reagir a ela de várias maneiras, tais como: ler a mensagem, clicar no link para mais informações, compartilhar o link com outros usuários em sua própria rede ou efetivamente

visitar o restaurante para jantar. O tipo de reação indica a força de influência que a mensagem teve no usuário. (Rao, Spasojevic, Li & Souza, 2015, p. 2282).

Na apresentação disponibilizada no *Slideshare*, os cientistas de dados do Klout resumem as pontuações como “a habilidade de um usuário em incitar ações entre outros usuários” (Rao, Spasojevic, Li & Souza, 2016, slide 6). A pontuação de alguém – de 0 a 100 – é apresentada universalmente, bem como em tópico, como: economia, política e viagem. Quando o método do Klout se tornou público, em 2015, as pontuações de celebridades, como Justin Bieber, e líderes políticos, como Barack Obama, não eram as mais interessantes, sobretudo em comparação mútua. Em vez disso, a atenção estava nas pontuações que poderiam reforçar um caso de negócios. Assim, na categoria de viagem estavam empresas e títulos. A pontuação máxima no Klout era da editora de guias de viagem Lonely Planet, seguida por American Airlines, Travel + Leisure, Condé Nast Traveler e BBC Travel. O próximo passo poderia ser a companhia monetizar as empresas e títulos do Klout ao oferecer o mesmo espaço privilegiado para outras empresas.



FIGURA 1 – Sinais da pontuação Klout. Fonte: RAO et al. (2016).

Como podemos interpretar a atração das pontuações fora de casos de negócios? Em 2012, conforme um artigo da revista *Wired*, uma alta pontuação no Klout possibilitaria que uma pessoa “embarcasse mais cedo em um avião, tivesse acesso VIP a *lounges* em aeroportos, ficasse em melhores quartos em hotéis e recebesse bons descontos em lojas” (Stevenson, 2012, parágrafo 7). Tal materialização

da pontuação do Klout não se concretizou. O interesse nas pontuações de *like* do Klout, contudo, encontra-se menos naquilo que Jeremy Rifkin (2000) uma vez chamou de uma estúpida “era do acesso” comercial do que na interpretação daquilo que as mídias sociais têm para oferecer ao *self* nesses dias.

Para atrair atenção e conexões, em primeiro lugar, performa-se nas mídias sociais de uma maneira caracterizada por Jenna Wortham (2012), no *The New York Times*, como “teatro de sucesso”, isto é, mostra-se aos outros que se tem sucesso. Marieke van Dijk (2014), a pesquisadora holandesa de design, de modo semelhante, chama as mídias sociais de *opschepmedia*, termo alemão para “mídias ostensivas”. Aqui, as mídias sociais se tornam um palco para a apresentação do “*self* que gostaríamos de ser” (Goffman, 1956, p. 12).

Em um segundo propósito, as mídias sociais podem ser usadas para construir uma “rede produtiva”, e é instrutivo se referir ao debate inicial se elas devem ser nomeadas “redes sociais” ou “software de rede social”. Este último termo representa uma descrição mais útil para o seu propósito, enquanto o primeiro retrata “amigos na vida real” (boyd & Ellison, 2007). Se for um software de rede social, então espera-se que a rede seja ativa, tornando o software mais produtivo para o *self*. A rede produtiva, agora migrada para o online, refere-se ao valor de conexão, em que laços valiosos são forjados. De fato, a rede produtiva foi construída na plataforma de conexão comercial LinkedIn. Também poderia ser dito que ela está presente no Twitter, bem como no Facebook, apesar de estes dois serem bem menos orientados para oportunidades lucrativas do que para a busca de informações profissionais e um ambiente em que se segue amigos.

Como essa rede poderia ser usada de modo produtivo? Há basicamente duas formas de redes produtivas com valor de conexão frequentemente associado às mídias sociais. Um conjunto, muitas vezes associado aos ensaios sobre mídias sociais, reintroduz o trabalho dos laços fracos, em referência ao clássico estudo de Granovetter (1973), que discute as oportunidades providas pela prática de networking com aqueles a quem se está ligado, mas não necessariamente próximo. O estudo também descreve como a rede de contatos com laços fortes (e recurso frequente a eles) pode ser prejudicial para a construção de um seguimento ou movimento mais amplo.

Em um popular debate desencadeado por Malcolm Gladwell (2010) acerca do valor das mídias sociais e o potencial delas para a mudança social, o principal ponto trata se, a partir do uso das mídias sociais, pode haver compromisso em participar de eventos na rua. O ato de seguir outros online atinge as ruas? Uma justificativa para o debate é reintroduzir e, finalmente, defender a tese do ativismo preguiçoso (*slacktivism*): a crítica do engajamento online que estipula que

as práticas de curtir e compartilhar são apenas formas de solidariedade de baixo custo (Mozorov, 2009). Todavia, uma contribuição para o debate se destaca pela identificação de onde se encontra o valor das mídias sociais: “Os desempregados devem passar seu tempo conversando com conhecidos que estão distantes, pelo Facebook” (Lehrer, 2010, parágrafo 7). Por mais grosseira que seja, a afirmação citada é uma tradução direta do valor dos laços fracos, e resume a ideia de que as plataformas são sites de redes sociais e têm valor de redes produtivas para o emprego empresarial (e não para os movimentos sociais).

Um terceiro propósito das mídias sociais (para aqueles interessados em desenvolver métricas como o Klout) é apreender e propagar o futurismo do consumidor, que se refere a uma das primeiras descobertas do uso das redes sociais para o marketing – a existência de “cliques” e a utilidade do mediador de fundos (*the broker*), ou o indivíduo “extremamente entre” (*highly between*), que influencia grupos com suas novas aquisições e interesses. Aqui, particularmente, usuários conectados semeiam desejo em outros, ou o que as pessoas vão querer consumir em um futuro próximo. Depois de uma conexão especialmente bem-sucedida (ou microcelebridade) nas mídias sociais, usuários também podem provocar uma “inveja de nicho”, em que um cobiça os privilégios de entrada e níveis de acesso do outro (Turow, 2006).

A combinação desses propósitos específicos das mídias sociais – teatro de sucesso e projeção, rede de relacionamento produtiva e futurismo de consumo – fornecem valor ao uso delas, bem como uma urgência para serem mensuradas. Quais tendências foram recentemente descobertas pelo aparente sucesso e boas conexões? Quando um seguidor reserva uma mesa no mesmo restaurante tal como recomendado por aquele que ele segue, esta é uma ação que vai na direção do sentido proposto pelo Klout, conforme o uso da pontuação por parte dessa empresa.

DAS MÉTRICAS DE VAIDADE À ANÁLISE CRÍTICA

De modo mais incisivo, com a pontuação do Klout, alguém pode se impressionar com a atratividade que uma pessoa tem sobre outras, o que se refere a como a vaidade é frequentemente definida. A publicação e o compartilhamento de uma pontuação, como o Klout, trazem à mente a noção proposta por Baudrillard (1990) referente às estatísticas como formas específicas de realização de desejos. As métricas de vaidade se tornam satisfatórias medidas de atratividade a outros. Uma consequência da métrica é que ela convoca uma pessoa a continuar

a performar para ser pontuada. Aqui a questão suscitada indaga se a métrica de fato mensura algo ou se ela induz o comportamento performático.

As métricas de vaidade não apenas medem, mas encorajam todos os desejos de vaidade: celebridade, influência e boa impressão. No teatro do sucesso das mídias sociais, a socialização intensificada (juntamente com a conectividade crescente) gera o que é denominado “microcelebridade”, ou o tratamento das audiências como “base de fãs” (Marwick & boyd, 2011; Senft, 2013). Daniel Boorstin (1961), o ex-bibliotecário americano do Congresso, uma vez definiu, reconhecidamente, o termo “celebridade” como a qualidade de ser bem conhecido por ser bem conhecido. A crítica que acompanha esse trabalho de Boorstin (1961) concerne a como a qualidade célebre é moldada pela fama em vez de ser concedida pela grandeza. As mídias sociais propagam esse ciclo de ser bem conhecido (e construção de celebridade como notoriedade em vez de grandeza) ao manterem a pontuação e ao exibirem-na em números, contas de seguidores e outros indicadores semelhantes. A palavra “micro”, no termo “microcelebridade”, é apropriada aqui no sentido de ser o resultado de fatores de fama reduzida.

As redes são produtivas não apenas por fabricarem exibições de microcelebridades metrificadas, mas também por produzirem influência. De fato, o segundo aspecto das métricas de vaidade deriva de ideias de influenciadores em redes ou da influência que uma determinada pessoa exerce, que é interpretada como influência palpável. Em *Linked: The New Science of Networks*, Barabasi (2002) discute como as redes de influência podem ser pensadas, em termos populares, como uma medida de distância de um aperto de mão. Quantos apertos de mão são de um diretor executivo de uma grande corporação e, portanto, da generosidade que ele pode ser capaz de dispensar? Um posicionamento na rede pode ser medido de acordo com o comprimento da trajetória ou sentido de outros nós, e estar “extremamente entre” significa estar particularmente bem posicionado, de tal forma que a distância de um em relação aos outros que interessam é curta (Freeman, 1977). A centralidade de intermediação (*betweenness centrality*) se torna uma pontuação influente em mídias sociais quando estas são vistas como plataformas produtivas de redes sociais. A influência de um é medida como tal, e os influenciadores se tornam, assim, identificados, valorizados e marcados.

Finalmente, outro desejo por trás da vaidade está relacionado ao futurismo do consumidor, às tendências e, então, à localização e disponibilização de boas impressões. A prática de previsões e mudanças de tendências, *cool hunting*, é uma habilidade altamente comercializável para analistas de rede sociais e gestores de comunidades (*community managers*). Como argumentam Frank (1997) e McGuigan (2009), a ascensão do “capitalismo legal” (*cool capitalism*) veio atrelada

à “conquista do legal” (*conquest of the cool*) pelos profissionais de marketing. Há medidas de rede (como os *trending topics* do Twitter) para aumentar a relativa novidade, bem como aquelas que a espalham a futuros consumidores (a extrema intermediação), mas a prática de *cool hunting* adiciona a concessão legal da tendência. Quem está agindo em uma tendência de um jeito ou de outro como uma afirmação de estilo? Isso pode ser rotulado como *coolness* (boa impressão) (Liu, 2004). Como Alan Liu apontou, aquilo que é legal, bacana, tem sido encontrado online desde o advento da web, e tem sido uma categoria de site desde que os primeiros diretórios foram feitos para organizarem a web por meio de premiações referentes ao “sites mais legais do mês” e outras taxonomias semelhantes, como “estranho” no antigo (e contemporâneo) *Webby Awards*. Online, aquilo que é ousado deve ser encontrado, rotulado e valorizado como tal.

Mais recentemente, com o ressurgimento do filtro bolha — em que o conteúdo é recomendado com base nas preferências dos usuários —, as tendências também se tornaram personalizadas (Pariser, 2011). Os *trending topics* do Twitter são baseados na localização geográfica e em outros “sinais” que, em certo sentido, deslocam e distribuem as tendências de uma escala e escopo sociais abrangentes para as mais pequenas redes de microcelebridades e em nicho, que então se espera que as espalhe.

Até agora, o argumento diz respeito às suposições acerca dos propósitos das mídias sociais — teatro de sucesso e projeção, redes produtivas e futurismo de consumidores —, que valorizam seu próprio uso e uma urgência para serem mensuradas. As mensurações são métricas de vaidade porque consideram, seriamente, as atividades de estar em cena, de modo bem-sucedido, nos sites de redes sociais, como produtivas e valiosas. A atratividade das métricas de vaidade está nos desejos que elas semeiam e refletem no *self*, mas também está na promessa de identificação da tendência e dos formadores de tendência, em uma realidade associada com aquilo que é legal ou bacana. As métricas, como as pontuações do Klout, gostariam de monetizar as escolhas impactantes de restaurantes dos influenciadores.

Proponho uma alternativa às métricas de vaidade, entretanto, algo que possa ser construído como uma agenda ou uma proposta a ser seguida como “métricas alternativas” para mídias sociais (em vez de um projeto para a análise de citações acadêmicas, de onde o termo é tomado de empréstimo). Já há usos alternativos, como “ativismo analítico” (*analytic activism*) (Karpf, 2017). Organizações não governamentais (ONGs), partidos políticos e agências governamentais semelhantemente mensuram a ressonância de suas campanhas e o significado das menções de seus slogans, bem como exercícios de monitoramento de lideranças

nas mídias. As ONGs, em especial, estão também desenvolvendo métricas para estudarem o alcance da eficácia, tal como *SumOfUs's MeRa* (Membros que Retornam para a Ação) ou o “número de membros únicos que tomaram uma ação diferente da primeira” (Karpf, 2017, p. 148). As análises críticas, apresentadas na discussão que segue, são menos um esforço para alcançar os objetivos das organizações de base, das campanhas políticas ou os empenhos de marcas corporativas, embora possam ser reaproveitadas como tal. Em vez disso, elas oferecem uma abordagem global (com exemplos) para estudar questões problemáticas nas mídias sociais, mais do ponto de vista do pesquisador social do que do profissional de marketing ou de campanhas (Marres, 2015).

As análises para os movimentos sociais (ou a mensuração do “poder dos movimentos”) também foram concebidas por meio da operacionalização do acrônimo WUNC, proposto por Charles Tilly, que indica dignidade, unidade, números e comprometimento (Tilly & Wood, 2016; Freelon, McIlwain & Clark, 2016). No estudo de caso referente à hashtag #blacklivesmatter no Twitter, a unidade é concebida como disciplina da hashtag (cultura de uso e apropriação), os números como colaboradores reconhecidos, e o compromisso como participação repetida. A dignidade não é quantificada. Apesar de esses termos serem apresentados como meios para mensurarem um movimento social online, eles também são uma forma de análise crítica, como descrito aqui, porque localizam uma rede de interesse e criam indicadores que são alternativas às métricas de marketing, tais como influência ou poder.

A análise crítica propõe uma alternativa às métricas de vaidade, e isso começa com uma mudança de perspectiva em relação às mídias sociais; vê-las não apenas como sites de redes sociais produtivos para autoapresentação. Em vez disso, as mídias sociais também podem ser vistas como espaços para se estudar redes sociais de interesse (*social issue networking*). Isso significa que as métricas podem ser criadas para as mídias sociais, que não são construídas como espaços de vaidade, mas como espaços para o trabalho de questões sociais. Embora as métricas propostas abaixo não estejam, de forma alguma, completamente estabelecidas, as redes sociais de interesse (*social issue networking*) deveriam, minimamente considerar: a) os atores específicos que dão voz à questão e com maior força; b) as áreas ou campos de questões que aceitam os interesses e aqueles que os ignoram; c) a longevidade ou durabilidade do interesse dos atores; 4) a articulação específica dos atores, bem como a contra-articulação; e 5) o conjunto de atores que especificam o interesse da mesma maneira, mas que podem não ser aliados. Nesta apresentação do espaço de questões (*issue space*) a ser cartografado, temos, portanto, uma voz dominante, interesse, compromisso, posicionamento e alinhamento. Explico cada um desses pontos com uma série de ilustrações.

Nas ruas, em reuniões, em panfletos, mas também em fóruns online, websites, blogs, espaços para comentários, e agora, cada vez mais, nas plataformas de mídias sociais, como Facebook, Twitter e Instagram, causas são promovidas, assumidas (ou ignoradas), prometidas, rearticuladas e reunidas, e os analistas monitoram e mensuram a atividade. Cinco medidas para mídias sociais e outras mídias online são apresentadas aqui. Primeiro, a *voz dominante* apreende as fontes consideradas mais impactantes (apesar de não necessariamente credíveis) no espaço de questões (*issue space*). Em segundo, o *interesse* considera se uma pessoa ou organização (ou conjuntos) estão presentes ou ausentes neste espaço. Quem está fazendo ou ocupando a questão problemática, e quem diferiu? Em terceiro, o *compromisso* é a longevidade ou persistência do interesse. Os atores se movem dentro ou fora do espaço de questões como os seguidores de tendências, sem náuseas, ou eles agem de acordo com seus interesses? Em quarto, o *posicionamento* é determinado por meio da escolha de palavras que são utilizadas para denotar e discutir a questão de interesse. Essas palavras são parte de uma agenda ou uma tomada de posição, ou são esforços conscientes para ficar fora da briga? Por último, o *alinhamento* é um termo para a formação de grupos por meio do posicionamento. Quer dizer, quem mais está usando a mesma linguagem e, assim, compartilha a mesma posição? Essas medidas podem ser companheiras peculiares, no sentido de não pertencerem a uma área, campo, coalizão ou parceria, mas a escolha da linguagem por parte delas, alinha elas com outras, que optam por empregar os mesmos termos. Juntas, estas e outras altimétricas para o trabalho de questões em mídias sociais oferecem alternativas para mensurar vaidade em espaços de autoapresentação; contudo, tais usos vaidosos das mídias não deveriam ser completamente dispensados ou mesmo totalmente separados daqueles descritos na próxima seção.

business

FIGURA 2 — Menções de “vacina contra o HIV” nas seções de negócios e saúde dos principais jornais dos EUA em 2009. Análise feita por estudantes da Universidade de Amsterdã em maio de 2009.

ANÁLISE CRÍTICA

Voz dominante

Quais fontes estão dadas em um espaço (oficial) de questões, e daquelas quais dominam e quais “assuntos falados” são cortados ou marginalizados (Foucault, 1972)? Uma ilustração particularmente gritante do estudo da voz dominante é uma análise das seções do jornal em que a vacina contra o vírus HIV é mais discutida. Uma questão, em 2009, envolveu a perspectiva de criação de uma vacina que protegesse aqueles sem o HIV-AIDS ou tratasse aqueles com o vírus. Como um grupo de pesquisadores e eu descobrimos, uma vacina contra o HIV era proporcionalmente muito mais discutida nas seções de negócios do que nas de saúde, nos principais jornais dos EUA (Figura 2). Aqui, o espaço autoritário diz respeito às notícias, e a voz dominante se refere às notícias sobre negócios – embora valha a pena afirmar que, enquanto a seção de saúde pode ser marginal, uma vacina contra o HIV não é apenas uma notícia; é também uma questão que faz parte da saúde global, em termos mais amplos, e também da agenda transnacional e não governamental, como mostra a Figura 3.

Gates and Gates grantee issues compared and ranked.

- HIV/AIDS
- Malaria
- Poverty
- Vaccines
- Tuberculosis
- Maternal Health
- Infectious Diseases
- Cancer
- Food Security
- Global Health
- Immunology
- Child Health
- Epidemiology
- Climate Change
- Public Health
- Cell Biology
- Innovations
- Microbiology
- Brain Sciences
- Molecular Biology
- Reproductive Health
- Water and Sanitation
- Hunger
- Parasitology
- Vaccine Development
- Agricultural Development
- Malnutrition
- Noncom Health
- Population Health
- Adolescent Health
- Sanitation
- HIV Prevention
- Infection



FIGURA 3 – Questões financiadas e não financiadas pela Gates Foundation, conforme uma análise das agendas de questões (encontradas em seus websites) da fundação, bem como de seus beneficiários. Redimensionada conforme a frequência de menções e colorida conforme financiamento (vermelho) e não financiamento. Análise feita pela Digital Methods Initiative (DMI), 2014-2015.

A *Gates Foundation* (como as notícias) também é um espaço autoritário, que está entre as maiores fundações privadas no mundo, as quais financiam projetos relacionados ao desenvolvimento e à saúde global. Em uma comparação, feita em 2014 entre questões presentes na agenda de financiamento da *Gates* e questões de interesse para os atores no campo de saúde global e desenvolvimento, de maneira mais ampla, nota-se que certas questões são nominalmente financiadas e outras inevitavelmente não são (Figura 3). Entre as questões não financiadas no topo da agenda dos atores em campo estão: pobreza, segurança alimentar e mudanças climáticas.

É importante mencionar que as questões não financiadas podem ser cobertas pela fundação em outros termos, porque a análise concerne o idioma utilizado pelos atores, em vez de categorias ou agrupamentos mais amplos. Tal ponto de partida etnográfico ou heurístico leva a sério como os atores em campo implementam termos, como eu noto na última discussão sobre posicionamento. As questões não financiadas de pobreza e mudanças climáticas também poderiam ser descritas como questões ainda mais amplas, fora do escopo reconhecidamente amplo da saúde global e da linguagem específica de atores transnacionais e intergovernamentais (*segurança alimentar* é um termo cunhado no *World Food Summit* (WWF), em 1974, conforme o artigo a respeito desse termo publicado na Wikipédia).

Como mostram as Figuras 2 e 3, os negócios dominam a discussão sobre a vacina contra o HIV como um item de notícias, e os campos de desenvolvimento e saúde global realçam questões não financiadas pela *Gates Foundation*. A análise de voz dominante não apenas abre criticamente a discussão acerca da potencialidade dos significativos deficits de atenção, mas também um meio de inserir (em uma tradição de contramapeamento) a linguagem e as agendas da voz não dominante (Crampton, 2010).

Fukushima nuclear disaster as environmental and species concern?

QUERY: site:greenpeace.org Fukushima site:worldwildlife.org Fukushima

METHOD: Query leading environmental and species NGOs for Fukushima



map generated by tools.digitalmethods.net

FIGURA 4 — Nuvem de questões acerca de meio-ambiente e espécies, mostrando qual tipo de questão de interesse é Fukushima, conforme o número de menções do termo “Fukushima” nos sites da *Greenpeace* e do *World Wildlife Fund*, respectivamente, de acordo com os resultados de contagem de páginas da Google, em março de 2011.

Interesse

Para quais atores uma questão problemática é uma questão de interesse? O devir de uma questão de interesse foi resumido como o “redirecionamento da atenção” (Latour, 2008, p. 48) pelos públicos. Como grupos de direitos infantis, justiça social, ICT4D ou outras questões, as ONGs elaboram campanhas, kits de ferramentas, vídeos de animação gráfica, jogos e outros formatos para atrair a atenção de outros profissionais que se interessam por questões problemáticas na bolha (como é comumente referenciado de maneira coloquial) e o público em geral, externo a ela. Os eventos se realizam onde uma decisão deve ser tomada: tornar ou não uma questão problemática uma questão de interesse. A atenção deveria ser redirecionada em direção a essa questão? Em outras palavras, deveria haver esforços combinados para tornar a questão problemática uma questão de interesse?

O significativo acidente nuclear de Fukushima, em 2011, no Japão, causado por um tsunami, levou à liberação de radiação, incluindo a disseminação de água contaminada na costa. Uma vez que os eventos aconteciam, um grupo de pesquisadores de pesquisa-ação e eu perguntamos: para quem essa questão problemática é uma questão de interesse? Com o auxílio do *Lippmannian Device*, um *scraper* da *Google* que pesquisa e gera nuvens de origens redimensionadas, conforme o número de páginas retornadas por uma pesquisa acerca de uma questão problemática, pesquisamos por uma ONG ambiental líder, bem como por uma ONG preocupada com espécies (Figura 4).

Profissionais que lidam com questões problemáticas e outros podem questionar que tipo de questão problemática é Fukushima. Trata-se essencialmente de uma questão nuclear? Ou é uma questão política no sentido de que um partido político japonês está revestido de poder nuclear enquanto outros são firmes oponentes? O que Fukushima poderia fazer a uma questão problemática? O acontecimento Fukushima ressoou muito com o *Greenpeace* e muito pouco com o *World Wildlife Fund* (WWF). O resultado poderia indicar uma distribuição enviesada de interesse com o meio ambiente e distante dos animais ou da vida marinha, pelo menos em relação às consequências imediatas, de acordo com a equipe técnica. Do ponto de vista das ONGs e da saúde dos problemas suscitados, deve-se notar que, desde o tsunami, tanto o *Greenpeace* quanto o WWF têm sido ativos em relação às questões antinucleares no Japão, com campanhas baseadas nas lições de Fukushima.

As análises críticas aqui apresentadas são registros que relatam não apenas se, mas especificamente para quem uma questão problemática é uma questão

de interesse em um dado momento. Além disso, pode-se questionar: quem está ausente e quem pode estar sujeito a persuadir a entrar em outro momento, quando outro registro puder ser feito? O tema do compromisso é o interesse longitudinal ou a junção de registros de interesse. Trata-se do interesse ao longo do tempo em vez de diminuir (e não aumentar) a atenção, como descrito anteriormente.

Compromisso

Por quanto tempo uma questão problemática é uma questão de interesse para os atores? De fato, além de quem considera uma questão problemática uma questão de interesse (ou se a questão está ou não presente nas comunicações dos atores), a questão relativa à questão problemática diz respeito à longevidade ou à perseverança, em vez de concernir à diminuição da atenção por outros. A longevidade pode ser inconveniente, sobretudo se surgirem outras questões problemáticas, que possam receber mais atenção e, portanto, parecerem mais atraentes. Os financiadores podem ter deixado de lado as questões prévias. Você também deveria fazer isso? Ao se descrever o compromisso, pode-se diferenciar o cidadão, que deve se preocupar com uma comunidade para além de si mesmo, de um consumidor, que precisa se preocupar apenas consigo (Suleiman, 2003).

Como ilustração do estudo do compromisso como uma forma de análise crítica, apresento uma análise da longevidade dos interesses do *Greenpeace*, empreendida com base em seu comportamento de campanha, tal como expresso em seu site. Essa análise recorre ao *Wayback Machine* do site *Internet Archive*, com uma técnica de captura e renderização desenvolvida para analisar a história de um site ao longo do tempo como uma história organizacional (Rogers, 2013). Tendo coletado as páginas iniciais dos sites durante um período de sete anos, quais questões persistem (*Greenpeace* como cidadão) e quais são fugazes (*Greenpeace* como consumidor)? Nesse modo de mapeamento de compromissos, o *Greenpeace* é notavelmente consistente em sua principal campanha: de 2006 a 2012 (período de mapeamento) não hesitou em defender “nuclear”, “oceanos”, “tóxico”, “florestas”, “clima” e “paz e desarmamento”, como mostram os itens de menu lidos (Figura 5). Em cada ano, essas principais questões são apresentadas, embora um conjunto de outras apareçam e desapareçam (engenharia genética e comércio sustentável) ou apareçam e permaneçam, apenas para deixar a página inicial anos depois da análise (agricultura, por exemplo). A campanha anual em torno da mesma questão demonstra uma notável resistência em permanecer no mesmo tema.



FIGURA 5 — A nuvem de questões problemáticas dos temas de compromisso do *Greenpeace* mostra campanha consistente para os mesmos problemas ao longo de sete anos, de 2006 a 2012. Análise feita por Anne Laurine Stadermann.

A aparição diária do menu, em um site, e do conteúdo que deve ser gerado para sustentá-lo, a fim de não se tornar “vazio”, também poderia ser descrita como um dispositivo de compromisso. Um outro exemplo dessa fórmula de sustentação de questões problemáticas é o calendário das Nações Unidas, com seu Dia Mundial de combate à AIDS, em 10 de dezembro (e muitos outros dias ao longo do ano).

Permanece, porém, a questão de saber se as atividades de conteúdo confirmam tal achado de compromisso. Ainda no âmbito da análise de sites, o site do *Greenpeace* pode ser consultado para as principais palavras-chave utilizadas nas campanhas a fim de criar um registro do número de páginas dedicadas a cada uma (Figura 6). Em 2012, o site do *Greenpeace* apresenta uma distribuição desigual de atenção para cada uma das questões (em que a atenção é mensurada por contagem de página por questão); contudo, “nuclear” e “clima” tiveram mais duplicações (pela soma de 2.000 a 3.000 páginas com menções) do que “tóxico”. Assim, o compromisso, bem como a longevidade de interesse, também é continuamente testado com o advento de novas causas e técnicas para medir a extensão de até que ponto as antigas questões ainda importam.

Greenpeace International's distribution of concern according to campaign word count, 2012, on Greenpeace.org

QUERY: site:http://www.greenpeace.org Nuclear, "Climate change", etc.

METHOD: Query Google Scraper / Lippmannian Device

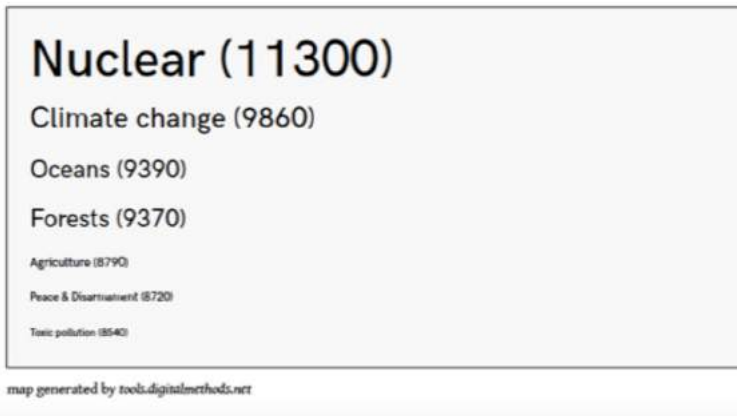


FIGURA 6 — Nuvem de questões problemáticas acerca da distribuição de interesse do *Greenpeace* com base nos resultados de contagem de páginas por campanha disponibilizados pelo *Google*, no site *Greenpeace.org*, em 2012. Análise feita por Anne Laurine Stadermann.

Posicionamento

Uma análise de posicionamento começa com a localização e colocação da escolha intencional de palavras-chave (ou questão de linguagem) por parte dos atores em relação a outras, no mesmo espaço (Williams, 1976). Em vez de um exercício em recepção, trata-se de detectar (e significativamente interpretar e delinear) inserções de questões feitas pelos atores, no espaço. Nas mídias sociais, quando a hashtag #blacklivesmatter é atrelada à #alllivesmatter, os atores injetam um contrapositionamento, ou, como descrito por Akrich e Latour (1992), um anti-programa no espaço de um programa. Poderia ser chamada de uma inserção patrocinada se fosse realizada por lobistas ou agentes políticos, ou poderia ser chamada de debate, se houvesse visualizações trocadas, mas muito trabalho de filtragem deve ser feito, a fim de desenterrar os engenhosos agentes, que trabalham no domínio subpolítico ou na organização da pilha de pequenos textos, os quais poderiam ser chamados de transcrições ordenadas de um debate.

Em vez de um trabalho de detetive ou de mapeamento de debate, o descrevermos como análise de posicionamento, e as questões a serem postas são: quem participa de um programa e quem participa de um antiprograma, com qual nuance e de qual lugar? O posicionamento como localização não é, portanto,

autônomo, mas topológico, no sentido geográfico. As coordenadas geográficas frequentemente acompanham as mídias digitais, e o conteúdo pode ser classificado (aparentemente) em simples termos geográficos, contudo, por mais que seja preciso identificar a localização (e visualizá-la), estão repletas de problemas de toponímia, densidade, modelagem e outros fatores complicadores.

Um exemplo que pode ilustrar tal análise de posicionamento se refere a quando a Suprema Corte dos EUA decidiu a favor do casamento entre pessoas do mesmo sexo, em 2015. No Instagram, bem como no Twitter e demais mídias sociais, poderíamos dizer que as hashtags #lovewins e #celebratepride e as contra-hashtags #jesuswins e #loveloses posiciona aqueles que as utilizam nesses espaços de questões societárias.

Alguns elogiam a decisão da corte com um programa, outros a ridicularizam com um antiprograma. Aqui, a análise crítica diz respeito à contabilidade de posicionamentos contra e a favor, mas tais oposições deveriam ser nuançadas porque há muitos antiprogramas e outras ações naquele espaço, tal como a propagação de *spams* por usuários, “trollagem” e redirecionamentos.

Aqueles que aderem aos programas e antiprogramas são frequentemente considerados como “públicos configurados por hashtags” e são descritos como aqueles que aderem ou deixam de aderir rapidamente (Bruns & Burgess, 2015; Rambukkana, 2015). Uma contribuição desses públicos, que demonstra a relação deles, é a memética, referente ao compartilhamento de conteúdos com características semelhantes em termos de forma e substância, feitas com conhecimento de outros conteúdos e que circulam no espaço (Shifman, 2013). Juntamente com a hashtag #jesuswins, vieram muitas imagens com cruzeiros e em tons avermelhados, que podem ser contrastadas com o uso abundante da imagem do arco-íris na hashtag #celebratepride (Figuras 7 e 8). Os memes e contramemes são disponibilizados pelos usos recorrentes de filtros, que são recursos originais do Instagram e sua fonte de popularidade.

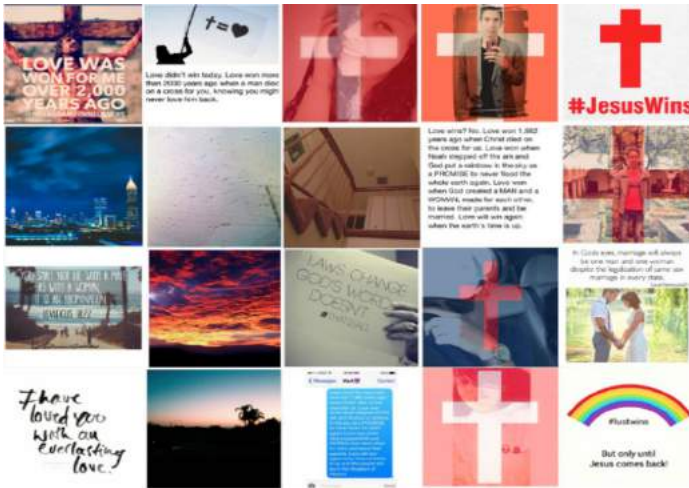
Ao trabalhar com as hashtags, podemos traçar o uso de #lovewins e #jesuswins (assim como as hashtags relacionadas a estas duas) no Instagram, onde a análise de hashtags confirmaria que os públicos agrupados em torno de hashtags nuançam seus posicionamentos em termos religiosos (#jesuswins com imagens da cruz) e contra a regulação do casamento entre pessoas do mesmo sexo, em que a hashtag #loveloses expressa sentimentos contra a validação desse tipo de casamento. Ao localizar (geograficamente) as postagens, notamos uma densa concentração de publicações em partes específicas dos EUA (incluindo a região do Cinturão Bíblico), embora com muito menos postagens do que #lovewins, que também ressoa fora dos EUA (Figura 9). Aqui a análise substantiva (por meio da

combinação de análise visual de hashtags e memes) e geograficamente posiciona as reações à decisão da Suprema Corte.



#celebratepride

FIGURA 7 — Representações visuais de filtros no Instagram depois a decisão da Suprema Corte dos EUA sobre o casamento de pessoas do mesmo sexo, em julho de 2015. Análise feita por Bastiaan Baccarne, Angeles Briones, Stefan Baack, Emily Maemura, Janna Joceli Omena, Peiqing Zhou e Humberto Ferreira, durante a *Digital Methods Summer School*, em 2015.



#jesuswins

FIGURA 8 — Representações visuais de filtros no Instagram depois a decisão da Suprema Corte dos EUA sobre o casamento de pessoas do mesmo sexo, em julho de 2015. Análise feita por Bastiaan Baccarne, Angeles Briones, Stefan Baack, Emily Maemura, Janna Joceli Omena, Peiqing Zhou e Humberto Ferreira, Durante a *Digital Methods Summer School*, em 2015.

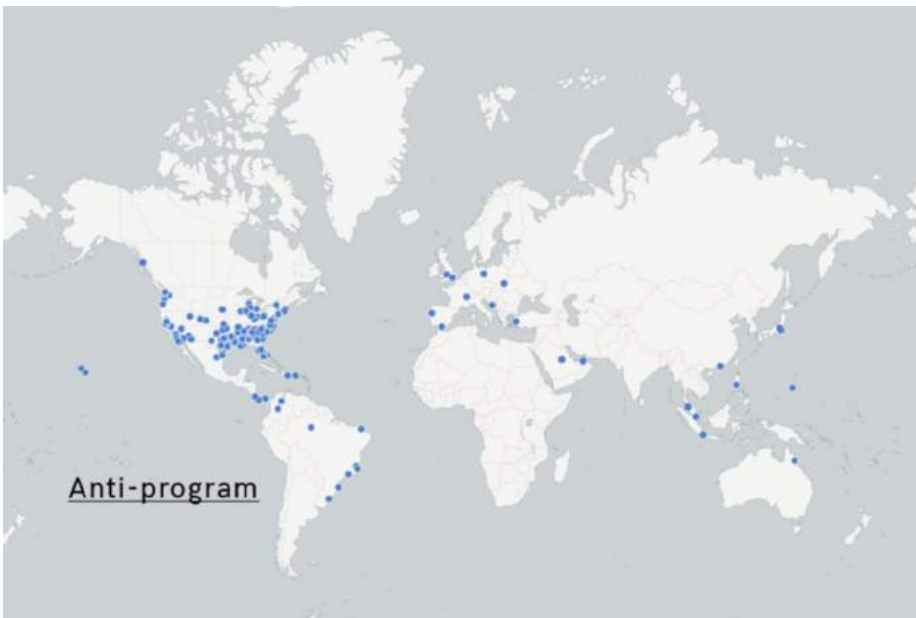


FIGURA 9 – Representações da geolocalização de usuários do Instagram de competição de hashtags (#celebratepride/#lovewins – programa – e #jesuswins/loveloses – antiprograma) após a decisão da Suprema Corte dos EUA sobre o casamento de pessoas do mesmo sexo, em julho de 2015. Análise feita por Bastiaan Baccarne, Angeles Briones, Stefan Baack, Emily Maemura, Janna Joceli Omena, Peiqing Zhou e Humberto Ferreira, Durante a Digital Methods Summer School, em 2015.

Alinhamento

O uso do termo “alinhamento” (e talvez a utilidade da análise crítica) é extraído da descrição feita por Walter Lippmann (1927) referente a como os públicos decidem de qual lado tomam partido em escândalos políticos. Em vez de se fazer estudos meticulosos acerca de uma questão social e das opções de enfrentamento delas nos negócios de fazer política, os públicos olham para os “sinais grosseiros de onde [suas] simpatias devem se transformar” (Lippmann, 1917, p. 64). Nessa leitura, os “sinais grosseiros” são palavras-chave ou termos específicos que os atores utilizam quando discutem uma questão. Quando múltiplos atores utilizam o mesmo idioma, ou quando os públicos também o fazem, eles se alinham. Devemos chamar a barreira entre Israel e os territórios palestinos uma cerca de segurança ou um muro de separação? Quando uma cerca de segurança é utilizada, esta é a posição de Israel, ao passo que, quando um muro de separação é usado, essa é a posição da Palestina. Outros termos poderiam ser analisados como medidas relativas à tomada de partido (muro de separação), esforços de reconciliação (cerca de separação) ou neutralidade (barreira) (Figura 10).

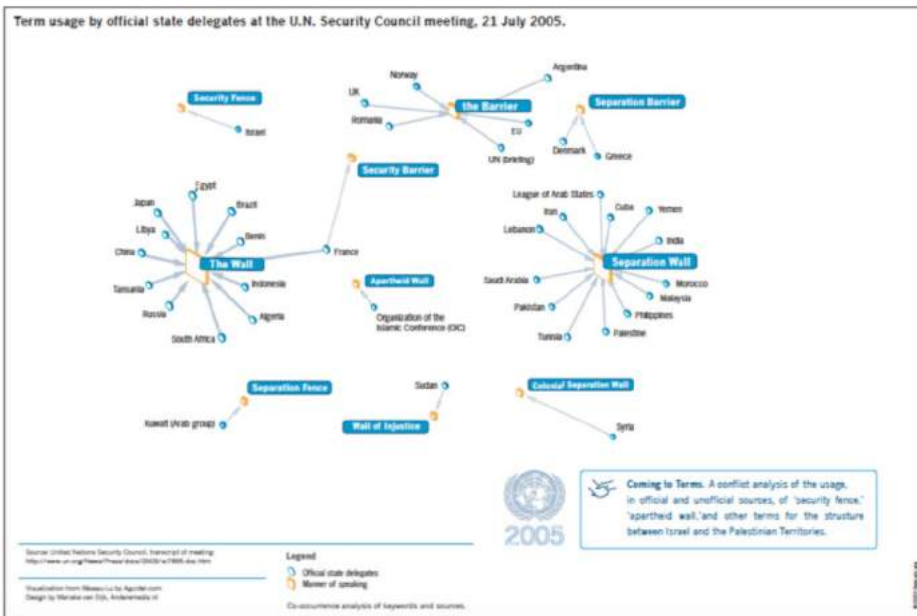


FIGURA 10 – Alinhamento de países no Conselho de Segurança da ONU, com base no uso de palavras-chave referentes à barreira entre Israel e territórios palestinos, 2005. Fonte: Rogers e Bem-David (2010).

O alinhamento se refere a forma como uma palavra-chave afeta o posicionamento. Quem mais utiliza o mesmo termo (ou palavra) para se referir a mesma questão e, portanto, associa-se de modo a estar em consonância com outros? O alinhamento se dá com base em terminologias ou palavras-chave compartilhadas. Seja em enunciados coletados pelas mídias ou reproduzidos do plenário parlamentar, os termos compartilhados se tornam os “sinais grosseiros” de Lippman. As análises apreenderiam sinais do alinhamento de atores. Como podemos considerar o estudo das correlações estabelecidas entre palavras-chave? Podemos perguntar quais ONGs (embora diferentes quanto às questões que as orientam) estão alinhadas quando se trata de discutir o acesso à Internet como direito humano, ou quais estados nacionais são a favor de medidas de austeridade na Europa (mesmo se discordam em resgates). Essas são análises de questões singulares com múltiplos atores, mas também podemos analisar múltiplas questões (palavras-chave) e seus atores, buscando por conjuntos de alinhamentos de atores.

O alinhamento de organizações (amplamente concebidas) pode ser ilustrado pelo tom político de “onde você permanece depende de onde você se assenta”, frase atribuída a Rufus Miles, um empregado do governo durante a administração de Truman nos EUA (Miles, 1978). Miles predisse, de modo significativo, como o ponto de vista de um colega mudaria abruptamente quando ele mudasse para outra agência federal. Assim, agências ou instituições devem estar, também, alinhadas com questões específicas.

CONCLUSÃO: MÉTRICAS ALTERNATIVAS PARA AS MÍDIAS SOCIAIS

Gostaria de sintetizar o argumento feito para um projeto alternativo de métricas para mídias sociais. Ao todo, eu faço cinco movimentos. Com base nas altimétricas para a ciência, que considero um projeto alternativo de métricas, proponho um outro projeto. Esta proposta é relativa às questões problemáticas sociais, ao invés de ser relativa às ciências. Para essa finalidade, conclamo a uma mudança nas redes em estudo por pesquisadores sociais, isto é, uma mudança da rede social (com suas métricas de vaidade) para uma rede de questões problemáticas. Essa mudança de redes possibilita focar nas oportunidades de métricas alternativas para o estudo do engajamento social em questões problemáticas, que eu nomeio “análise crítica”. Em uma aplicação de métodos digitais, que procura redirecionar dispositivos online e seus métodos de pesquisa, minha proposta aqui é reutilizar as pontuações de altimétricas e outras medidas de engajamento

para a pesquisa social (Rogers, 2013). A análise crítica — o projeto alternativo de métricas para as mídias sociais — busca mensurar o “engajamento de outra maneira” ou modos de engajamento (outros que não aqueles feitos com métricas de vaidade), como voz dominante, interesse, comprometimento, posicionamento e alinhamento.

Um dos *insights* mais originais do projeto de altimétricas na ciência diz respeito à importância das mídias sociais para a organização da atenção no novo trabalho. As novas métricas antecipariam o interesse acadêmico em artigos publicados ou republicados, e mesmo em projetos em desenvolvimento ou aqueles que estão em tópicos atuais de blogs. Isso concede à web, e mais especificamente às plataformas de mídias sociais, o *status* de um fluxo de dados quase em tempo real que não apenas anteciparia o interesse em novo trabalho, mas fornecer indicações de impacto antes que o tempo necessário para as citações impressas se acumulem e serem contadas (Thelwall et al., 2013). A análise crítica é igualmente considerada uma alternativa aos projetos de métricas existentes em mídias sociais, mas a relação entre as altimétricas e a análise de citações (*Web of Science*) é diferente daquela entre análise crítica e métricas de vaidade.

A análise crítica toma de empréstimo o *insight* de que há trabalho profissional sendo organizado e disseminado por meio de mídias sociais que não concernem, principalmente, à apresentação de si e à vaidade (embora esses aspectos existam, claro, tanto na ciência quanto no trabalho referente às questões sociais). Da mesma maneira que as altimétricas enxergam as redes de ciência nas mídias sociais, a análise crítica propõe enxergar redes de questões problemáticas. Por conseguinte, as redes sociais são produtivas não apenas para o *self* e aquilo que alguém gostaria de ser, mas para a análise de engajamento em torno de questões problemáticas.

A análise crítica também toma das práticas de mensuração de engajamento online, tal como o monitoramento de mídias, a ideia de que o engajamento nas mídias sociais é significativo e digno de ser mensurado para além do alcance da vaidade. O impacto dos atores nos espaços de questões, ou seja, a métrica de *voz dominante*, é um exercício relativo à identificação de atores e demarcação de fontes que poderia ser empreendido em ciência ou no monitoramento de mídias, onde os atores são autores ou líderes de opinião, e as fontes são revistas ou jornais renomados. O *interesse*, como redirecionamento da atenção por parte dos públicos, similarmente, pode ser estudado em termos de novos investimentos em inovação na ciência (Callon et al., 1983). O *comprometimento* poderia ser concebido como uma “mais que esperada” atenção para uma notícia ou um paradigma científico fora de moda, ainda que se escreva sobre isso depois do término

do ciclo de atenção (Downs, 1972). O *posicionamento* deve ser pensado como os termos específicos usados repetidas vezes em artigos noticiosos que sinalizam uma inclinação política ou política editorial no enquadramento de prestação de contas sobre um evento (Etman, 1991; Herman & Chomsky, 1988). O *alinhamento* poderia ser concebido como enquadres comuns em jornais, mostrando como os jornais se alinham a outros. Em outras palavras, cada uma das análises críticas poderia ser descrita como formas de redirecionamento de métricas, em ciência e em monitoramento de mídias, aplicadas às mídias sociais.

Também sustentamos aqui que o estudo do engajamento nas mídias evolui com o advento das mídias sociais. Para cada uma das análises críticas há combinações de atores e linguagem, com relações entre eles, que são delineadas substancialmente, topologicamente e estilisticamente, e onde as particularidades são próprias a cada meio pesquisado. Assim, há importantes hashtags, postagens geolocalizadas e fotos, bem como memes expressos por meio de filtros. Além disso, os resultados de análise compartilham uma linguagem visual com o meio do qual derivam. Há nuvens de palavras, mapas com marcadores de localização, grids de imagem e outras visualizações do meio. Desse modo, as análises e seus resultados, enquanto compartilham comprovações e recursos com métricas de análise de citações científicas e monitoramento de mídias, também contêm objetos nativos digitais (do próprio meio) cujos meios de mensuração são comumente referidos como analíticos. Não apenas a apresentação de uma alternativa às métricas de vaidade, mas também o estudo do engajamento em questões problemáticas oferece um atributo crítico.

REFERÊNCIAS

- Akrich, M., & Latour, B. (1992). A summary of a convenient vocabulary for the semiotics of human and nonhuman assemblies. In W. Bijker & J. Law (Eds.), *Shaping technology/building society: Studies in sociotechnical change* (pp. 259-264). Cambridge, MA: MIT Press.
- Barabasi, A.-L. (2002). *Linked: The new science of networks*. New York, NY: Plume.
- Baudrillard, J. (1990). *Cool memories* (C. Turner, Trans). London, UK: Verso.
- Beer, D. (2016). *Metric power*. Basingstoke, UK: Palgrave Macmillan.
- Boorstin, D. (1961). *The image: A guide to pseudo-events in America*. New York, NY: Harper Colophon Books.
- boyd, d., & Ellison, N. B. (2007). Social network sites: Definition, history, and scholarship. *Journal of Computer-Mediated Communication*, 13(1).
- Bruns, A., & Burgess, J. (2015). Twitter hashtags from ad hoc to calculated publics. In N. Rambukkana (Ed.), *Hashtag publics: The power and politics of discursive networks* (pp. 13-28). New York, NY: Peter Lang.

- Callon, M., Courtial, J. P., Turner, W., & Bauin, S. (1983). From translations to problematic networks: Na introduction to co-word analysis. *Social Science Information*, 22(2), 191-235.
- Crampton, J. (2010). *Mapping: A critical introduction to cartography and GIS*. Chichester, UK: Wiley-Blackwell.
- Downs, A. (1972). Up and down with ecology. *Public Interest*, 28, 38-50.
- Entman, R. (1991). Framing U.S. coverage of international news. *Journal of Communication*, 41(4), 6-27.
- Foucault, M. (1972). *The archaeology of knowledge and the discourse on language* (A. M. Sheridan Smith, Trans). New York, NY: Pantheon.
- Frank, T. (1997). *The conquest of cool: Business culture, counter-culture and the rise of hip consumerism*. Chicago, IL: University of Chicago Press.
- Freelon, D., McIlwain, C. D., & Clark, M. D. (2016). Quantifying the power and consequences of social media protest. *New Media & Society*. Advance online publication. doi:10.1177/1461444816676646
- Freeman, L. C. (1977). A set of measures of centrality based on betweenness. *Sociometry*, 40, 35-40.
- Gerlitz, C., & Helmond, A. (2013). The like economy: Social buttons and the data-intensive Web. *New Media & Society*, 15(8), 1348-1365.
- Gerlitz, C., & Lury, C. (2014). Social media and self-evaluating assemblages: On numbers, orderings and values. *Distinktion: Journal of Social Theory*, 15(2), 174-188.
- Gladwell, M. (2010, October 4). *Small Change: Why the revolution will not be tweeted*. The New Yorker. Recuperado de <https://www.newyorker.com/magazine/2010/10/04/small-change-malcolmgladwell>
- Goffman, E. (1956). *The presentation of self in everyday life* (Monograph no. 2). Edinburgh, UK: University of Edinburgh Social Sciences Research Center.
- Granovetter, M. (1973). The strength of weak ties. *American Journal of Sociology*, 78(6), 1360-1380.
- Herman, E. S., & Chomsky, N. (1988). *Manufacturing consent*. New York, NY: Pantheon.
- Karpf, D. (2017). *Analytic activism: Digital listening and the new political strategy*. New York, NY: Oxford University Press.
- Latour, B. (2008). *What is the style of matters of concern? Two lectures in empirical philosophy* (Spinoza Lectures). Assen, Netherlands: Van Gorcum.
- Lehrer, J. (2010, September 29). *Weak ties, Twitter and revolution*. Wired. Recuperado de <https://www.wired.com/2010/09/weak-ties-twitter-and-revolutions/>
- Lippmann, W. (1927). *The phantom public: A sequel to "Public Opinion."* New York, NY: Macmillan.
- Liu, A. (2004). *The laws of cool*. Chicago, IL: University of Chicago Press.
- Marres, N. (2015). Why map issues? On controversy as a digital method. *Science, Technology and Human Values*, 40(5), 655-686.
- Marwick, A., & boyd, d. (2011). To see and be seen: Celebrity practice on Twitter. *Convergence: The International Journal of Research Into New Media Technologies*, 17(2), 139-158.
- McGuigan, J. (2009). *Cool capitalism*. London, UK: Pluto Press.
- Miles, R. E. Jr. (1978). The origin and meaning of Miles' Law. *Public Administration Review*, 38(5), 399-403.
- Morozov, E. (2009, September 5). From slacktivism to activism. *Foreign Policy*. Recuperado de <http://foreignpolicy.com/2009/09/05/from-slacktivism-to-activism/>
- Pariser, E. (2011). *The filter bubble*. New York, NY: Penguin.
- Priem, J., Taraborelli, D., Groth, P., & Neylon, C. (2010, October 26). *Altmetrics: A manifesto*. Recuperado de <http://altmetrics.org/manifesto>
- Rambukkana, N. (2015). Introduction: Hashtags as technosocial events. In N. Rambukkana (Ed.), *Hashtag publics: The power and politics of discursive networks* (pp. 1-12). New York, NY: Peter Lang.

- Rao, A., Spasojevic, N., Li, Z., & Souza, T. (2015). Klout score: Measuring influence across multiple social networks. In *Proceedings of Big Data 2015* (pp. 2282-2289). New York: IEEE.
- Rao, A., Spasojevic, N., Li, Z., & Souza, T. (2016, January 20). *Klout score: Measuring influence across multiple social networks*. Recuperado de <http://www.slideshare.net/adithya.srao/klout-scoremeasuring-influence-across-multiple-social-networks>
- Ries, E. (2009, May 19). *Vanity metrics vs. actionable metrics* (Web log post). Recuperado de <http://fourhourworkweek.com/2009/05/19/vanity-metrics-vs-actionable-metrics/>
- Rifkin, J. (2000). *The age of access*. New York, NY: Putnam.
- Rogers, R. (2013). *Digital methods*. Cambridge, MA: MIT Press.
- Rogers, R., & Ben-David, A. (2010). Coming to terms: A conflict analysis of the usage, in official and unofficial sources, of “security fence,” “apartheid wall,” and other terms for the structure between Israel and the Palestinian Territories. *Media, Conflict and War*, 2(3), 202-229.
- Senft, T. (2013). Microcelebrity and the branded self. In J. Hartley, J. Burgess, & A. Bruns (Eds.), *A companion to new media dynamics* (pp. 346-354). Oxford, UK: Wiley-Blackwell.
- Shifman, L. (2014). *Memes in digital culture*. Cambridge, MA: MIT Press.
- Stevenson, S. (2012, April 24). What your Klout score really means. *Wired*. Recuperado de https://www.wired.com/2012/04/ff_klout/
- Suleiman, E. (2003). *Dismantling democratic states*. Princeton, NJ: Princeton University Press.
- Thelwall, M., Haustein, S., Larivière, V., & Sugimoto, C. R. (2013). Do altmetrics work? Twitter and ten other social Web services. *PLOS ONE*. doi:10.1371/journal.pone.0064841
- Tilly, C., & Wood, L. J. (2016). *Social movements 1768-2012*. London, UK: Routledge.
- Turow, J. (2006). *Niche envy*. Cambridge, MA: MIT Press.
- Van Doorn, N. (2014). The neoliberal subject of value: Measuring human capital in information economies. *Cultural Politics*, 10(3), 354-375.
- Van Es, K., & Schaefer, M. (2017). Introduction: Brave new world. In M. Schaefer & K. Van Es (Eds.), *The datafied society: Studying culture through data* (pp. 13-22). Amsterdam, Netherlands: Amsterdam University Press.
- Venturini, T. (2016). *Controversy mapping: A travel companion* (HDR dissertation). Ecole Normale Supérieure, Lyon, France.
- Williams, R. (1976). *Keywords: Vocabulary of culture and society*. London, UK: Fontana.
- Wortham, J. (2012, December 28). Digital diary: Facebook poke and the tedium of success theater. *The New York Times*. Recuperado de <https://bits.blogs.nytimes.com/2012/12/28/digital-diaryfacebook-poke-and-the-tedium-of-success-theater/>

*Tradução por Tiago Salgado.



TRÈS

REDES DIGITAIS E
AFFORDANCES VISUAIS



EXPLORAÇÃO VISUAL DE REDES PARA JORNALISTAS DE DADOS

TOMMASO VENTURINI
MATHIEU JACOMY
LILIANA BOUNEGRU
JONATHAN GRAY

Poucas pessoas sabem, tão bem quanto os jornalistas, que o mundo é feito de relações. Seguir alianças, desvelar conexões, desenrolar tópicos é, e por muito tempo tem sido, uma parte central de suas investigações. Se os cientistas sociais podem especular sobre estruturas duradouras e arranjos globais, os jornalistas não possuem tal prazer. O trabalho deles consiste em rastrear as associações específicas que conectam indivíduos e instituições para descobrir como somas de dinheiro, influência e conhecimento são trocadas por meio deles, e onde comportamentos antiéticos, corrupção, fraude e influência política injusta pode ocorrer. O advento de tecnologias digitais tornou esse trabalho tanto fácil quanto mais difícil. Mais fácil porque aumentou a rastreabilidade de associações políticas e econômicas. Mais difícil porque afundou os jornalistas em mais informações do que com as quais seus quites de ferramentas para investigações estão acostumados a lidar.

Quando, por exemplo, os repórteres do Consórcio Internacional de Jornalistas Investigativos (ICIJ) receberam 2,6 terabytes e 11,5 milhões de documentos compondo os então chamados *Panamá Papers*, eles certamente não puderam processá-los manualmente (Baruch & Vaudano, 2016). Percebiam que este não é apenas um problema relacionado ao *Big Data*. O embaraço com o vazamento não dizia respeito apenas ao tamanho dele, mas ao fato de que o interesse por ele estava nas conexões que estabeleceu entre indivíduos peculiares e paraísos fiscais específicos. A extração de figuras-chave por meio de agregação estatística ou modelos computacionais abstratos poderia perder o cerne de muitas das histórias que os

jornalistas estavam mais interessados em explorar. A investigação não poderia simplificar o conjunto de dados (*dataset*), mas tinha que explorar cada uma das conexões exposta por tal conjunto. Isso foi feito, entre outras formas, por meio de uma ferramenta chamada *Linkurious* (<http://linkurio.us>), cujo interesse é prove-niente menos de sua potência computacional do que da maneira pela qual possi-bilita aos seus usuários ver e seguir as conexões de uma rede.

O caso dos *Panamá Papers* é interessante, mas também o é de maneira iso-lada. Apesar do interesse de longa data, o uso de redes no jornalismo permanece, comparativamente, à margem (para uma visão geral dos usos emergentes de redes no jornalismo, conferir: Bounegru et al., 2016). As razões não são difíceis de serem imaginadas. Os grafos matemáticos demandam mais habilidade e esforço, são menos conhecidos do que as abordagens estatísticas tradicionais, e não são acompanhados do mesmo vocabulário facilmente acessível e publicamente reco-nhecido de temas visuais. Com toda a sua potência computacional, os gráficos matemáticos não se adequam às necessidades jornalísticas porque tendem a ser pouco claros tanto para repórteres quanto para leitores.

Neste capítulo, abordamos essa dificuldade ao sugerirmos uma técnica para a exploração visual de redes. Como tentaremos mostrar, quando corretamente per-formada, a representação visual de rede traduz algumas das mais importantes estruturas de grafo em variáveis gráficas (assim, auxilia o trabalho investigativo) e possibilita a interpretação de redes com convenções similares àquelas desen-volvidas por mapas geográficos (desse modo, legíveis para amplas audiências). Após termos introduzido as bases matemáticas e históricas de nossa abordagem, apresentaremos nossa técnica para a exploração visual de redes. Tomamos como exemplo a rede da esfera de informação francesa para ilustrar o trabalho recur-sivo de interpretação e categorização que possibilita ler a rede como um território organizado. A exploração visual de rede, que cresce proeminentemente entre os pesquisadores de métodos digitais para a pesquisa social e cultural, pode ser útil não apenas para o estudo de cenários midiáticos, mas também para praticantes do jornalismo digital que estão interessados em explorar e contar histórias com redes e dados relacionais.

LAYOUTS DIRECIONADOS POR FORÇA

Longe de ser meramente estática, a representação gráfica de redes tem um valor hermenêutico intrínseco. Se você já usou um mapa de transporte público, sabe como funciona. Tais mapas são muito distintos dos mapas de rodovias e cidades.

Não se trata apenas de os mapas de transporte serem mais simples (o nível de detalhes depende somente da resolução do mapa), a questão está no fato de representarem uma rede e não um território geográfico. Uma ilustração dessa diferença pode ser encontrada no famoso mapa do metrô de Londres, tal como desenhado por Harry Beck, em 1933. Depois de Beck redesenhar o mapa, ele se tornou uma rede de correspondências nas quais as estações estão posicionadas de acordo com a sua proximidade e conectividade relativas. O ganho em legibilidade é evidente, visto que a função do mapa de transporte não é situar estações no espaço urbano, mas situá-las umas em relação as outras, bem como ajudar os usuários a se moverem de uma para outra (um tipo de orientação que inegavelmente se assemelha a uma utilizada pelos tradicionais navegadores — conferir, por exemplo, Turnbull, 2000, pp. 133-165).



FIGURA 1 — Mapa do metrô de Londres antes (1920, acima) e após (1933, abaixo) o redesenho de Beck.

Outro exemplo de tal abordagem de mapeamento é oriundo de trabalhos iniciais em Análise de Redes Sociais (ARS) (Freeman, 2000). Jacob Moreno, fundador da ARS, em entrevista concedida ao *The New York Times*, em 1933, é claro a respeito da importância da visualização: “O processo de elaboração de gráficos – o sociograma – foi concebido pelos sociometristas, e é mais do que meramente um método de apresentação. Ele é, antes de tudo, um método de exploração.” Na entrevista, Moreno apresenta a análise de redes como a “nova geografia”. Mais importante que o título, no entanto, é a imagem que acompanha a entrevista, que representa as amizades entre os alunos da quarta série. O sociograma, apresentado pelas figuras, fortemente revela como a amizade não é igualmente distribuída na classe. É preciso apenas entender que os triângulos representam os meninos e os círculos representam as meninas para ver como as relações entre gêneros são desencorajadas nesta idade específica (ou ao menos a declaração de tal amizade). O artifício, claro, só funciona porque os nós não são posicionados aleatoriamente no espaço, mas de uma forma que minimiza a sobreposição de linhas. Nas palavras de Moreno (1953, p. 141): “Quanto menor o número de linhas cruzadas, tanto melhor será o sociograma”. Pelo fato de os triângulos serem empurrados de um lado, e os círculo de outro, é fácil identificar a existência de uma única conexão entre gênero.

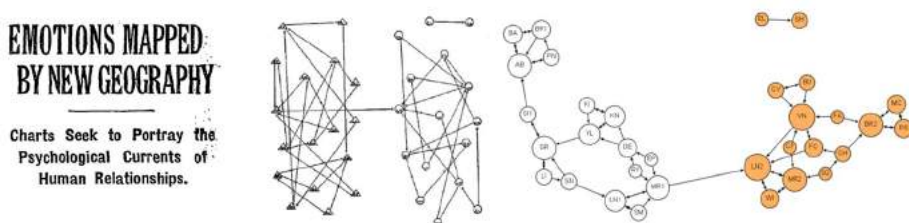


FIGURA 2 — Sociograma que representa a amizade entre alunos (o título e a imagem original acompanham a entrevista de Moreno ao *The New York Times*) — versão original à direita e versão moderna de espacialização direcionada à esquerda.

A regra de espacialização de Moreno é fácil de ser seguida em um grafo com dezenas de nós e arestas, mas impraticável em redes mais largas. Os grafos com milhares de nós e arestas são tão complexos que a contagem direta de cruzamentos de linhas se torna inviável, de tão demorada. Uma abordagem indireta consiste em atrair os nós conectados a fim de minimizar o comprimento das arestas e, sendo assim, a possibilidade de cruzamentos ou sobreposições. Mesmo neste caso, contudo, uma vez que cada nó pode estar conectado a vários outros, os

quais estão conectados a muitos outros, minimizar o comprimento das arestas está longe de ser um exercício trivial.

Nestas condições, podemos explorar a rede ao usarmos uma técnica chamada “especialização dirigida-por-força”. Tal especialização segue uma analogia física: os nós são carregados com uma força repulsiva que os separa, enquanto as arestas agem como molas, ao ligarem os nós que elas conectam. Quando o algoritmo é iniciado, ele altera a disposição dos nós até alcançar um equilíbrio entre tais forças (Jacomy et al., 2014). Esse balanceamento reduz o cruzamento de linhas e aprimora a legibilidade do grafo. Früchterman e Reingold (1991), que propuseram o primeiro algoritmo dirigido-por-força eficiente, mencionam o cruzamento de linhas como o segundo critério estético por eles proposto.

Apesar disso, os estudiosos que trabalham com redes logo perceberam que evitar o cruzamento de linhas não é o efeito mais interessante de layouts dirigidos-por-força. Em equilíbrio, a densidade visual de nós e arestas se torna uma aproximada, embora segura, representante da estrutura matemática do grafo (para uma prova matemática detalhada, conferir: Venturini et al., a ser publicado). Grupos de nós reunidos no layout tendem a corresponder aos clusters identificados por técnicas de detecção de comunidades em redes (Noack, 2009); buracos estruturais (Burt, 1995) tendem a se parecer com zonas esparsas; nós centrais se movem em direção a posições intermediárias; e pontes estão posicionadas, de algum modo, entre diferentes regiões (Jensen et al., 2015).

O artifício dos algoritmos dirigidos-por-força é ainda mais notável, dado que o espaço das redes é relativo, em vez de absoluto (ele pode ser rodado ou espelhado sem distorção de informação). Isso é consequência, e não uma condição do posicionamento do elemento. Em representações geográficas tradicionais, o espaço é definido *a priori* pelo modo como os eixos horizontal e vertical são construídos. Os pontos são projetados em tal espaço pré-existente de acordo com um conjunto de regras que atribuem uma posição unívoca a um par de coordenadas. O mesmo é válido para qualquer diagrama cartesiano (gráficos de dispersão, por exemplo), mas não para redes, nas quais o espaço é definido pela posição dos nós, e não o contrário.

Apesar dessas diferenças (que não devem ser esquecidas), algoritmos dirigidos-por-força possibilitam a leitura de redes como mapas geográficos ao traduzirem conceitos matemáticos complicados em um vocabulário mais convencional de regiões e margens, trajetórias e pontos de referência, centros e periferias (Lynch, 1960). Esta é uma vantagem crucial que explica por que algoritmos se tornaram o padrão da visualização de rede: eles facilitam a exploração de redes e as relações por meio de metáforas espaciais mais familiares e intuitivas, bem como por meio de métricas computacionais e estatísticas menos conhecidas.

O DÉCODEX: UM ESTUDO DE CASO CONTROVERSO

Nas próximas páginas, iremos ilustrar a técnica de exploração visual de rede por meio de um exemplo concreto. Nosso estudo de caso é uma rede de websites extraídos de uma lista compilada pelo jornal francês *Le Monde*. Desde 2009, um grupo de jornalistas, reunidos sob o nome de *Les Décodeurs* (Os decodificadores),¹ tem verificado a precisão de milhares de histórias que circulam na blogosfera francesa e nas mídias sociais. Em janeiro de 2017, ao início da campanha eleitoral francesa, o grupo lançou uma ferramenta online chamada *Décodex*,² que possibilita aos leitores pesquisarem sobre as mais importantes fontes de informação online relevantes para os debates públicos franceses (apesar de não necessariamente em francês). Cada fonte está acompanhada de uma breve descrição e, crucialmente, de uma avaliação de sua credibilidade, conforme os jornalistas do *Le Monde*.



FIGURA 3 — Interface de usuário da ferramenta Décodex do Le Monde.

1 Disponível em: <www.lemonde.fr/les-decodeurs/article/2014/02/12/l-equipe-des-decodeurs_4365082_4355770.html>.

2 Disponível em: <www.lemonde.fr/verification/>.

Não é surpreendente que a classificação fornecida pelos Decodificadores tenha suscitado muito debate nas esferas midiáticas francesas. Muitas das fontes categorizadas como imprecisas ou duvidosas, juntamente com outros jornais e blogs, contestaram o *Décodex*, com críticas que vão desde desafiar a maneira como websites são classificados de modo simplista, até questionar o direito do *Le Monde* (uma fonte rival de informação) de observar a confiabilidade de outros websites, bem como disputar a legitimidade e interesses de tal classificação, em geral (argumentam que alguns websites na lista circulam opiniões em vez de informação). Os Decodificadores admitem a dificuldade de seu trabalho, as muitas ambiguidades entre as quais eles são obrigados a escolher, bem como os erros e imprecisões que podem ter derivado delas. Ao mesmo tempo, eles defendem seu trabalho ao apontarem para a crescente quantidade de informações falsas ou partidárias que circulam online e ao afirmarem que se abrem para discutir as classificações que fazem, de modo a revisá-las, se necessário.

A controvérsia em torno do *Décodex* é um bom exemplo das dificuldades relacionadas à detecção de notícias falsas online (Bounegru et al., 2017), mas também dos debates mais gerais acerca de todo tipo de classificações. Categorizar coisas nunca é uma prática autoevidente ou inocente (Bowker & Star, 1999), e deveria sempre ser feita com o maior cuidado. Isso é verdade para a classificação inicial do *Décodex*, mas isso também é verdadeiro para a rede extraída dele. Como veremos nas próximas páginas, a exploração visual de redes envolve constantes idas e vindas referentes à categorização e à observação, à tipologia e à topologia.

Para construir nosso exemplo de rede, extraímos, em colaboração com os Decodificadores, todos os websites contidos no *Décodex*, e investigamos a maneira como eles citam um ao outro. Para fazer isso, recorremos ao *Hyphe*,³ um *crawleador* web desenvolvido pelo *médilab* da *SciencesPo*, que facilita a exploração de websites e segue os hiperlinks presentes nas páginas. Todos os websites que compõem o corpus *Décodex* foram *crawleados* até grau um de profundidade, referente ao clique iniciado na página inicial. Assim, obtivemos uma rede com 653 nós e 5.943 arestas. Enquanto os Decodificadores focam em juízos editoriais sobre como *classificar* websites no panorama das mídias francesas, nossa exploração de rede examina as relações entre eles e outros websites por meio de suas *práticas de link* (*linking practices*). Enquanto alguns pesquisadores focam em como as redes são mantidas unidas por meio de laços financeiros, afiliações organizacionais, relacionamentos comerciais e relações familiares e sociais, nós consideramos suas relações de acordo com o hiperlink, conforme uma longa tradição

3 Disponível em: <<http://hyphe.medialab.sciences-po.fr>>.

dos métodos digitais, da sociologia digital e da pesquisa sobre novas mídias (cf. Marres & Rogers, 2005; Rogers, 2013).

O tratamento de plataformas sociais — Facebook, Twitter, YouTube etc. — em nosso *crawler* requer uma explicação adicional. Tais plataformas são tanto fontes de informação como um todo quanto containers de múltiplas fontes individuais na forma de páginas ou contas. Uma vez que extrair todos os hiperlinks de um site tão amplo quanto o Facebook teria sido impossível, nós apenas *crawleamos* as contas que foram especificamente mencionadas no *DécodeX*. Contudo, mantivemos um arquivo com todos os links que apontavam para as principais plataformas de mídias sociais para investigarmos como elas são citadas por outros websites de nosso corpus.

UMA EXPLORAÇÃO VISUAL DA REDE DÉCODEX

A exploração visual de redes observa três variáveis visuais para representar graficamente suas características: posição, tamanho e matiz (para uma definição dessas variáveis e suas affordances semióticas, conferir: Bertin, 1967). Pelas razões discutidas antes, a posição é crucial na tradução das características matemáticas dos grafos. Layouts direcionados-por-força criam regiões onde vários nós são densamente agregados e regiões que são menos movimentadas. Essas diferenças de densidade, determinadas pela distribuição irregular de links, revelam a associação assimétrica entre as entidades da rede. Tudo pode estar conectado neste mundo, mas nem tudo está *igualmente* conectado.

Discernir a estrutura espacial das redes, todavia, nem sempre é simples. Nos casos mais fáceis, a diferença na densidade das associações é tal que os clusters aparecem como grupos de nós e arestas separados por zonas vazias (ou quase vazias). Essas zonas são chamadas de “buracos estruturais” (Burt, 1995) e, quando existem, oferecem uma orientação crucial para a interpretação da rede. Graças às rupturas criadas pelos buracos estruturais, as barreiras de clusters podem facilmente serem detectadas, como falésias que separam um planalto de um vale. Muitas das redes sociais ou naturais não exibem essa separação clara, e os limites de seus clusters tendem a ser graduais, com os declives das encostas. A imprecisão das fronteiras dos clusters não é necessariamente um obstáculo ao seu reconhecimento (pode-se apontar para uma colina mesmo quando é impossível dizer exatamente onde ela começa e termina), mas isso certamente torna a identificação deles mais difícil. Por esse motivo a análise visual de redes é mais parecida com uma expedição exploratória — onde significados e descobertas são

progressivamente e hermeneuticamente geradas – do que com a confirmação estatística de um conjunto de hipóteses pré-existentes (a respeito da diferença entre análises exploratórias e confirmatórias, conferir: Tuckey, 1997; Behrens e Chong-Ho, 2003).

Certamente este é o caso da nossa rede *Décodex*, que em um primeiro olhar não apresenta nenhum buraco estrutural manifesto ou qualquer estrutura espacial clara. Para visualizar nossa rede usamos duas ferramentas principais: o *Gephi*,⁴ para filtrar e espacializar a rede (usamos, em especial, o algoritmo dirigido-por-força *ForceAtlas2*), e o *Graph Recipes*,⁵ para ajustar a renderização visual da rede. Apesar de os buracos estruturais não serem evidentes na rede *Décodex*, olhar de modo mais próximo o layout, torna possível notar que a rede não se espacializa como um círculo perfeito, mas, ao contrário, como na forma de um abacate, com um topo menor e uma base maior. Tais irregularidades, tão fracas e sutis quanto possam parecer, frequentemente sugerem a presença de efeitos de polarização que podem interessar a pesquisas futuras.

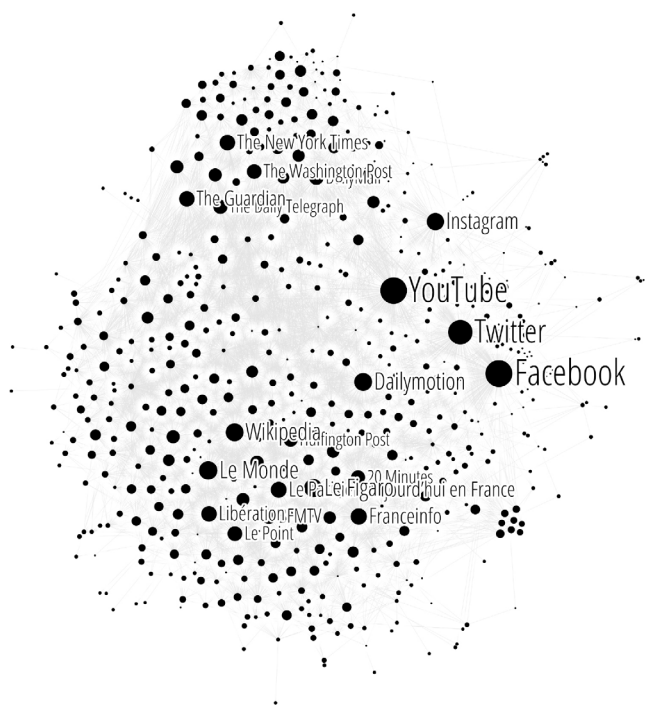


FIGURA 4 — Rede Décodex espacializada com o ForceAtlas2. O tamanho dos nós é proporcional ao grau de entrada.

4 Disponível em: <<https://gephi.org>>.

5 Disponível em: <<http://tools.medialab.sciences-po.fr/graph-recipes>>.

A primeira e a mais fundamental maneira de explorar nossa rede é olhar para a identidade dos nós que ocupam as diferentes regiões dela. Essa é uma distinta vantagem da própria exploração visual se comparada com outras formas de análise estatística, pois não agrega as entidades individuais que compõem seu corpus: cada e todo nó é visível no layout e pode ser questionado pelo pesquisador. Mesmo em uma rede pequena, como a do nosso exemplo, entretanto, a quantidade de nós pode dificultar (e tornar demorada) a observação de todos eles.

É aí que a segunda variável de nossa exploração visual (o tamanho) se torna útil. Posto que, nas redes, os nós são definidos primeiro e, acima de tudo, por suas conexões, eles foram ranqueados conforme o número de arestas que apontam para eles. No jargão da análise de redes, esse número é chamado de “grau de entrada” (*in-degree*), e os nós com um elevado grau de entrada são chamados “autoridade”, pois são reconhecidos e referenciados por muitos outros. Na figura anterior e em todas que seguem, dimensionamos os nós de acordo com o grau de entrada deles, de modo que uma grande autoridade literalmente é traduzida em maior destaque visual.

Ao lermos os nomes dos websites que ocupam os dois polos de nosso grafo em forma de abacate, parece natural supor que a separação deles deriva de uma fratura linguística. Os websites na parte mais baixa são predominantemente franceses, enquanto que aqueles na parte mais alta são mais internacionais. Uma maneira de realçar isso é mostrar a distribuição desigual do TLD (*Top Level Domain* – Domínio de topo ou Domínio de nível superior) na rede.

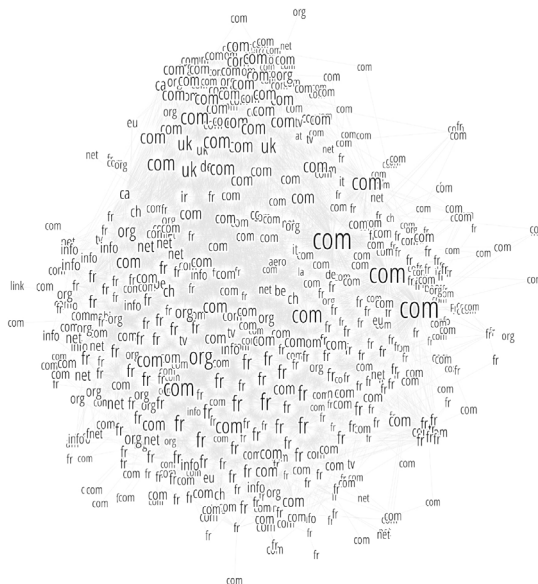


FIGURA 5 — Distribuição do *top level domains* na rede *Décodex*.

A separação linguística que acabamos de ressaltar, contudo, não é particularmente surpreendente ou interessante. Esse tipo de divisão é regularmente observado em rede de websites e hiperlinks. Detectá-lo é importante, mas ao contrário, de uma forma negativa – ele nos torna conscientes de que, para gerar descobertas mais interessantes, teremos de olhar *para além* dele.

Ao explorarmos ainda mais a rede, podemos notar o papel não apenas dos idiomas, mas também de plataformas de redes sociais, tais como YouTube, Facebook, Instagram e *Dailymotion*. Com a notável exceção da *Wikipedia*, todas as principais plataformas de mídias sociais estão localizadas ao centro e à direita do layout – em algum lugar entre os sites em inglês e francês (o que poderia ser esperado, tendo em vista a variedade de idiomas), mas também separadas de ambos por sua natureza distinta (e possivelmente pela maneira distinta como foram tratados no *crawl*).

Além disso, ao focar na parte inferior e maior da rede, podemos reconhecer dois subpolos diferentes, com fontes nacionais (tais como *Le Monde*, *Le Figaro*, *FranceInfo*, *Libération* etc.), que ocupam a maior parte da região inferior, e a imprensa regional, na parte inferior direita do layout.

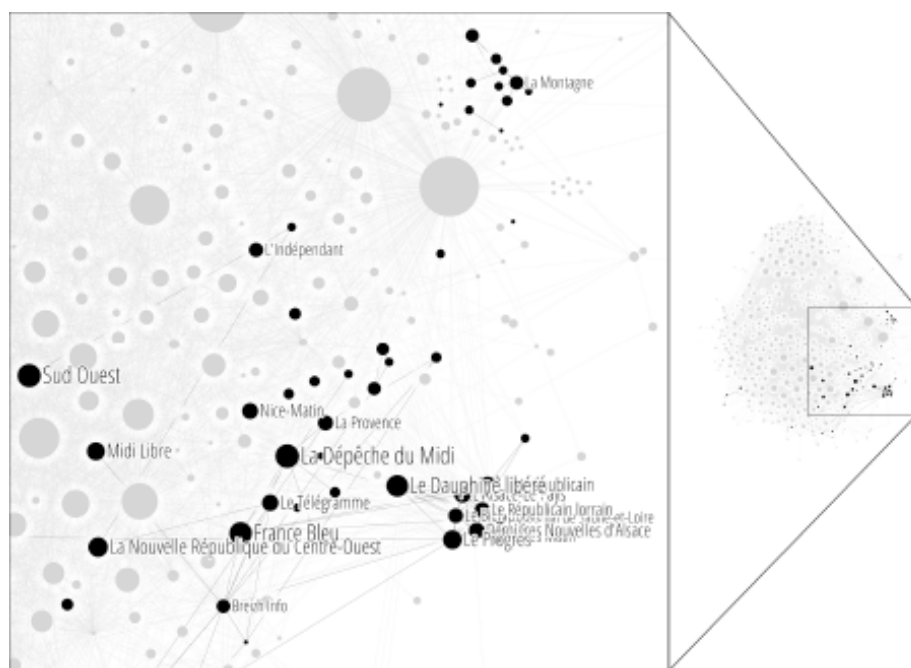


FIGURA 6 – Zoom na imprensa regional francesa.

A posição distinta das plataformas e da imprensa nacional/regional são dois achados interessantes e não triviais, mas podemos ir além com nossa análise. A maneira de fazer isso é jogarmos com a terceira variável visual da análise visual de redes: a matiz de um nó. Essa é uma parte trabalhosa, mas reveladora de nossa exploração visual. Ela consiste em categorizar os nós da rede conforme múltiplas classificações e visualizar essas classes, na rede, por diferentes cores ou, como neste trabalho, por diferentes tons de cinza. É importante notar que a classificação de nós e a leitura da disposição das classes não são operações separadas, mas realizadas ao mesmo tempo.

Como ficará claro, nossa técnica não consiste apenas na projeção de um conjunto de categorias pré-existentes em um layout baseado em conectividade, mas em utilizar recursivamente as categorias, de modo a conferir sentido ao layout e possibilitar que ele defina as categorias. É fundamental lembrar que a cor é uma variável visual que “não se mistura”. Um nó pode ser vermelho ou azul, por exemplo, mas não as duas cores ao mesmo tempo. Quando categorizamos nós, é necessário, portanto, utilizar categorias exclusivas. Um site, por exemplo, pode ser classificado na categoria “notícias” ou “sátira”, mas não nas duas. No caso (bastante comum) de nós que resistem a uma classificação única, os pesquisadores podem introduzir uma categoria residual como “múltiplo” ou “miscelânea”.

Como um primeiro passo em nossa exploração combinada de topologia e tipologia, iremos colorir os nós da rede conforme as categorias originais do *Décodex*. Essas categorias se referem à confiabilidade das fontes, manualmente avaliadas pelos jornalistas do *Le Monde* em quatro categorias: confiável, imprecisa, não confiável e satírica. Justamente porque tais categorias foram definidas antes e de modo independente da extração da rede, a disposição delas não segue a articulação espacial da rede. Certamente, é possível encontrar nós de todas as categorias em quase todas as regiões da rede. Uma exceção notável são os sites satíricos que são encontrados no lado direito do layout, tanto na parte superior quanto inferior. Indiscutivelmente, essa posição não se deve aos hiperlinks entre os sites satíricos, que não citam muito um ao outro, mas à forte conexão com as plataformas de mídias sociais, às quais todos esses sites estão amplamente conectados.

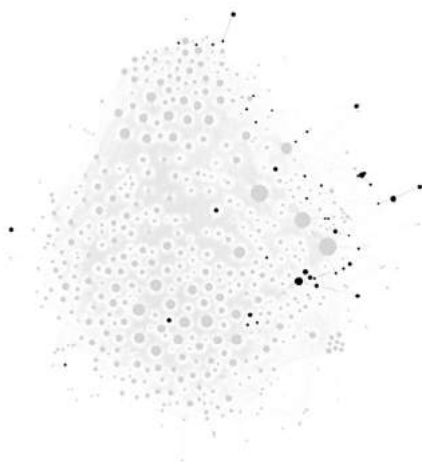


FIGURA 7 — Os sites “satíricos”, conforme a classificação original *Décodex* (o nó é enfatizado pela cor preta e pela duplicação de seu raio, apesar de seu baixo grau).

As outras classes são distribuídas mais uniformemente, mas não aleatoriamente. Os sites “confiáveis” tendem a ocupar o centro de ambos os polos (internacional e francês), enquanto os sites “imprecisos” e “pouco confiáveis” ocupam uma posição mais periférica. Curiosamente, ao olharmos para a parte inferior da rede, observamos dois grupos de fontes “imprecisas” e “pouco confiáveis”. Embora a maioria desses nós esteja posicionada acima dos principais sites nacionais e “confiáveis” (portanto, entre os sites franceses e internacionais), uma minoria significativa está localizada abaixo deles.

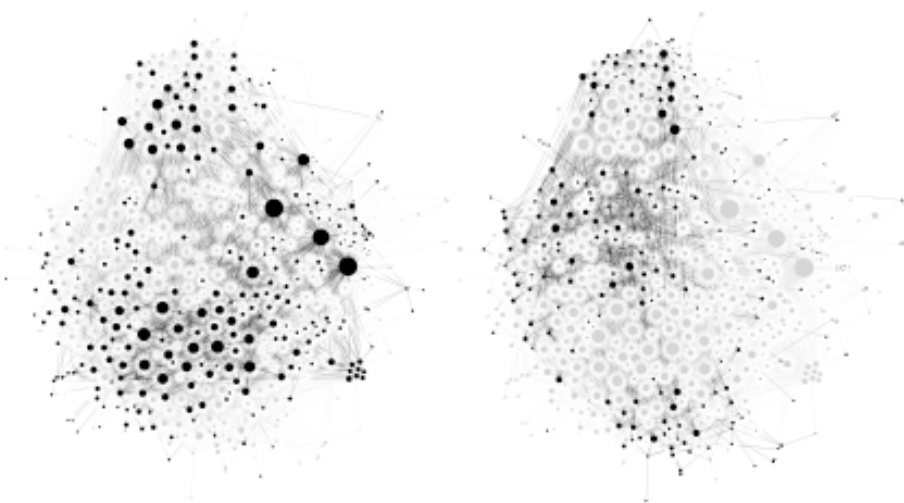


FIGURA 8 — Realce dos sites “confiáveis” (esquerda) e dos sites “não confiáveis” e “imprecisos” (direita).

Para explicar essa separação, introduzimos uma categorização adicional baseada na inclinação política dos sites. Em particular, diferenciamos os sites que disseminam informações não confiáveis ou imprecisas, pois eles seguem uma agenda de direita ou de extrema-direita (ocupam o centro da rede), e os sites que exibem uma atitude conspiratória mais geral (ocupam a parte inferior da rede).

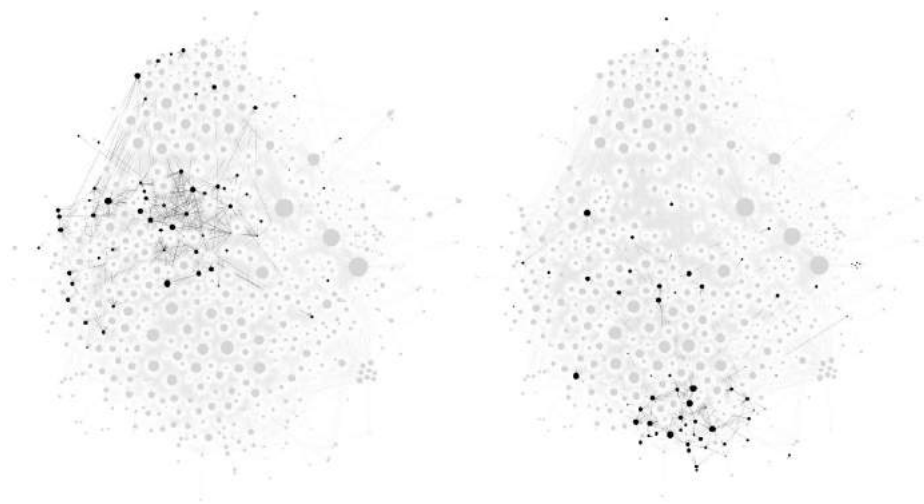


FIGURA 8 — Realce dos sites “conspiratórios” (esquerda) e dos sites de “direita” e “extrema-direita” (direita).

Por meio de nossa iterativa exploração da tipologia e da topologia, acabamos por revelar um seccionamento da rede que, embora invisível à primeira vista, possibilita-nos interpretar alguns dos principais contornos da paisagem midiática francesa. Apesar de esses territórios não estarem separados por buracos estruturais nítidos, os nós que eles contêm são bastante consistentes. Curiosamente, nossa classificação final produz uma divisão homogênea do layout, não apesar de sua heterogeneidade, mas por causa dela, que mistura categorias linguísticas, classes confiáveis e inclinações políticas. O fato de uma categorização não homogênea oferecer a melhor caracterização da estrutura de nossa rede deveria ser uma surpresa. As redes são objetos complexos que articulam diversos elementos por meio de lógicas distintas. Nesse sentido, nos lembramos de uma passagem de Jorge Luis Borges citada por Foucault como um perfeito exemplo de uma classificação heterogênea que, embora defina nossas categorias tradicionais, é altamente eficiente para descrever a cultura na qual foi elaborada:

Esse texto cita “uma certa enciclopédia chinesa” onde será escrito que “os animais se dividem em: a) pertencentes ao imperador, b) embalsamados, c) domesticados, d) leitões, e) sereias, f) fabulosos, g) cães em liberdade, h) incluídos na presente classificação, i) que se agitam como loucos, j) inumeráveis, k) desenhados com um pincel muito fino de pelo de camelo, l) et cetera, m) que acabam de quebrar a bilha, n) que de longe parecem moscas”. No deslumbramento dessa taxinomia, o que de súbito atingimos, o que, graças ao apólogo, nos é indicado como o encanto exótico de um outro pensamento, é o limite do nosso: a impossibilidade patente de pensar isso. (Foucault, [1970]/1999, p. IX).

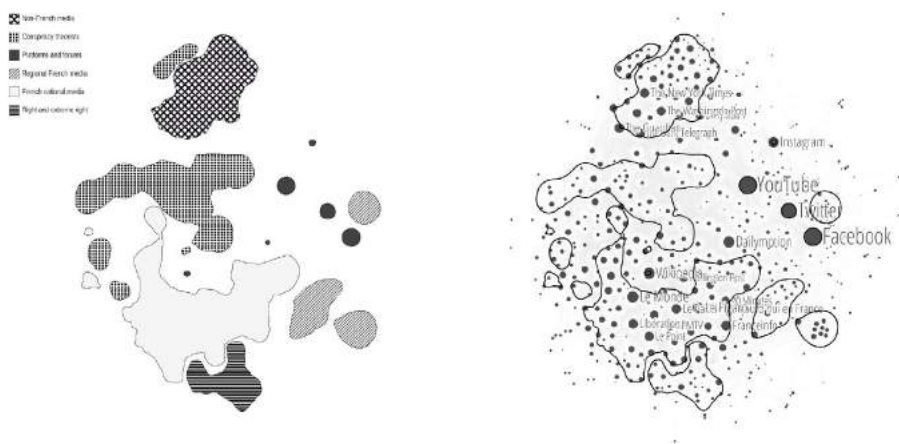


FIGURA 10 — Territórios heterogêneos da rede *DécodeX*.

CONEXÃO DE PADRÕES NA REDE DÉCODEX

Agora que definimos uma divisão heterogênea e hermenêuticamente robusta de nossa rede, por meio da exploração visual, podemos utilizá-la como base para nossa análise estatística. Ao passo que elogiávamos as vantagens da interpretação visual, também estávamos cientes de que nem todas as propriedades estruturais podem ser renderizadas visualmente. A direção das arestas ou a conexão entre diferentes classes, em particular, não são facilmente lidas em imagens de rede. Contudo, essas questões podem ser investigadas por outros meios, uma vez que a partição da rede tenha sido definida.

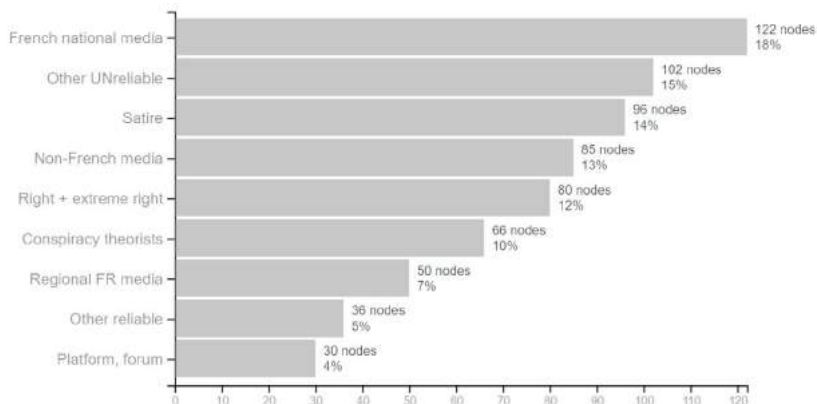


FIGURA 11 – Distribuição do número de nós por categoria.

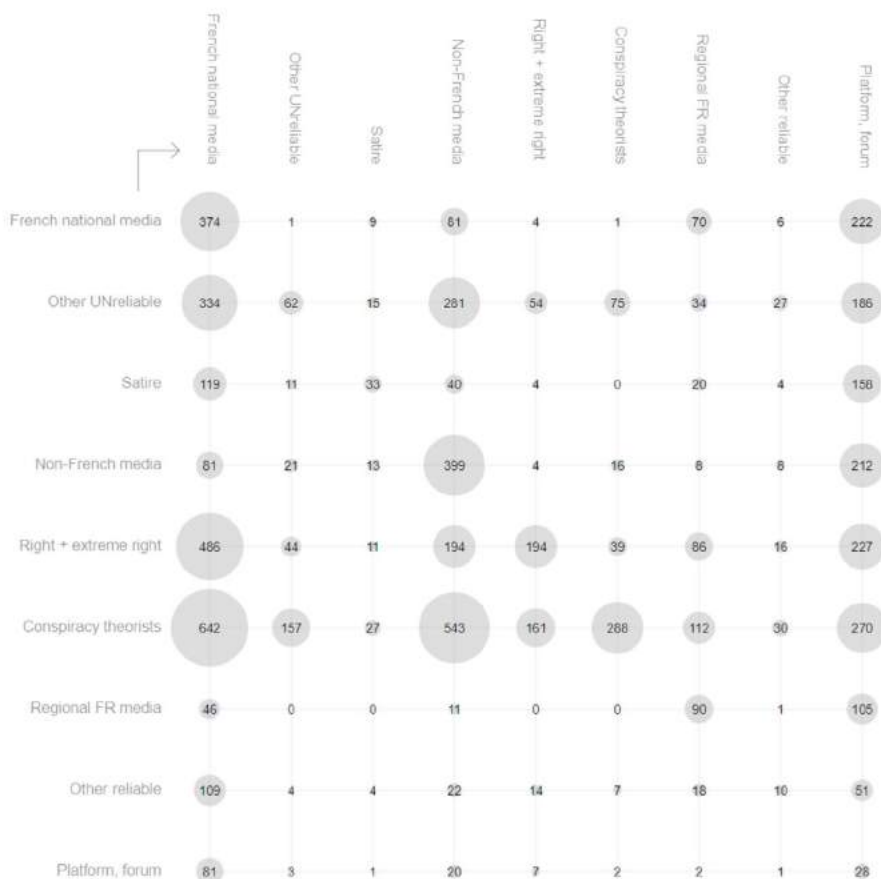


FIGURA 12 – Conectividade entre as categorias da nossa classificação final. As linhas indicam quantas vezes os nós de uma determinada categoria citam os nós de outras categorias. As colunas apontam quantas vezes os nós de uma determinada categoria são citados pelos nós de outras categorias.

A Figura 11 mostra a distribuição de nós em regiões identificadas em nossa classificação final (ver Figura 10), a qual adicionamos os sites “satíricos” (que discutimos antes, mas não incluímos na Figura 10 por uma questão de legibilidade), bem como “outros confiáveis” e “outros não confiáveis”. Essas duas categorias residuais compreendem, juntas, cerca de um quinto dos nós da rede. Este valor, relativamente alto, não é algo incomum. Dada a heterogeneidade das redes com as quais trabalham, cientistas sociais e jornalistas devem buscar classificações robustas e perspicazes (capazes de delinear zonas homogêneas no grafo) em vez de classificações abrangentes.

As nossas categorias empíricas são ferramentas poderosas para revelarem diferentes estratégias de conexão na rede. A Figura 12 apresenta os links no corpus, agregados por categorias. Como podemos ver, nem todas as categorias citam ou são citadas da mesma maneira. “Mídias nacionais francesas” e “plataformas” são amplamente citadas, e por muitos atores (suas colunas contêm círculos largos), enquanto sites “satíricos” são raramente citados (suas colunas estão quase vazias). Plataformas não são muito citadas, mas isso é meramente uma

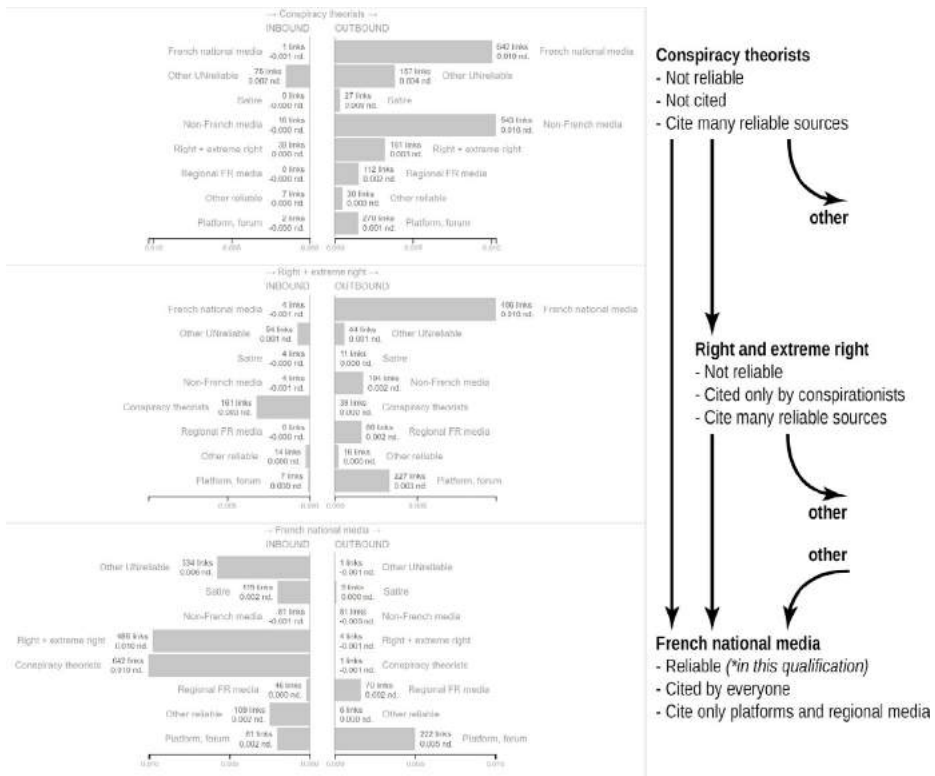


FIGURA 13 — Estrutura hierárquica no corpus, baseada em nossas categorias finais. As setas pretas no lado direito sintetizam a estrutura de links entre essas categorias.

consequência de nosso método, posto que muitas delas (como explicado anteriormente) não foram rastreadas (*crawled*). Sites de direita, de teorias da conspiração e outros sites “não confiáveis” são, ao contrário, a origem do maior número de citações e, de maneira muito interessante, eles parecem favorecer fontes “confiáveis” em detrimento daquelas “não confiáveis”. Como esperado, os sites confiáveis não se conectam de volta aos não confiáveis, e essa assimetria revela uma importante hierarquia. Para investigar esse padrão de conexão, iremos comparar os links de entrada e de saída das categorias mais interessantes.

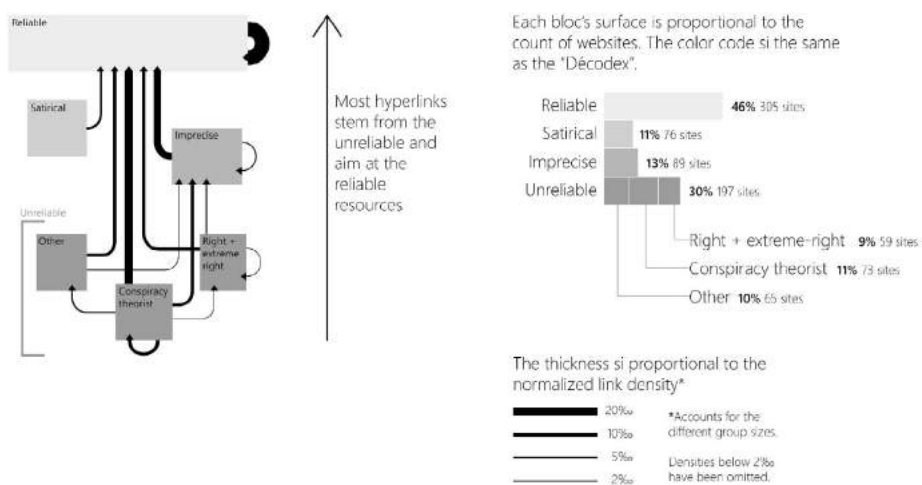


FIGURA 14 — Versão simplificada da análise estatística apresentada na Figura 13.

Esse tipo de estrutura hierárquica é comum na web e tem sido explicada como uma consequência da “anexação preferencial” (*preferential attachment*) (Barabási & Albert, 1999): atores tendem a se conectar a outros sites que eles percebem como mais elevados na hierarquia e evitam se ligar àqueles que percebem como inferiores. Tal estilo de anexação preferencial, por meio do qual os atores menores se vinculam aos atores da estrutura social, sem reciprocidade de vinculação, tem sido chamada de “vinculação almejada” (*aspirational linking*) (Rogers, 2013). Nem sempre os links de uma rede produzem uma hierarquia de categorias, mas este comportamento o faz. Este padrão de conexão e a maneira como ele se encaixa em nossas categorias empíricas pode sugerir uma maneira alternativa para a categorização da credibilidade que é investigada pelos Decodificadores: fontes confiáveis são citadas por todos os tipos de sites, embora fontes não confiáveis sejam citadas apenas por poucos outros (se alguns).

Esta observação está, em muitos aspectos, em desacordo com o que é frequentemente afirmado a respeito da “era da pós-verdade”, na qual supostamente aterrissamos. Enquanto as notícias falsas são usadas para alavancar a horizontalidade das mídias digitais a fim de borrarem os limites entre o verdadeiro e o falso, os padrões de conexão das esferas de informação (francesas) sugerem uma imagem diferente. Apesar de suas diferentes inclinações ideológicas, todos os sites concordam com a hierarquia geral de confiabilidade ao citarem em um sentido e não em outro. Os sites de “direita”, por exemplo, tentam borrar as linhas ao citarem seus pares e mais fontes confiáveis, mas também tentam traçar uma linha entre eles e os sites de “teorias da conspiração”, que são ainda menos confiáveis. Seja qual for a sua posição na pirâmide de hiperconexão, cada ator tenta melhorar a sua situação ao se ligar às autoridades acima, e não se ligando aos sites menos conceituados, que estão abaixo, o que reforça, assim, a hierarquia.

CONCLUSÃO

Este capítulo discutiu a exploração visual de redes com o intuito de aprimorar o entendimento de uma das formas dominantes de análise visual de nossa era digital – o diagrama de rede – e seu potencial papel em relação ao estudo e à prática do jornalismo digital. Inspirado na semiótica dos grafos e na cartografia tradicional, este capítulo propôs um modelo de acordo com o qual a interpretação da topologia de rede, com suas regiões, seus caminhos, núcleos e periferias, é guiada por três variáveis: posição, tamanho e matiz. O processo que descrevemos é aquele que enfatiza o caráter exploratório e iterativo da investigação. Embora contraintuitivo ao início, enfatizamos que, a fim de enfrentar as múltiplas lógicas desenvolvidas na estrutura de um grafo de rede, a análise não deve se limitar a um princípio de classificação. Vários critérios heterogêneos de classificação são frequentemente necessários para caracterizar a topologia de um mapa de rede. Finalmente, defendemos métodos mistos, complementando a exploração visual de redes com análises estatísticas, a fim de caracterizar ainda mais as propriedades da rede.

Por meio do estudo de caso do mapa de hiperlinks das mídias francesas, tentamos mostrar como a exploração visual de redes revela novos ângulos deixados de lado por outras análises. Neste caso, o presente capítulo ilustrou uma maneira alternativa de avaliar a confiabilidade dos sites, que complementa a tradicional abordagem de verificação de fatos e qualificação de conteúdos a partir da análise dos padrões de conexão entre diferentes regiões da rede como marcadores

de reputação (Rogers, 2013). Nesta análise, portanto, combinamos a classificação manual de confiabilidade realizada pelos jornalistas do *Le Monde* com a posição de uma fonte conforme os hiperlinks que nela entram e saem. Essa análise nos possibilitou trazer discussões frescas ao debate sobre notícias falsas. Apesar da proliferação de diversos conteúdos fabricados, as hierarquias de reputação na web parecem ser mantidas (ao menos até certo ponto), como a implementação de estilos de hiperconexão almejada por parte de falsos e hiperpartidários sites, a qual favorece, de forma surpreendente, talvez, as fontes autoritárias.

REFERÊNCIAS

- Barabási, A. L., & Albert, R. (1999). Emergence of scaling in random networks. *Science*, 286 (5439), 509. Recuperado de <http://www.sciencemag.org/cgi/content/abstract/sci;286/5439/509>
- Baruch, J., & Vaudano, M. (2016, April 8). « Panama papers » : un défi technique pour le journalisme de données. *Le Monde*. Paris. Recuperado de <http://data.blog.lemonde.fr/2016/04/08/panama-papers-un-defi-techniquepour-le-journalisme-de-donnees>
- Behrens, J. T., & Chong-Ho, Y. (2003). Exploratory Data Analysis. In I. B. Weiner (Ed.), *Handbook of Psychology* (pp. 33-64). London: Wiley. <http://doi.org/10.1002/0471264385.weio202>
- Bounegru, L., Gray, J., Venturini, T., & Mauri, M. (2017). *A Field Guide to Fake News*. Recuperado de fakenews.publicdatalab.org
- Bounegru, L., Venturini, T., Gray, J., & Jacomy, M. (2016). Narrating Networks: Exploring the Affordances of Networks as Storytelling Devices in Journalism. *Digital Journalism*
- Bowker, G. C., & Star, S. L. (1999). *Sorting Things Out: Classification and Its Consequences (Inside Technology S.)*. Cambridge MA: MIT Press.
- Burt, R. S. (1995). *Structural Holes: The Social Structure of Competition*. Cambridge MA: Harvard University Press. Recuperado de <http://books.google.com/books?id=E6vocVy8hVIC&pgis=1>
- Foucault, M. (1970). *The Order of Things*. New York: Pantheon Books.
- Foucault, M. (1999). *As palavras e as coisas: uma arqueologia das ciências humanas*. Trad. Salma Tannus Muchail. 8ª ed. São Paulo: Martins Fontes. (Coleção tópicos)
- Freeman, L. C. (2000). Visualizing Social Networks. *Journal of Social Structure*, 1 (1).
- Fruchterman, T. M., & Reingold, E. M. (1991). Graph drawing by force-directed placement. *Software: Practice and Experience*, 21 (NOVEMBER), 1129-1164. Recuperado de <http://onlinelibrary.wiley.com/doi/10.1002/spe.4380211102/abstract>
- Jacomy, M., Venturini, T., Heymann, S., & Bastian, M. (2014). ForceAtlas2, a Continuous Graph Layout Algorithm for Handy Network Visualization Designed for the Gephi Software. *PLoS One*, 9 (6), e98679. <http://doi.org/10.1371/journal.pone.0098679>
- Jensen, P., Morini, M., Karsai, M., Venturini, T., Vespignani, A., Jacomy, M., ... Fleury, E. (2015). Detecting global bridges in networks. *Journal of Complex Networks*, cnv022. <http://doi.org/10.1093/comnet/cnv022>
- Lynch, K. (1960). *The image of the city*. Cambridge MA: MIT Press. Recuperado de http://books.google.com/books?hl=it&lr=&id=_phRPWssPAGC&pgis=1

- Marres, N., & Rogers, R. (2005). Recipe for Tracing the Fate of Issues and Their Publics on the Web. In B. Latour and P. Weibel (Eds.), *Making Things Public: Atmospheres of Democracy* (pp. 922-935). Cambridge, MA: MIT Press.
- Moreno, J. (1953). *Who Shall Survive? (Second Edition)*. New York: Beacon House Inc.
- Noack, A. (2009). Modularity clustering is force-directed layout. *Physical Review E*, 79 (2). <http://doi.org/10.1103/PhysRevE.79.026102>
- Rogers, R. (2013). *Digital Methods*. Cambridge, MA: MIT Press
- The New York Times. (1933). Emotions Mapped by New Geography. *The New York Times*, 3 April.
- Tukey, J. W. (1977). *Exploratory Data Analysis*. Reading, MA: Addison-Wesley.
- Turnbull, D. (2000). *Masons, Tricksters and Cartographers*. London: Routledge.
- Venturini, T., Jacomy, M., & Jensen, P. (2019). What do we see when we look at networks. arXiv preprint arXiv:1905.02202. Recuperado de <https://arxiv.org/abs/1905.02202>

*Tradução por Tiago Salgado.

SISTEMA DE LEITURA DE REDES DIGITAIS MULTIPLATAFORMA

JANNA JOCELI OMENA
INÊS AMARAL

Tudo é rede! Esta antiga premissa acompanha o conceito de redes. Para desconhecidos, as redes são uma combinação de pontos ligados por linhas que podem representar processos de comunicação, e para especialistas, como o motor de busca Google, “um grupo ou sistema de pessoas e coisas interligadas”. Sempre que clicamos, vemos, avaliamos, gostamos, fazemos uma publicação ou descarregamos uma aplicação, tornamos as nossas ações quotidianas passíveis de serem lidas digitalmente. Estas ações digitais (também automatizadas) tornam-se rastreáveis e prontas para serem disponibilizadas em forma de rede, através da interface de programação de aplicações (API) ou através das *affordances*¹ de dados digitais. Neste sentido, as redes digitais são “representações de alguma coisa”, são processos técnico-sociais constituídos por ligações humanas e automatizadas (Latour *et al.*, 2012; Brandes, Freeman, & Wagner, 2014; Venturini, Munk, & Jacomy, 2019). Não surpreende que o estudo das redes digitais tenha vindo a ocupar um lugar proeminente nas práticas de Humanidades Digitais e Sociologia Digital. Então, na prática, como abordar estas redes? O que está em causa?

¹ Tania Bucher e Anne Helmond (2017) apresentam diferentes conceitos de *affordances*; da ontologia relacional proposta por James Gibson ao olhar atento às estruturas sociais que são formadas *por e através* da tecnologia (ver Baym, 2010; boyd, 2011; Ellison & Vitak, 2015). Assim como vemos a descrição do território semântico relacionado com o termo ‘plataforma’ proposto por Tarleton Gillespie em 2010 (computacional, arquitetónico, figurativo e político), Bucher e Helmond lembram-nos a rica e diversa história intelectual, *status* e valor analítico de *affordances*. Em conformidade com as autoras, o conceito de *affordance* neste capítulo faz referência a um pensamento ou raciocínio “muito mais relacional e composto por múltiplas camadas, por reter a sensibilidade da plataforma (*platform-sensitivity*) ao passo que considera também a especificidade dos media (*medium-specificity*)” (2017, p. 16).

A abordagem de análise multiplataforma serve com eficácia a prática de interpretar redes digitais, mas introduz alguma complexidade devido à gramatização da plataforma – processo de materialização técnica de atividade online (ver Gerlitz & Rieder, 2018; Omena, Rabello & Mintz, no prelo). Por um lado, disponibiliza perspectivas genéricas e específicas sobre o objeto de estudo; estas *affordances* revelam diferentes histórias visuais e textuais sobre o mesmo assunto ou problema de investigação (Rogers, 2017; Pearce *et al.*, 2018; Niederer e Colombo 2019). Por outro lado, possibilita uma reflexão sobre as qualidades e oportunidades de cada plataforma (Gillespie, 2010; Gibbs *et al.*, 2015; Rogers, 2017; Pearce *et al.*, 2018).

Diferentes plataformas sugerem distintas formas leitura de redes digitais. Assim sendo, perguntamo-nos: como combinar o conhecimento da gramatização da plataforma com a prática de recolha e análise de dados? Como explicar a espacialização da rede a partir dos *layouts* direcionados por força? Para responder a estas questões, propomos uma análise multiplataforma de redes digitais integrada numa abordagem do tipo *medium-specific*² (Rogers, 2013), em vez de nos focarmos na circulação dos conteúdos ou nos modelos de negócio e gestão das plataformas.

Neste capítulo, defende-se que a leitura multiplataforma de redes digitais pode advir de um triplo entendimento das seguintes fases, interligadas entre si: i) o exercício de mapeamento de disparidades na infraestruturas da multiplataforma digital; ii) a representação gráfica das redes digitais através de um algoritmo de *layout* direcionado por força, designadamente do algoritmo ForceAtlas2; e iii) do seu valor hermenêutico. Esta abordagem representa metáforas espaciais mais intuitivas e menos métricas computacionais ou estatísticas (Venturini *et al.*, 2018). Tem ainda em atenção a arquitetura *vernacular* das plataformas ou a “multicontextualidade” das interações homem-máquina (Gibbs *et al.*, 2015; Mares, 2017), e as múltiplas valências dos registos de ação das atividades online (Gillespie, 2010; Gerlitz, 2016; Highfield, 2018). No sentido de contextualizar a nossa abordagem, começamos por uma breve exposição teórica dos estudos de rede até às redes digitais. Em seguida, avançamos com uma proposta de um sistema de leitura de redes digitais multiplataforma.

DOS ESTUDOS DE REDES ÀS REDES DIGITAIS

Os estudos de redes têm como base os trabalhos de Jacob Moreno, George Simmel, John Arundel Barnes, entre outros. Em 1934, Moreno, o fundador da análise

² Construído do ponto de vista epistemológico (método).

de redes sociais, apresentou conceitos e instrumentos de sociometria que pretendiam registrar a observação e análise da interação social em pequenos grupos. No seu famoso estudo sobre os motivos da fuga de 14 adolescentes da Escola de Formação Feminina de Nova Iorque, expôs o comportamento humano através de sociogramas (representações gráficas compostas de vértices e linhas). Simmel (1955), como precursor da análise de redes, defende o primado das relações, fundamentado no argumento de que o mundo social resulta de interações, e não de uma agregação de indivíduos. O autor considera que a sociedade não é mais do que uma rede de relações e a interseção dessas relações assumidas como provedoras de estruturas sociais e unidades individuais. Por fim, mas igualmente importante, o antropólogo social Barnes (1954) desenvolveu um estudo pioneiro no qual, pela primeira vez, o termo “rede social” foi referido em contexto científico. A sua investigação pretendeu estudar a importância das interações individuais através de uma análise da organização social de uma comunidade piscatória norueguesa. Barnes concluiu que a maioria das ações individuais seria formada por laços familiares, de amizade e de conhecidos, e não pela pertença a um grupo particular³.

Em conformidade com estes exemplos clássicos, as redes sociais constituem-se como metáforas da sociabilidade ou como base para a compreensão da formação de interações individuais, do comportamento humano e de relações. Duas abordagens, que têm vindo a ser utilizadas por académicos de diferentes disciplinas para o estudo das redes digitais, destacam-se das restantes: a Análise de Redes Sociais (ARS)⁴ e a Teoria Ator-Rede (ANT). A primeira estabelece deduções sociais com base em fundamentos matemáticos – designadamente a teoria dos grafos. A ideia central é a de que as estruturas sociais resultam de um processo dinâmico (Giddens, 1987), partindo do pressuposto da complexidade do mundo relacional (Wellman, 1983). Os métodos e técnicas de análise de redes sociais impulsionaram um novo campo académico: a Ciência das Redes. Ao tentar interpretar a sociedade contemporânea como uma rede complexa (Watts, 2004) e sob a premissa de que tudo está interligado e nada acontece de forma isolada

3 Barnes concebeu esta estrutura partindo do princípio de que as interações constituem uma rede de relações, o que permite compreender os processos sociais fundamentais na formação e descrição de comunidades (ver Amaral, 2016).

4 Análise de Redes Sociais (ARS) é uma metodologia multidisciplinar que estuda as relações de um conjunto de atores com o objetivo de detetar padrões de interação, esclarecer porque é que estes acontecem e quais as suas consequências. O foco das abordagens da ARS está diretamente relacionado com a sua tipologia: redes ego – as conexões são rastreadas a partir de um nó em particular; ou redes globais – observam-se as ligações num grupo de atores previamente definido (Garton, Haythornthwaite, & Wellman, 1997).

(Barabási, 2003), a Ciência das Redes propõe-se a compreender “como tudo está interligado a tudo e o que isso significa para o comércio, para a ciência e para o dia a dia (Barabási, 2003).

Em contrapartida, a segunda abordagem rejeita as divisões clássicas da teoria social (Callon & Latour, 1981), os conceitos tradicionais da sociologia (Callon, 1986). AANT assume uma perspectiva sociotecnológica para defender que pessoas (humanos) e objetos (não humanos) assumem a figura de ator social, coabitando numa rede semiótica, interdependentes mas também autossuficientes, construindo uma estrutura de rede (Callon, 1986; Latour, 1987). Com o propósito de estudar “redes capturadas por dados digitais”, Venturini e colegas (2019) fazem uma crítica destas duas abordagens. Após um estudo exaustivo, concluem que a Análise de Redes Sociais é “um método sem uma teoria” e que a Teoria Ator-Rede é “uma teoria sem método”⁵. É um facto que as redes digitais nos trazem novas formas e formatos de compreender a sociedade e as nossas ações quotidianas⁶.

Os estudos de redes digitais sugerem “uma nova forma de rastrear a vida coletiva” ou mesmo “o próprio tecido da nossa existência” (Venturini, Jacomy, & Pereira, 2015; Venturini & Latour, 2010), onde participam conexões entre humanos, atividades automatizadas e objetos nativos digitais. Na mesma linha de pensamento, entendemos as redes digitais como esquemas representativos de conexões/ações rastreáveis/recuperáveis, formadas pela mobilização de atores humanos e não-humanos em torno de uma questão ou tema e mediada por interfaces técnicas. Neste sentido, nas redes digitais, as interfaces técnicas e os softwares de extração assumem-se como organizadores e mediadores entre o processo de recolha e a análise. Assim sendo, entendemos que as redes digitais podem emergir de duas formas: i) redes proporcionadas por interfaces técnicas (e.g. APIs) – como as plataformas armazenam, renderizam e disponibilizam traços digitais em forma de rede (e.g. uma rede de likes composta pelas conexões entre Páginas do Facebook; uma rede utilizador-fonte no Twitter); ii) redes criadas a partir das *affordances* de dados digitais e software – não necessariamente um esquema representativo das conexões organizadas e distribuídas via interfaces técnicas, mas algo que podemos construir (e.g. uma rede de coterminos do Instagram gerada a partir das legendas vinculadas às imagens ou vídeos publicados; uma rede bipartida de imagens e *labels* facultada por *Vision APIs* e

⁵ Para uma noção mais aprofundada destas abordagens, consultar Venturini e colegas (2019). Os autores apresentam uma perspectiva geral seguida de uma análise crítica das características principais e das diferenças entre os três tipos de abordagens: Ator-Rede, análise de redes e redes digitais.

⁶ O sociólogo Bruno Latour refere-se às redes digitais como parte de uma nova teoria social, em contraste com uma “velha teoria social, nascida de dados descontinuados” (2010, p. 161).

produzida a partir de uma lista de URLs que aponta para a localização das imagens na web). Consequentemente, argumentamos que uma análise minuciosa de redes digitais apela à utilização de métodos sociais específicos que vão além do que é definido pela ARS ou pela ANT. Neste capítulo mostramos como a natureza relacional dos dados digitais em conjunto com o meio (*medium*) (Rogers, 2013) e as *affordances* dos algoritmos de espacialização (Venturini, Jacomy, & Pereira, 2015; Venturini, Jacomy, Bounegru, & Gray, 2018) devem ser tidas em consideração na abordagem e leitura das redes digitais.

(i) Mapeamento de disparidades

O primeiro passo para a leitura de redes digitais multiplataforma é compreender a infraestrutura digital e os dados prontos dos media sociais. Isto remete para uma investigação do tipo *medium-research*, que pressupõe que se tenha um domínio da *maneira de ser* dos dispositivos técnicos e de ter noção do que implica uma “mecanização da metodologia” (Rieder e Röhle, 2012). A título de exemplo, consideremos as interfaces técnicas de plataformas web, mais especificamente as APIs, e a forma como estas lidam com os dados digitais. As APIs impõem, organizam, recolhem, renderizam e disponibilizam atividade digital para recolha e análise de dados digitais, dentre outras funcionalidades (Murugesan, 2007; Berlind, 2015). Consideradas as principais impulsionadoras da *web como plataforma*⁷ (Murugesan, 2007; Helmond, 2015) e reconhecidas como vias de acesso comuns à investigação digital, as APIs de media social continuam ao serviço da investigação digital: seja como fonte e método de recolha de dados ou como forma de criticar os regimes de acesso a dados ao longo do tempo. Apesar da emergência de um movimento que procura restringir o acesso à informação via APIs, estas interfaces técnicas deveriam servir como fonte de conhecimento para os estudos de

⁷ Esta questão remete à infraestrutura das APIs que responde ao princípio da ocultação de informação (*information hiding*) ou aos critérios aplicados para dividir o sistema em módulos (modularidade), proposto por David Lorge Parnas em 1972 (de Souza *et al.*, 2004). Sob esta organização, as APIs têm propriedades públicas e não públicas (privadas), separando a função da implementação. As propriedades públicas são visíveis para o cliente e devem incluir especificações sobre a funcionalidade; por outro lado, propriedades não públicas devem ser secretas porque encerram a implementação dessa funcionalidade (Meyer, 1988). Isto significa que os módulos de software assentam no princípio *open-closed*: são simultaneamente “abertos” (para extensão e adaptação) e “fechados” (para evitar modificações que afetem os clientes) (Meyer, 1988, p. 52). As APIs de *media social* funcionam na mesma lógica.

redes⁸. Contudo, no processo de leitura das redes digitais, raramente se tem em consideração a forma como as APIs operaram e organizam atividade online. Os investigadores recorrem às APIs como simples fontes de dados, negligenciando as camadas de informação técnica existentes que podem interferir no processo da investigação digital.

As interfaces técnicas (APIs) têm duas formas de comunicação interligadas, uma com os utilizadores (*front-end*) e a outra com software (*back-end*). O que podemos fazer na interface *front-end* (e.g. gostar, partilhar, carregar imagens) é armazenado, organizado e parcialmente disponibilizado no *back-end*. Entretanto, o que for modificado na interface *back-end* reflete-se naquilo que podemos fazer, ver e pesquisar na zona *front-end*. Um exemplo disso é a criação das cinco reações no Facebook, além do *like*; *amor*, *tristeza*, *surpresa* (wow), *raiva* e *risos* (ahah). Daqui decorre que as APIs permitem que a atividade seja discreta e reproduzida. É a este fenómeno que o filósofo francês de tecnologia Bernard Stiegler (2006, 2012) designa de processo de gramatização digital: a produção e “discretização” de estruturas de comportamento; ou seja, qualquer processo de materialização técnica. Nesta linha de pensamento, e na sequência do trabalho de Philip E. Agre’s (1994) sobre sistemas de vigilância e captura, as *gramáticas de ação* nas plataformas digitais adaptar-se-iam às atividades rastreáveis/recuperáveis disponibilizadas pelas interfaces técnicas; formas representativas de “discurso-feito-maquínico” (*discourse-made-machinery*) (ver também Gerlitz, 2016; Gerlitz & Rieder, 2018; Omena, Rabello & Mintz, no prelo).

As redes digitais exibem ações ator-software-plataforma e as suas gramáticas *multivalentes* (e.g. identificador de objetos digitais, hashtags, likes, sistemas de recomendações, reconhecimento automatizado de imagens); não só apresentam diferentes valores para diferentes *stakeholders* (Gillespie, 2010; Gerlitz, 2016; Marres, 2017), como também proporcionam diferentes redes e formas de rastrear a vida coletiva em tempo real. Por isso, os investigadores devem procurar dar resposta à forma como as interfaces técnicas lidam com a gramatização digital: quais são os pontos de entrada para recolher dados? Quanto tempo atrás os dados podem ser recuperados? Como as conexões são feitas? Do mesmo modo, quais as possíveis características da rede: direcionada ou não direcionada? Monopartida ou bipartida? Que gramáticas (ou ações) ligam nós e arestas? O que pode ser medido? Para além de conhecimento na área que está a ser objeto de estudo (e.g. alterações climáticas, ativismo político, saúde), o processo de interpretação

⁸ Para uma discussão mais crítica sobre *API research*, consultar o capítulo proposto por Venturini e Rogers em 2019, intitulado: “API-based research” or how can digital sociology and journalism studies learn from the Cambridge Analytica affair.

das redes digitais exige, não só literacia tecnológica (Rieder & Röhle, 2017) mas uma *prática técnica crítica* (Agre, 1997).

O segundo passo para a leitura de redes digitais multiplataforma consiste em combinar a gramatização da plataforma com a *praxis* de captura, análise e visualização de dados. É fundamental ter em consideração os diferentes tipos de redes — facultadas por interfaces técnicas ou pelas *affordances* dos dados digitais e software — e perceber que diferentes plataformas demandam distintas formas de leitura de redes digitais. Da mesma forma, a infraestrutura da plataforma intervém na formação da rede, já que as estruturas de rede que acontecem num determinado contexto social respondem também a diferentes gramáticas, conexões e significados gerados pelas plataformas. Por exemplo, a forma como se analisa e interpreta redes de hashtags não pode ser reproduzida em contexto de rede *page-like* no Facebook. Esta remete para o ato de fazer *likes* em outras páginas e receber *likes* de volta — as conexões podem exprimir interesse, sentimento de identificação ou apoio. Mas nas redes de coocorrência de hashtags as conexões observadas são indiretas, estas permitem explorar associações, termos relacionados entre si e conteúdo (visual e textual) em torno de um assunto ou problema de investigação.

O passo final para a interpretação de redes digitais multiplataforma compreende um processo que acontece antes e durante a análise da rede. A este propósito, devemos considerar como nós e arestas de uma determinada plataforma se relacionam com as métricas baseadas em plataforma (ou métricas predefinidas) e as *affordances* dos softwares de visualização e análise (ver Figura 1). Isto leva-nos a questionar sobre se devemos considerar a gramatização da plataforma ou as *affordances* do Gephi, e também quando e como combinar estas a fim de avaliar conexões, posicionamentos e atributos dos nós. Por exemplo, quando se analisa os sistemas de recomendações de *App Stores* em redes digitais, deve-se ignorar o grau (total de ligações incidentes num nó), já que evidenciaria as aplicações mais recomendadas pela plataforma e não necessariamente a sua lógica de recomendação algorítmica⁹. Na Figura 1 apresentamos uma discussão mais detalhada sobre estas questões.

⁹ Ver trabalho desenvolvido sobre “Objectionable Queries: searching for Porn in App Stores” disponível em: <https://digitalmethods.net/Dmi/SummerSchool2018AppStoresBiasObjectionableQueries>



FIGURA 1 – Processo de análise de redes digitais: o que se deve ter em conta?

Paralelamente, há camadas de conhecimento técnico associadas ao método de análise de redes digitais multiplataforma (ver Figura 2). Que tipos de redes emergem de diferentes plataformas? Quais as características das redes? O que pode ser estudado? Que histórias contam estas plataformas? Por exemplo, no Twitter, Instagram ou Tumblr é possível examinar redes de hashtags bem como redes de utilizador-post/tweet, utilizador-imagem ou link-imagem. Em redes de co-hashtag, observamos ligações indiretas em que o posicionamento de grupos (e.g. apoiantes e não apoiantes de uma causa social)¹⁰ e a sua escolha de palavras podem ser analisados – a partir do espaço semântico visualizado. Por outro lado, nas redes utilizador-post/tweet, o foco do estudo pode ser o fluxo de informação ou o conjunto de reações coletivas e respetivas respostas a um particular assunto/problema. No YouTube e lojas de apps, a ênfase pode ser o mapeamento de vídeos (ou apps) relacionados ou a observação do sistema de recomendações destas plataformas representado em rede. Por fim, ao construir uma rede a partir de Vision APIs, tomamos em consideração alguns estudos exploratórios, como o estudo de circulação de imagem (ver Omena, 2018; d’Andrea & Mintz, 2019; Omena, Chao, Pilipets *et al.* 2019). Fazemos ainda o exercício de reclassificar os resultados derivados da visão computacional (*relabelling the machine*); o espaço semântico das redes viabilizadas através de aprendizagem da máquina (*machine learning*) necessita de uma leitura crítica sobre a sua capacidade de descrição automatizada de

¹⁰ A deteção de *clusters* é uma técnica amplamente aplicada nas redes de co-tags. Permite revelar vários grupos sociais que utilizaram a mesma tag mas que visual e textualmente se expressam de forma diferente – por exemplo, vegans, artistas ou ativistas.

conteúdo visual (ver também Mintz *et al.*, 2019; Silva, 2019). Estudamos também as relações entre os atores (e.g. utilizador, link domains, *web entities*) e imagens (ver Omena, Chao, Pilipets *et al.* 2019), bem como a prática crítica de interrogar a aprendizagem da máquina e os seus vieses.

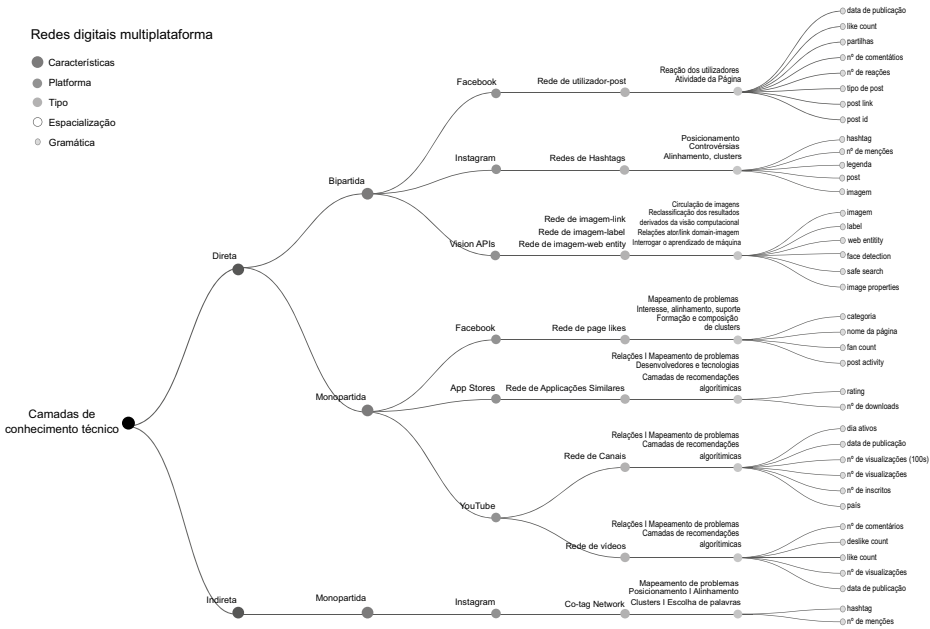


FIGURA 2 – Análise de redes digitais multiplataforma: camadas de conhecimento técnico.

Antes de iniciarmos a interpretação daquilo que observamos nas redes digitais, ou seja, o que de facto vemos na sua representação gráfica, é necessário fazermos um exame à gramatização da plataforma e *affordances* de software (utilizado para extração e análise de dados), ou seja, um mapeamento de disparidades. Aqui, as camadas de conhecimento técnico são relevantes.

(ii) A representação gráfica das redes

Uma rede é definida como um conjunto de nós com laços entre eles que resulta num fluxo de informação direto ou indireto. As redes indiretas são assimétricas quando têm conexões unilaterais, que podem ou não ser recíprocas (e.g. redes de *tweets*). As redes diretas são simétricas (e.g. redes de *page like*), o que implica uma

ligação intencional entre os nós. A representação visual das redes é feita através de grafos que podem ser monopartidos (um tipo de nó) ou bipartidos (dois tipos de nós). Como esquema representativo de uma rede, um grafo pode revelar padrões de relacionamento ou quantificar relevantes propriedades estruturais. Na Figura 2 demonstrámos as características das redes digitais em função de diferentes plataformas.

Os padrões das relações formam um espaço social que pode ser mapeado e analisado pela perspetiva relacional dos dados (Wasserman & Faust, 1994). As propriedades estruturais e topológicas das redes (grau, diâmetro, proximidade, afastamento, *pagerank*, modularidade, *hubs*) remetem diretamente para os seus elementos (nós e arestas) e para as suas dinâmicas (densidade, distância geodésica, densidade, centralização, coeficiente de *clustering*). Estas propriedades topológicas fomentam a análise das dinâmicas sociais e relacionais das redes. Diferentes softwares de análise de redes (e.g. Gephi, NodeXL Pro ou Network Workbench) utilizam diferentes métricas para estudar as propriedades dos sistemas. Contudo, as definições de uma rede quando ajustadas aos ambientes das plataformas digitais, devem ter em conta a sua infraestrutura e gramatização digital, que atribuem características específicas às redes digitais.

Para a representação espacial de redes recorreremos aos mecanismos dos algoritmos direcionados por força (Fruchterman & Reingold, 1991; Bannister *et al.*, 2012; Jacomy *et al.*, 2014). Estes algoritmos são largamente utilizados para a análise de media sociais e outras plataformas. Venturini *et al.* (2018) salientam a diferença entre a representação tradicional geográfica e a “espacialização direcionada por força” operacionalizadas para leitura e interpretação de redes. No primeiro caso, o espaço “é definido *a priori* pela posição como os eixos horizontal e vertical são desenhados. Os vértices projetam-se num espaço pré-existente de acordo com um conjunto de regras que atribuem uma posição universal a um par de coordenadas” (Venturini *et al.*, 2018, p. 268). No segundo caso, o espaço é “uma consequência, e não uma condição para posicionamento de um elemento”; “é relativo e não absoluto (pode ser rodado ou refletido sem distorção da informação)” (Venturini *et al.*, 2018, p. 268). Deste modo, a distância espacial é relevante para a examinação das redes digitais.

Uma leitura mais técnica explica-nos que os algoritmos de *layout* direcionados por força i) minimizam o cruzamento de arestas (tornando a representação gráfica esteticamente agradável) e ii) funcionam numa lógica de forças de atração e repulsão — a posição de cada nó irá depender da posição dos outros nós e da densidade das suas conexões (ver Fruchterman & Reingold, 1991; Kobourov, 2013; Jacomy *et al.*, 2014). No caso do ForceAtlas2, um algoritmo largamente utilizado

nos estudos de Internet, “os nós repulsam-se como partículas carregadas, ao passo que as pontas atraem os nós como elásticos. Estas forças criam um movimento que conduz a uma condição de equilíbrio. Em princípio, esta configuração final ajuda na interpretação dos dados” (Jacomy *et al.*, 2014, p. 2). Mas como podemos aplicar a espacialização direcionada por força, em particular o ForceAtlas2, às redes digitais multiplataforma?

No plano teórico, a espacialização do ForceAtlas2 dá resposta ao que Barabási e Albert (1999) postularam no modelo de redes de escala livre. Ou seja, permite uma leitura espacial assumindo que existe uma ordem na estruturação dinâmica das redes, que segue uma lei de potência e uma ligação preferencial (i.e. padrão de estrutura “rich get richer”) (Barabási & Albert, 1999). No plano prático¹¹, e diferente dos demais algoritmos de layout direcionados por força, a posição do nó numa rede responde as forças regulares de repulsão versus a força de repulsão calculada pelo grau (*degree*). Desta forma, ForceAtlas2 informa a importância de dois aspetos essenciais para leitura de uma rede: a característica relacional dos dados – como as conexões entre nós são feitas; e as potencialidades e riqueza da espacialização. Todavia, pouca atenção tem sido dada ao potencial de narrativa deste layout direcionado por força em estudos multiplataforma bem como em investigação no campo dos media digitais. Neste contexto, e na sequência de estudos exploratórios somados à realização de três workshops, defendemos que as redes digitais, ao serem espacializadas com o ForceAtlas2, podem ser lidas a partir de *camadas fixas* de interpretação (centro, ponto intermédio, periferia e elementos isolados) que proporcionam *múltiplas formas* de leitura¹² (ver Figura 3).

11 O ForceAtlas2 foca-se no grau da repulsão por grau (Jacomy *et al.*, 2014), i.e. a soma das ligações feitas por um nó. O grau de distribuição ($P(k)$) é uma probabilidade estabelecida a partir da frequência de cada grau. Quanto mais pequeno o expoente da lei de potência ($P(k) \propto k^{-\gamma}$), mais provável é que haja esse tipo de conexões. Esta lei determina que a maioria dos nós tem poucas interações, com poucos mas largos hubs e um número elevado número de conexões (Barabási, 2003). Esta definição remete para o conceito de rede de escalas livre, i.e. onde o número de conexões de nós obedece a uma distribuição regulada por leis de potência.

12 A proposta de leitura de redes digitais aqui apresentada responde não apenas a experiência prática das autoras em análise de redes digitais, mas também a trabalhos desenvolvidos em contexto colaborativos de *data sprints* nas Universidades de Amesterdão (2018 e 2019) e Nova de Lisboa (2019). A organização de três workshops ao longo de 2019 também reforçam a perspetiva teórica e prática proposta, sendo dois deles desenvolvidos na Universidade Nova de Lisboa e um no Center of Advanced Internet Studies (CAIS), em Bochum, Alemanha.

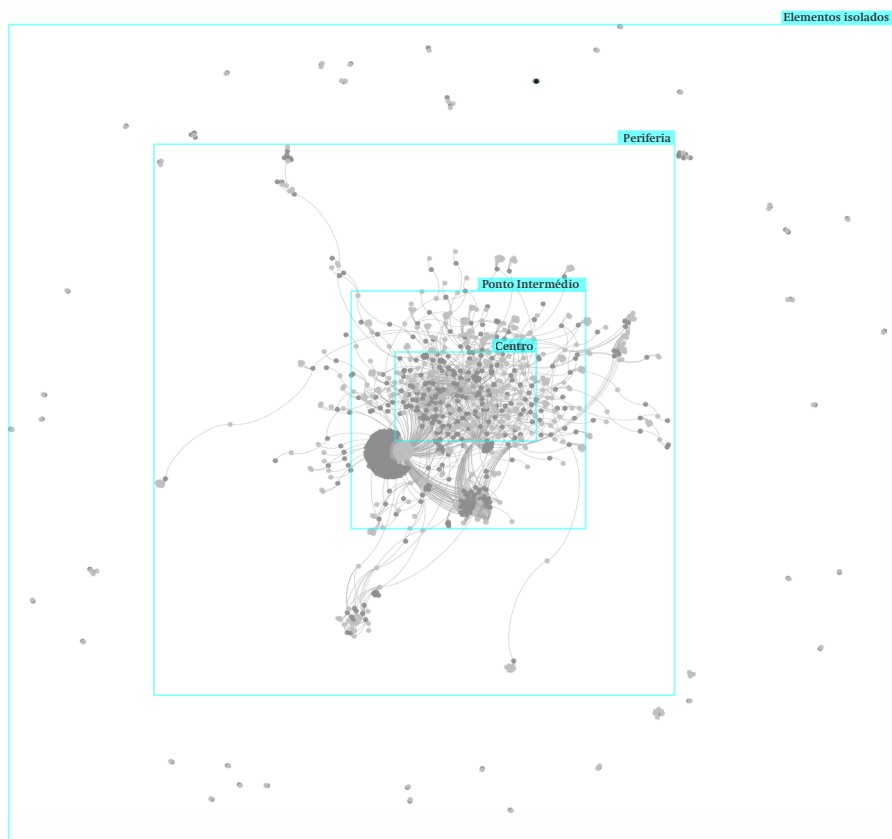


FIGURA 3 – Interpretação da representação gráfica a partir da espacialização do *ForceAtlas2*.

O centro da rede é a primeira camada, onde tendem a estar os nós mais conectados ou os nós que agrupam maior diversidade e variedade nas suas conexões, e.g. as hashtags mais recorrentes, os vídeos ou apps mais recomendados. Nesta dimensão, encontram-se atores mais empenhados ou ativos em torno de um tema específico, bem como nós mais populares. Na segunda camada, ou ponto intermédio, é possível identificar atores ou gramáticas influentes, *gatekeepers*, *bridging nodes* (i.e. nós que conectam diferentes *clusters*), bem como zonas vazias (com poucas conexões). A periferia da rede é um espaço onde se podem revelar diferentes perspetivas ou particularidades do tema principal. Os nós isolados, cuja existência se justifica, mas que por alguma razão não estão ligados ao componente central, completam o círculo que explica a espacialização do *ForceAtlas2*. Exemplos práticos sobre esta prática de leitura de redes digitais podem ser encontrados em projetos desenvolvidos no contexto de *data sprints* fundamentados na

perspetiva dos métodos digitais (Rogers, 2013). Por exemplo, redes digitais para leitura de camadas de recomendações algorítmicas da Google Play Store (Helmond *et al.*, 2018); interrogar os espaços semânticos proporcionados por diferentes Vision APIs através da visualidade de representações nacionais disponíveis em *stock websites* (Mintz *et al.*, 2019); verificar as imagens vinculadas a contas automatizadas (*botted accounts*) pertencentes ao Instagram e Tumblr e, em paralelo, investigar a circulação de imagens — que tipo de conteúdo visual permanece na plataforma e o que circula além da mesma (Omena *et al.*, 2019).

Esta interpretação está associada ao modelo de redes de escala livre (Barabási & Albert, 1999), que assenta no crescimento e ligações preferenciais, de acordo com leis de potência — estruturas que são orientadas numa ordem específica, que se traduz num mecanismo de conexão adequada. A característica principal é a distribuição desigual de ligações entre nós e a tendência para novos nós se conectarem a nós com elevado nível de conectividade, o que se deve às “leis de potência para a distribuição de graus” (Jacomy *et al.*, 2014). Portanto, as leis de potência associam-se a uma simetria particular que aponta para o argumento de que as redes não são igualitárias (Barabási & Albert, 1999). Nesta perspetiva, as redes possuem alguns nós muito conectados (*hubs*), mesmo que a maioria dos atores tenha poucas conexões. Os nós que conectam outros são os *bridging nodes* que têm laços fracos (Granovetter, 1973) de extrema importância na medida em que estabelecem a ligação entre diferentes grupos e *clusters*. Estes *bridging nodes* podem exercer influência, serem *gatekeepers* ou buracos estruturais (Burt, 1982; Venturini, Jacomy, & Pereira, 2015) que ligam a camada central à periferia.

Ao analisar a representação gráfica é fundamental ter em conta que as plataformas digitais configuram características específicas das redes digitais. Neste processo, é importante ter em consideração a espacialização da rede e as diferentes formas de interpretação.

(iii) o valor hermenêutico

Em concordância com estudos multiplataforma (*cross-platform studies*), entendemos que não são apenas aspetos práticos que devem servir de guia orientador para a pesquisa¹³ (ver Rogers, 2017) ou de como o conteúdo visual relacionado a um mesmo assunto/problema pode ser apresentado de forma diferente por dife-

¹³ Num artigo sobre análise multiplataforma, Richard Rogers (2017) sugere que os investigadores tenham em consideração: i) como a plataforma fornece o acesso aos dados — i.e. as suas *affordances* de interface e as suas limitações, e como a plataforma apresenta os conteúdos; ii) as culturas de uso

rentes plataformas (ver Pearce *et al.*, 2018). Em relação aos estudos de redes digitais multiplataforma, argumentamos que devemos atender a i) como as conexões se estabelecem nos diferentes tipos de redes – sejam diretamente oriundas de interfaces técnicas ou das *affordances* de dados digitais; ii) os diferentes significados e contextos que acarretam; iii) a lógica da representação gráfica e o seu potencial narrativo.

Nesta secção, propomos uma prática de interpretação de redes digitais multiplataforma a partir do potencial narrativo do *ForceAtlas2* baseado nas técnicas de exploração visual de redes – *visual network analysis* (Venturini, Jacomy, & Pereira, 2015; Venturini *et al.*, 2018; Venturini, Jacomy, & Jensen, 2019). Para fundamentar esta posição, defendemos que o valor da interpretação metodológica vem da consciencialização para a gramatização (e infraestrutura) da plataforma, combinada com a praxis de captura e análise de dados. A nossa sugestão está em concordância com a proposta de Latour (2010, 2012), para quem o conceito de *navigational and exploration practice* é um princípio básico para compreender redes digitais.

A posição dos nós, o seu tamanho e cor ocupam um papel proeminente na análise visual de redes digitais. Enquanto a posição do nó remete para os dados da plataforma e atuação dos *layouts* direcionados por força, o tamanho e cor refletem uma decisão analítica; algo que requer descrições detalhadas (*thick descriptions*) e camadas de conhecimento técnico, conforme explicámos na secção anterior (ver Figura 2). Neste sentido, devemos ter em conta que a proximidade e conetividade dos nós dentro da rede pode ser imposta pela gramatização da plataforma e, simultaneamente, moldada pelo software de análise e visualização (ver Figura 1). O que aqui indicamos como a narrativa das redes digitais (asseguradas pelo *ForceAtlas2*) remete mais para a “natureza relacional dos dados”, ou seja, que a sua proeminência não reflete necessariamente o volume de dados digitais (Marres & Rogers, 2005), mas a sua relação com uma determinada plataforma, um contexto particular e o processo de tomada de decisão.

Além disso, no processo analítico deve-se ter em conta o significado das ligações dentro da rede. Por exemplo, a forma como as conexões são feitas numa rede de vídeos de YouTube é diferente daquelas estabelecidas numa rede de *page-like* no Facebook. No caso do YouTube, os nós com mais vídeos recomendados (com um grau *indegree* elevado) seriam maiores. Por outro lado, no caso de páginas do Facebook, nós grandes indicariam páginas que gostaram de um maior número de páginas (com um elevado *outdegree*) ou que receberam muitos *likes*. Ou seja,

da plataforma; iii) e, com base nos dados disponíveis, as histórias que os dados da plataforma podem narrar.

quando se faz o mapeamento de um fenômeno social ou cultural no YouTube, observar a trajetória desde o centro da rede até à sua periferia pode ser mais relevante do que focalizar nos vídeos mais recomendados. Enquanto no caso das páginas do Facebook, a periferia da rede pode revelar grandes *clusters* devido à capacidade de uma determinada página de “gostar” de muitas outras. Seguindo esta linha de raciocínio, o tamanho dos nós ou as suas aglomerações dentro de uma rede não serão de certo um problema no processo analítico.

Através de uma *navigational practice* fornecida pela análise visual da rede, “os significados e resultados são gerados progressiva e hermenêuticamente” (Venturini *et al.*, 2018, p. 270), sendo que durante o processo analítico se devem equacionar algumas questões. Por exemplo: 1. O que é importante dentro da rede? (Como é que as conexões são estabelecidas? Atores influentes? Formação de *clusters*?) 2. Quais as plataformas ou métricas do Gephi que se devem considerar e/ou ignorar? 3. É necessário remover nós e/ou *clusters*? Porquê? 4. O que é visível mas irrelevante? 5. O que está escondido mas é importante? 6. O que é que precisa de ser descrito ou enfatizado? Estas questões devem ir a *par e passo* com a narrativa proporcionada pelos *layouts* direcionados por força, como o *ForceAtlas2*.

Considerando a lei de potência inerente às redes digitais, inferimos que a lógica da espacialização de rede com base no *ForceAtlas2* (ver Tabela 1) propõe algo definitivo: uma espécie de estrutura “fixa” que vai desde o centro até aos elementos isolados (e vice-versa); enquanto que a sua composição (os nós e a forma como se conectam) tem um caráter transitório, que muda de acordo com a plataforma. Portanto, esta estrutura permite *múltiplas formas* de leitura pois tem como essência a natureza relacional dos dados. Sublinhamos a relevância não só das distintas zonas da rede, mas sobretudo para a importância da zona periférica e da trajetória até aí (do centro para a periferia e vice-versa). Este caminho habitualmente fornece informações muito interessantes à investigação. Por exemplo, no YouTube ou nas App Stores podemos ler as redes através de camadas de recomendação, enquanto que nas Vision APIs (e.g. Google Vision, IBM Watson e Microsoft), podemos preferir olhar para a organização do espaço semântico de uma coleção de imagens e suas respectivas classificações. A partir daí, e graças às *affordances* das redes digitais, o estudo pode facilmente ser ampliado e aprofundado — por exemplo, é possível ir além da análise visual da rede e proceder a uma análise de conteúdo visual ou textual aprofundada.

| | Facebook Rede de page likes | Instagram Rede de coocorrências | Youtube Rede de recomendações algorítmicas | Vision APIs Rede de imagem-label |
|-------------------------------|---|--|---|---|
| Centro | (tendem a estar) Páginas mais conectadas ou populares | Hashtags/Termos com frequência elevada de coocorrência | Vídeos/Canais mais recomendados ou populares | (tendem a estar) Labels utilizadas com maior frequência na classificação automatizada do conteúdo visual (i.e. product, indoor, no person) |
| Ponto Intermédio | Páginas influentes, gatekeepers ou bridging pages | Bridging hashtags Camada complementar de conversação | Vídeos/Canais influentes Gatekeepers | Labels de ligação |
| Periferia | Clusters com formação baseada no ato de 'gostar' (like) | Clusters com temáticas específicas ou diferentes idiomas | (formação de cluster a partir de) Temáticas específicas ou diferentes idiomas ou interesse particular de um determinado vídeo/canal | (tendem a estar) Labels utilizadas com menor frequência na classificação automatizada das imagens. Indicam um particular conteúdo visual |
| Elementos Isolados | Páginas com a ausência de conexões | Hashtags/Termos com a ausência de coocorrência | ----- | ----- |

TABELA 1 – Matriz: Leitura de redes digitais multiplataforma de acordo com a narrativa fornecida pela representação espacial do ForceAtlas2.

A matriz da Tabela 1 exemplifica como podemos ler redes digitais multiplataforma de acordo com a narrativa fornecida pela espacialização do *ForceAtlas2*. A matriz está ancorada à perspectiva de um procedimento analítico que tem mais a ver com especificidades de plataforma e pesquisa digital prática, e menos com a migração de métodos tradicionais para o estudo de redes digitais. Neste sentido, a proposta é a de discutir o papel das conexões e a espacialização que se assumem como elementos-chave do processo. A investigação digital deve considerar o potencial narrativo da espacialização baseado em *layouts* direcionados por força como o *ForceAtlas2* e a sua relação com a gramatização da plataforma.

CONCLUSÃO

Neste capítulo apresentámos uma proposta de leitura de redes digitais multiplataforma com o propósito de compreender o significado de uma rede digital na sua totalidade. Ou seja, a nossa proposta vai além daquilo que se pode ver com recurso a softwares de visualização e análise; ou de como criar redes esteticamente agradáveis com recurso a *affordances* de software e as métricas baseadas em plataforma. Com base numa abordagem *medium-specific* (Rogers, 2013) e na tecnicidade das plataformas sociais (Omena, 2016), apresentámos uma abordagem que inclui um triplo entendimento para leitura de redes digitais multiplataforma. Este sistema permite mapear disparidades nas infraestruturas digitais multipla-

taforma, potenciando uma compreensão sobre a gramatização da plataforma com a prática de recolha e análise de dados. A proposta apresenta também é uma forma de *pensar* a representação gráfica das redes através de *layouts* direcionados por força. Em última instância, o sistema de leitura de redes digitais multiplataforma torna explícita uma prática específica de interpretação: a análise visual de redes e o encadeamento narrativo fornecido pelo *ForceAtlas2*. Trata-se de uma abordagem com caráter epistêmico, metodológico e relacional de dados.

Estas três fases interligadas têm em conta os princípios da análise multiplataforma e os “vernáculos de plataforma”, considerando tanto o conhecimento prático quanto o técnico no estudo das redes digitais como formações sociotécnicas. Considerando que “há mais complexidade nos elementos do que nas agregações” (Latour, 2012, p. 591), a nossa proposta inclui uma reflexão na trajetória que torna as redes digitais visíveis e prontas a explorar: o que simultaneamente precede e ocorre *com* e *no* procedimento analítico de leitura de redes digitais multiplataforma.

A abordagem apresentada surge como resposta ao apelo urgente pela literacia digital – também referida como “digital *Bildung*” (Rieder & Rohle, 2017) ou como “outra forma de cultura do conhecimento” (Marres, 2017) – nas Ciências Sociais Computacionais e Humanidades Digitais. Ao propor uma estratégia de leitura de redes digitais multiplataforma adequada à investigação, este capítulo abre novas possibilidades para o estudo e exploração de redes digitais. Contudo, salientamos que neste processo não se deverá negligenciar o papel da cultura de uso de cada plataforma em questão nem a contextualização do objeto de estudo.

REFERÊNCIAS

- Agre, P. E. (1997). *Computation and Human Experience*. Cambridge: Cambridge University Press.
- Agre, P. E. (1994). Surveillance and capture: Two models of privacy. *The Information Society* (10)2, 101-127.
- Amaral, I. (2016). *Redes sociais na internet: sociabilidades emergentes*. Covilhã: Labcom.IFP.
- Bannister, M., Eppstein, D., Goodrich, M. T., & Trott, L. (2013). Force-Directed Graph Drawing Using Social Gravity and Scaling. In W. Didimo & M. Patrignani (Eds.), *Graph Drawing GD 2012, Lecture Notes in Computer Science, vol. 7704* (pp. 414-425). Berlin, Heidelberg: Springer.
- Barabási, A. L. (2003). *Linked: The new science of networks*. Cambridge, Massachusetts: Perseus Publishing.
- Barabási, A. L., & Albert, R. (1999). Emergence of scaling in random networks. *Science* (286)5439, 509-512. doi: <https://doi.org/10.1126/science.286.5439.509>

- Barnes, J. A. (1954). Class and committees in a Norwegian island parish. *Human relations*, 7(1), 39-58. doi: <https://doi.org/10.1177%2F001872675400700102>
- Berlind, D. (2015). "What Are APIs and Do They Work?," *ProgrammableWeb*. Recuperado de <https://www.programmableweb.com/api-university/what-are-apis-and-how-do-they-work>
- Brandes, U., Linton, C., Freeman, & Wagner, D. (2014). Social Networks. In R. Tamassia (Ed.), *Handbook of graph drawing visualization* (pp. 805-839). Boca Raton: CRC Press.
- Burt, R. (1992). *Structural Holes*. Cambridge, MA: Harvard University Press.
- Callon, M. (1986). The sociology of an actor-network: The case of the electric vehicle. In M. Callon, A. Rip & J. Law (Eds), *Mapping the dynamics of science and technology* (pp. 19-34). Palgrave Macmillan, London.
- Bucher, T., & Helmond, H. (2017). "The Affordances of Social Media Platforms." In *The SAGE Handbook of Social Media*, edited by Jean Burgess, Thomas Poell, and Alice Marwick. London and New York: SAGE Publications Ltd. Retirado de https://www.annehelmond.nl/wordpress/wp-content/uploads/2016/07/BucherHelmond_SocialMediaAffordances-preprint.pdf
- Callon, M., & Latour, B. (1981). Unscrewing the big Leviathan: how actors macro-structure reality and how sociologists help them to do so. In K. Knorr-Cetina & A. V. Cicourel (Eds.), *Advances in social theory and methodology: Toward an integration of micro-and macro-sociologies* (pp. 275-303). Boston and London: Routledge and Kegan Paul.
- de Souza, C. R., Redmiles, D., Cheng, L. T., Millen, D., & Patterson, J. (2004, Novembro). Sometimes you need to see through walls: a field study of application programming interfaces. In *Proceedings of the 2004 ACM Conference on Computer Supported Cooperative Work* (pp. 63-71). ACM. doi: <https://doi.org/10.1145/1031607.1031620>
- doi: <https://doi.org/10.1109/MITP.2007.78>
- Fruchterman, T. M. J., & Reingold, E. M., (1991). Graph drawing by force-directed placement. *Software: Practice and Experience* (21)11, 1129-1164. doi: <https://doi.org/10.1002/spe.4380211102>
- Garton, L., Haythornthwaite, C., & Wellman, B. (1997). Studying online social networks. *Journal of computer-mediated communication*, 3(1). doi: <https://doi.org/10.1111/j.1083-6101.1997.tb00062.x> 2 January 2019.
- Gerlitz, C. (2016). What counts? Reflections on the multivalence of social media data. *Digital Culture & Society*, 2(2), 19-38. doi: <https://doi.org/10.14361/dcs-2016-0203>
- Gerlitz, C., & Rieder, B. (2018). Tweets are not created equal: Investigating Twitter's client ecosystem. *International journal of communication*, 12, 528-547.
- Gibbs, M., Meese, J., Arnold, M., Nansen, B., & Carter, M. (2015). # Funeral and Instagram: death, social media, and platform vernacular. *Information, Communication & Society*, 18(3), 255-268. doi: <https://doi.org/10.1080/1369118x.2014.987152>
- Giddens, A. (1987). *Social Theory and Modern Sociology*. Cambridge: Polity Press.
- Gillespie, T. (2010). The politics of 'platforms'. *New Media & Society*, 12(3), 347-364. doi: <https://doi.org/10.1177%2F1461444809342738>
- Granovetter, M. (1973). The Strength of Weak Ties. *American Journal of Sociology* (78)6, 1360-1380.
- d'Andrea, C., & Mintz, A. (2019). Studying the Live Cross-Platform Circulation of Images With Computer Vision API: An Experiment Based on a Sports Media Event. *International Journal Of Communication*, 13, 21. Retirado de <https://ijoc.org/index.php/ijoc/article/view/10423>
- Helmond, A. (2015). The platformization of the web: Making web data platform ready. *Social Media + Society* (1)2, 1-11. doi: <https://doi.org/10.1177%2F2056305115603080>
- Helmond, A., van der Vlist, F., Weltevrede, E., Christensen, C. M. D., Blaiotta, E., Omena, J.J., MacDonald, M., Bharati, S., & de Smale, S. (2018). Objectionable Queries: Search-

- ing for Porn in App Stores, *Digital Methods Initiative Wiki*. Retirado de <https://wiki.digitalmethods.net/Dmi/SummerSchool2018AppStoresBiasObjectableQueries>
- Highfield, T. (2018). Emoji hashtags/hashtag emoji: Of platforms, visual affect, and discursive flexibility. *First Monday*, 23(9). doi: <https://doi.org/10.5210/fm.v23i9.9398>
- Jacomy, M., Venturini, T., Heymann, S., & Bastian, M. (2014). ForceAtlas2, a Continuous Graph Layout Algorithm for Handy Network Visualization Designed for the Gephi Software. *PLoS One* (9)6, e98679. doi: <https://doi.org/10.1371/journal.pone.0098679>
- Kobourov, S. (2013). Force-Directed Drawing Algorithms. In R. Tamassia (Ed.), *Handbook of Graph Drawing and Visualization* (pp. 383-408). Boca Raton: CRC Press.
- Latour, B. (1987). *Science in action: How to follow scientists and engineers through society*. Cambridge, Massachusetts: Harvard university Press.
- Latour, B. (2010). Tarde's idea of quantification. In M. Candea (Ed.), *The Social after Gabriel Tarde: debates and assessments* (pp. 145-162). London: Routledge.
- Latour, B., Jensen, P., Venturini, T., Grauwin, S., & Boullier, D. (2012). 'The whole is always smaller than its parts' – a digital test of Gabriel Tarde's monads. *The British journal of sociology*, 63(4), 590-615. doi: <https://doi.org/10.1111/j.1468-4446.2012.01428.x>
- Marres, N. (2017). *Digital sociology: The reinvention of social research*. John Wiley & Sons.
- Meyer, B. (1988). *Object-oriented software construction*. New York: Prentice Hall.
- Murugesan, S. (2007). Understanding Web 2.0. *IT Professional Magazine*, 9(4), 34-41.
- Niederer, S., & Colombo, G. (2019). Visual Methodologies for Networked Images: Designing Visualizations for Collaborative Research, Cross-platform Analysis, and Public Participation. *Diseña*, (14), 40-67. doi: <https://doi.org/10.7764/diseña.14.40-67>
- Marres, N., & Rogers, R. (2005). Recipe for Tracing the Fate of Issues and their Publics on the Web. In B. Latour & P. Weibel (Eds.), *Making Things Public: Atmospheres of Democracy* (pp. 922-935). Cambridge, MA: MIT Press.
- Mintz, A., Silva, T., Gobbo, B., Pilipets, E., Azhar, H., Takamitsu, H., Omena, J.J., & Oliveira, T. (2019). Interrogating Vision APIs. Data Sprint Report. *SMART Data Sprint 2019: Beyond Visible Engagement*. Retirado de <https://smart.inovamedialab.org/smart-2019/project-reports/interrogating-vision-apis/>
- Omena, J. J. (2016). Social Media Technicity. What to look for in the APIs? *Encontro com Ciência e Tecnologia em Portugal*, 4-6 julho de 2016, Centro de Congressos de Lisboa.
- Omena, J.J., Chao, J., Pilipets, E., Kollanyi, B., Zilli, B., Flaim, G., Sivori, H., Ruiven, K., Rademakers, L., Li, M., & Del Nero, S. (2019). Bots and the Black Market of Engagement, *Digital Methods Initiative Wiki*. Retirado de <https://wiki.digitalmethods.net/Dmi/SummerSchool2019Botsandtheblackmarket>
- Omena, J.J., Rabello, E. & Mintz, A. (2020). Digital Methods for Hashtag Engagement Research. *Social Media and Society*, special issue 'Studying Instagram Beyond Selfies. (no prelo)
- Pearce, W., Özkula, S. M., Greene, A. K., Teeling, L., Bansard, J. S., Omena, J. J., & Rabello, E. T. (2018). Visual cross-platform analysis: digital methods to research social media images. *Information, Communication & Society*, 1-20. doi: <https://doi.org/10.1080/1369118X.2018.1486871>
- Rieder, B., & Röhle, T. (2012). Digital methods: Five challenges. In D. M. Berry (Ed.), *Understanding Digital Humanities* (pp. 67-84). London: Palgrave Macmillan.
- Rieder, B., & Röhle, T. (2017). Digital methods: From challenges to Bildung. In M. T. Schäfer & K. van Es (Eds.), *The Datafied Society: Studying Culture through Data* (pp. 109-124). Amsterdam: Amsterdam University Press.
- Rogers, R. (2013). *Digital methods*. Cambridge, MA: MIT Press.
- Rogers, R. (2017). Digital methods for cross-platform analysis. In J. Burgess, A. Marwick, & T. Poell (Eds.) *The SAGE handbook of social media* (pp. 91-110). London: Sage.
- Simmel, G. (1955). *Conflict & The Web of Group-Affiliations*. New York: The Free Press.

- Silva, T. (2019). Visão Computacional e Vieses Racializados: branquitude como padrão no aprendizado de máquina. II COPENE Nordeste: *Epistemologias Negras e Lutas Antirracistas*, 29-31 maio de 2019, Universidade Federal da Paraíba. Retirado de https://www.academia.edu/40084537/Vis%C3%A3o_Computacional_e_Vieses_Racializados_branquitude_como_padr%C3%A3o_no_aprendizado_de_m%C3%A1quina
- Stiegler, B. (2006). Anamnesis and Hypomnesis: The Memories of Desire. In A. Bradley & L. Armand (Eds.), *Technicity* (pp. 15-41). Prague: Litteraria Pragensia.
- Stiegler, B. (2012). Die Aufklärung [The Enlightenment] in the Age of Philosophical Engineering. *Computational Culture – a journal of software studies*. Recuperado de <http://computationalculture.net/die-aufklarung-in-the-age-of-philosophical-engineering/>
- Venturini, T., & Latour, B. (2009). The Social Fabric: Digital footprints and quali-quantitative methods. In *Proceedings of Future en Seine*, 87-103. Retirado de https://medialab.sciencespo.fr/publications/Venturini_Latour-The_Social_Fabric.pdf
- Venturini, T., & Rogers, R. (2019). “API-Based Research” or How can Digital Sociology and Journalism Studies Learn from the Facebook and Cambridge Analytica Data Breach. *Digital Journalism*, 1-9.
- Venturini, T., Jacomy, M., Bounegru, L., & Gray, J. (2018). Visual Network Exploration for Data Journalists. In S. Eldridge II & B. Franklin (Eds.), *The Routledge Handbook of Developments in Digital Journalism Studies* (pp. 265-283). London: Routledge.
- Venturini, T., Mathieu, J., & Pereira, D. (2015). *Visual Network Analysis*. SciencesPo Media-Lab Working Paper. Retirado de http://www.tommasoventurini.it/wp/wp-content/uploads/2014/08/Venturini-Jacomy_Visual-Network-Analysis_WorkingPaper.pdf
- Venturini, T., Jacomy, M., & Jensen, P. (2019). What do we see when we look at networks. arXiv preprint arXiv:1905.02202. Retirado de <https://arxiv.org/abs/1905.02202>
- Venturini, T., Munk, A., & Jacomy, M. (2019). Actor-Network versus Network Analysis versus Digital Networks: Are We Talking about the Same Networks? In J. Vertesi & D. Ribes (Eds.), *DigitalSTS: A Field Guide for Science & Technology Studies* (pp. 510-524). PRINCETON; OXFORD: Princeton University Press.
- Wasserman, S., & Faust, K. (1994). *Social Network Analysis: Methods and Applications*. Cambridge: Cambridge University Press.
- Watts, D. (2004). The ‘New’ Science of Networks. *Annual Review of Sociology* (30)1, 243-270.
- Wellman, B. (1983). Network analysis: Some basic principles. *Sociological Theory* (1)1, 155-200.

*Tradução por Filipa Felipe.



QUATRO

O PENSAMENTO NATIVAMENTE DIGITAL



MÉTODOS DIGITAIS NOS ESTUDOS EM SAÚDE

MAPEANDO USOS E PROPONDO SENTIDOS

ELAINE TEIXEIRA RABELLO
FÁBIO CASTRO GOUVEIA

Em tempos em que tudo está sob monitoramento e controle, a centralidade das questões de saúde para compreender os modos de vida contemporâneos é inegável. Adicione-se a este cenário o papel da Internet e a miríade de plataformas através das quais as pessoas (profissionais, cientistas, agências, governo, o público em geral, etc.) podem falar e interagir quase “livremente” sobre um tópico em tendência. Ao longo dos últimos dez anos, diversas questões que exemplificam como isso acontece no campo da saúde: movimentos sociais por direito à saúde, epidemias, movimento antivacina, profusão de tratamentos alternativos à biomedicina, debates acirrados sobre políticas de saúde, uso da Internet e dispositivos para monitoramento e avaliação de comportamentos de saúde, etc. Na interface saúde-sociedade, cada vez mais a Internet e as diferentes plataformas digitais têm sido usadas pela população em geral para buscar, coletar ou até mesmo checar informações e “verdades” para tomada de decisão em relação à sua saúde (Camargo & Grant, 2015).

Nesta última década, os usos e papéis da Internet e das mídias digitais na saúde também têm tido centralidade nos estudos acadêmicos, que buscam compreender o papel e o impacto do fenômeno da Internet na perspectiva psicossocial, e na saúde. Porém, por estarem em uma área interdisciplinar, diversas abordagens teóricas e metodológicas têm sido adotadas em estudos e publicações do campo, muitas sob a mesma denominação de “digitais”, mas abordando diferentes questões de pesquisa e estratégias de investigação. Embora a epide-

miologia e a saúde pública lidem basicamente com dados para realizar estudos populacionais, nos estudos que acompanhamos na nossa rotina acadêmica, ainda há potencialidades inexploradas de usos de tipos diferentes de dados, de uma fonte diferente e de uma maneira diferente: não-estruturados, extraídos de plataformas sociais, disponíveis e ligados em múltiplas mídias e formatos (texto, legenda, imagens, vídeo, hashtags, etc.), produzidos e divulgados por pessoas em geral, tudo isso tentando interpretar os significados desses dados em termos de engajamento. Isso exige de nós, pesquisadores da saúde, novas estratégias para lidar com dados, novas formas de coleta / rastreamento, organização / visualização, compreensão / análise de dados em grandes quantidades ou brutos. Como mencionado em Rogers (2019, p.3), “métodos digitais são estratégias de investigação para lidar com a natureza efêmera e instável dos dados online”, em uma sociedade na qual esta esfera é parte da estrutura social e cultural.

Este texto pretende lançar um debate que não se encerra em si, mas levanta pontos para a discussão sobre tais perspectivas sob as quais a Internet e as plataformas têm figurado nos estudos e publicações em saúde pública e coletiva cujo objetivo, processo ou conteúdo tenha relação com os métodos digitais e discussões afins.

Neste sentido, nosso objetivo é mapear o terreno dos estudos em saúde que vêm adotando, nominalmente, “métodos digitais” de investigação, entendendo melhor que características esses estudos, pesquisadores, objetos e contribuições vêm dando ao campo dos estudos digitais em saúde. A expectativa é que este mapeamento dos estudos em saúde que apontam a adoção dos métodos digitais possa nos dar uma noção de como essa perspectiva está sendo incorporada na área, assim como sinalizar para a potencialidade desta ser uma inovação para o campo da saúde pública, onde a pesquisa e a ciência tradicional geralmente entendem os dados em um sentido mais restrito e ligado ao campo da epidemiologia, da computação e da informação.

Paralelamente, debatemos que perguntas e que contribuições a perspectiva dos métodos digitais discutidas neste livro, assim como a adoção dos seus objetos de investigação, podem ou poderiam agregar a esses estudos em saúde, defendendo o posicionamento da utilização crítica e defesa das potencialidades dos métodos digitais para estudos do tema.

A POLISSEMIA DO “DIGITAL”

Esse interessante campo que se desenha (com diversos nomes e abarcando variadas correntes epistemológicas) nos coloca inicialmente uma questão terminológica. Rogers (2019), ao apresentar a abordagem dos métodos digitais, defende sua acepção para além da ideia do uso de algumas ferramentas, sugerindo inclusive que a abordagem poderia ser denominada de “metodologias digitais”, porque se refere mais a um processo de pesquisa online, sem, no entanto ser protocolar em termos de técnica. “Ciências Sociais Computacionais”, “Humanidades Digitais”, “Estudos de Novas Mídias”, entre outros, são todos nomes para campos que, por vezes, estão em sobreposição e são evocados para classificar e até mesmo legitimar alguns estudos que navegam na interface da saúde com o digital nas formas, ferramentas e conteúdos comunicados em saúde. Entretanto, entre eles, diferem muito os objetos investigados, a compreensão de noções como mídia, método digital, ferramenta e dados; mas, principalmente, diferem o grau e o papel de atuação das ciências sociais ao formular as perguntas do estudo, bem como analisar os dados trabalhados.

Veja abaixo na Figura 1 um diagrama de Venn¹ com um cruzamento das coocorrências de termos nos títulos, resumos ou palavras-chave para “Digital Humanities” (DigHumanities), “Computational Social Sciences” (CompSocialScien), “Digital Methods” (DigMethods) e “New Media Studies” (NewMStudies) nos registros na base Scopus. Podemos ver que, apesar de existirem interseções, os termos estabelecem interpretações e possivelmente campos com apropriações do “digital” que dialogam com um referencial teórico próprio dada a baixa sobreposição.

Os métodos digitais são definidos por Rogers (2015, p. 1) como “técnicas para o estudo das mudanças sociais e condições culturais com dados online”. No campo da inovação em saúde, por exemplo, trabalhar com grandes quantidades de dados geralmente significa “minerar” conjuntos e bancos buscando padrões, diferenças e semelhanças entre as entidades, pois eles se relacionam com os resultados operacionais desejados. No entanto, os métodos digitais, como afirma Rogers e sua equipe, oferecem uma oportunidade de ir além desse uso de “big data”.

Nessa perspectiva, os dados derivados da web não refletem necessariamente padrões ou informações estruturadas. Pelo contrário, a maneira como são montadas online é, por si só, um reflexo de como a sociedade lidou com as tecnologias

¹ Diagrama gerado no Interactivenn: <http://www.interactivenn.net/> Heberle, H.; Meirelles, G. V.; da Silva, F. R.; Telles, G. P.; Minghim, R. InteractiVenn: a web-based tool for the analysis of sets through Venn diagrams. BMC Bioinformatics 16:169 (2015). DOI: 10.1186/s12859-015-0611-3

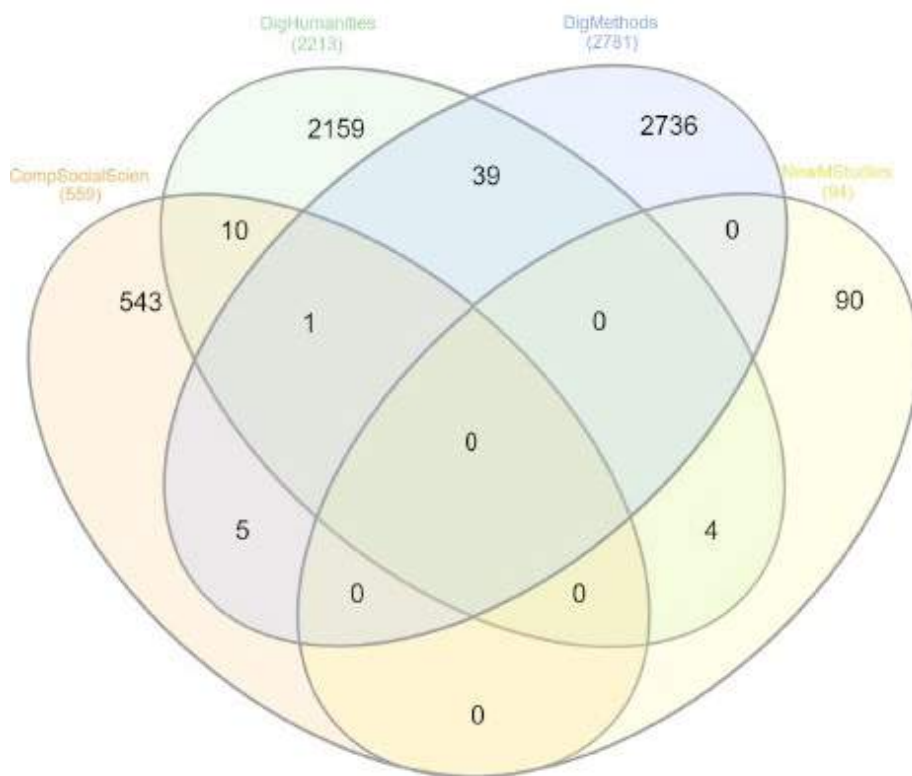


FIGURA 1 – Diagrama de Venn com a coocorrências de termos nos títulos, resumos ou palavras-chave para “Digital Humanities” (DigHumanities), “Computational Social Sciences” (CompSocialScien), “Digital Methods” (DigMethods) e “New Media Studies” (NewMStudies) em registros na base Scopus.

de comunicação e informação (TIC). Os mecanismos de busca, por exemplo, não necessariamente são “fonte” de informações sobre saúde, mas certamente são fontes de “como as pessoas pesquisam e encontram” informação de saúde. Como estudar motores de busca e plataformas? Como suas interfaces estão influenciando as maneiras de interagir? Nesta equação, interesses, tecnicidade, lógica de algoritmos e vieses, tudo isso deve ser levado em consideração na pesquisa digital, não necessariamente como algo ruim que interfere na qualidade dos dados, mas como um objeto a ser considerado em si.

Conforme afirma Rogers (2013; 2015), isso significa que devemos tornar o digital não apenas uma questão de abordagem metodológica ou uma tendência ao uso de ferramentas computacionais, mas principalmente uma maneira de garantir a disponibilidade das condições culturais da produção de dados. Esse é o potencial dos métodos digitais quando aplicados a questões da saúde, considerando que a web e a Internet não tiveram apenas um impacto sobre os seres

humanos e a sociedade como uma maneira de interagir, mas também influenciaram o modo como a ciência e a política atuam na área da saúde.

Tudo isso considerado, as pesquisas em saúde sob a perspectiva dos métodos digitais dão hoje aos cientistas sociais grandes oportunidades para entender os usos e impactos sociais de dispositivos, gadgets e a própria web nos seres humanos e na sociedade. Dá, aos sanitaristas, excelente suporte teórico e metodológico não apenas para conduzir a pesquisa, mas principalmente para elaborar pesquisas voltadas para a imiscuidade do online na rotina de cuidados, práticas e noções circulantes de saúde.

No entanto, lidar com esses dados online tem sido um novo desafio que, pelo que vimos nesta exploração, foi abraçada pela área da saúde recentemente, a partir de 2016. Técnicas ou ferramentas que visam coletar, analisar, rastrear e organizar dados das plataformas da web e de mídia social têm exigido um tipo de alfabetização e uma infraestrutura sociotécnica mais ampla, pois cada plataforma tem sua maneira de organizar (e fornecer) dados, suas *affordances*. Tradicionalmente, isso é visto como um viés para muitos cientistas sociais e da saúde, quando se assume a ideia de seguir o ator e considerar a plataforma de conteúdo da web acaba sendo uma maneira de pré-formatar os discursos dos usuários. No entanto, como Marres e Weltvrede (2013) e Rogers (2013, 2019) propõem, de acordo com a lógica de seguir o meio, o enorme potencial de usar a Internet como um objeto em si, sendo a investigação do meio uma parte (crucial) da investigação do “social” (Marres e Waltevrede, 2013, p. 13), é algo que não podemos perder no horizonte especialmente na saúde pública e coletiva.

Para o campo da saúde, nossa hipótese inicial era de que os estudos que adotam nominalmente a perspectiva digital derivam das discussões clássicas no campo acerca de sistemas de informação, informetria, big data, etc., e mais recentemente dos estudos das Tecnologias de Informação e Comunicação em Saúde e dos estudos de educação em saúde. Como pesquisadores do campo, observamos que trabalhar na perspectiva “digital” significa, na maioria das vezes, manejar bancos de dados populacionais e de unidades de saúde, lidar com métricas e dados epidemiológicos, realizar avaliações sistematizadas ou automatizadas de serviços e políticas públicas ou, ainda, aplicar alguma tecnologia para fins de educação ou vigilância em saúde. Apenas mais recentemente o conteúdo circulante em mídias sociais, por exemplo, passou a ser considerado como relevante para entender o cenário da saúde populacional, para além da perspectiva puramente comunicacional.

Apesar das múltiplas acepções possíveis, seguindo a discussão de Rogers (2013) sobre métodos digitalizados e métodos nativamente digitais, nossa

hipótese é de que o que têm se propagado como estudos que utilizam métodos digitais em saúde pública e coletiva, na verdade, são em sua maioria releituras ou reaplicações de metodologias já estabelecidas no campo que, em alguma medida, envolvem técnicas computacionais, manipulação e gerenciamento de dados, uso de dispositivos eletrônicos, mas não necessariamente adotam a perspectiva dos métodos digitais em toda sua potencialidade. Isso porque o papel que as ciências sociais assumem numa perspectiva de métodos digitais é central, mas não necessariamente é adotado em todos os estudos de saúde. Portanto, surge-nos a primeira pergunta que conduzirá nosso debate a seguir: o que os estudos em saúde têm referido como métodos digitais?

Para explorar essa questão, objetivamos:

1. Mapear o campo de publicação de pesquisas em saúde para ver quais aceções de “métodos digitais” estão sendo adotadas;
2. Apresentar a partir desse mapeamento as principais características dessas pesquisas;
3. Ilustrar como a noção de “métodos digitais” amplamente discutida neste livro pode aportar às discussões da saúde uma perspectiva mais crítica de pesquisa dos objetos nativamente digitais.

O QUE SE CHAMA DE MÉTODOS DIGITAIS NOS ESTUDOS DE SAÚDE?

Para explorar o cruzamento entre as publicações sobre métodos digitais em saúde, foi feita uma busca na base Scopus² pelo termo “{digital methods}” dentro das categorias título, resumo e palavras-chave (*title, abstract and keywords*) para todas as tipologias documentais publicadas depois de 2006³. Obtivemos uma lista de 873 documentos, que serviram de base para a posterior filtragem pela coocorrência do termo “health” restando 24 documentos. Após uma leitura de cada um dos documentos para limpeza (exclusão de artigos que embora usem o termo *digital*

² Escolhemos uma base bibliográfica de amplo alcance por possibilitar acesso a estudos em âmbito global, não restritos à língua; além disso, por contemplar diversos tipos de publicações acadêmicas, nos dá um cenário sobre a aceção do termo “métodos digitais” no âmbito científico em geral.

³ Consideramos o marco do ano de 2007 como momento de consolidação da proposta dos métodos digitais na perspectiva do Digital Methods Initiative (DMI), entendendo que é a partir deste momento que surge a ideia de pesquisa nativamente digital. Ainda cientes da historicidade do termo e do próprio grupo fundador do DMI, descartar menções anteriores à expressão “métodos digitais” na publicação científica nos pareceu uma forma de refinar a busca, já que o termo em si é bastante polissêmico, conforme discutimos.

methods, o usam em uma acepção completamente diferente da que nos serve), chegamos à uma lista final de 17 documentos que estariam adotando métodos digitais para investigação de temas em saúde, conforme Quadro 1 abaixo:

| ID | Título | Ano | Autor Principal | Revista |
|----|--|------|-------------------|---|
| A | Tracing medicinal agencies: Antidepressants and life-effects | 2019 | Ruckenstein M. | Social Science and Medicine |
| B | Restor(y)ing Health: A Conceptual Model of the Effects of Digital Storytelling | 2019 | Fiddian-Green A. | Health Promotion Practice |
| C | Big Data Challenges from an Integrative Exposome/Exposure Perspective | 2019 | Martin-Sanchez F. | Lecture Notes in Bioengineering |
| D | Disease detection, epidemiology and outbreak response: the digital future of public health practice | 2018 | Velasco E. | Life Sciences, Society and Policy |
| E | Using digital storytelling to promote the sexual health and well-being of migrant and refugee young people: A scoping review | 2018 | Botfield J.R. | Health Education Journal |
| F | Technology use in mental health practice and research: Legal and ethical risks | 2018 | Lustgarten S.D, | Clinical Psychology: Science and Practice |
| G | The walkthrough method: An approach to the study of apps | 2018 | Light B. | New Media and Society |
| H | Political hashtag publics and counter-visibility: a case study of #fertilityday in Italy | 2018 | Vicari S. | Information Communication and Society |
| I | Environmental justice in the age of big data: challenging toxic blind spots of voice, speed, and expertise | 2017 | Mah A. | Environmental Sociology |
| J | DERBI: A digital method to help researchers offer "Right-to-know" personal exposure results | 2017 | Boronow K.E. | Environmental Health Perspectives |
| L | Mapping the new nordic issue-scape: How to navigate a diffuse controversy with digital methods | 2016 | Munk A.K. | Tourism Encounters and Controversies: Ontological Politics of Tourism Development |
| M | Advancing digital methods in the fight against communicable diseases | 2015 | Chabot-Couture G | International Health |
| N | Cancer information seeking in the digital age: Effects of angelina jolie's prophylactic mastectomy announcement | 2015 | Noar S.M. | Medical Decision Making |

| ID | Título | Ano | Autor Principal | Revista |
|----|--|------|----------------------------------|--|
| O | Collaborative care for co-morbid major depressive disorder in chronically ill outpatients in a general hospital [Collaborative care voor de behandeling van comorbide depressieve stoornis bij chronisch lichamelijk zieke patienten op een polikliniek van een algemeen ziekenhuis] | 2015 | Van Steenbergem-Weijnenburg K.M. | Tijdschrift voor Psychiatrie |
| P | A situated practice of ethics for participatory visual and digital methods in public health research and practice: A focus on digital storytelling | 2014 | Gubrium A.C, | American Journal of Public Health |
| Q | New technology in dietary assessment: A review of digital methods in improving food record accuracy | 2013 | Stumbo P.J. | Proceedings of the Nutrition Society |
| R | Digital communication to support clinical supervision: considering the human factors. | 2013 | Mather C. | Studies in health technology and informatics |

QUADRO 1 – Listagem dos 17 estudos selecionados para análise qualitativa da interface entre “saúde” e “métodos digitais”.

A partir das palavras-chave, títulos e resumos desses 17 documentos, fizemos primeiramente dois grafos importando e tratando os termos pelo software VOSviewer⁴ e gerando as visualizações pelo software Gephi⁵. O primeiro (Figura 2) representa a coocorrência das palavras-chave atribuídas pelos autores dos textos sem restrições à sua frequência mínima e com pesos fracionários (uma palavra-chave dentre 5 contaria 0,2 enquanto uma dentre 2 contaria 0,5) o que resultou em 69 palavras-chave diferentes. A colorização dos agrupamentos foi efetuada a partir do cálculo de modularidade.

⁴ Disponível em: <http://www.vosviewer.com>

⁵ Disponível em: <https://gephi.org>

Em seguida, fizemos a leitura completa e a categorização manual do conteúdo destes 17 estudos na interface métodos digitais e saúde. Nosso objetivo, com isso, foi analisar qualitativamente as acepções de “métodos digitais” trazidas em cada um desses estudos.

Em uma primeira leitura, percebemos que há, basicamente, três acepções aplicadas ao que é digital e que, conseqüentemente, justificam a adoção da expressão “método digital” na descrição dos estudos. São elas em ordem de frequência neste levantamento:

- Digital como ferramenta ou recurso: esses estudos consideram o termo “digital” e “métodos digitais” para descrever e discutir a o desenvolvimento, o uso e a aplicação de recursos tecnológicos que vão desde produção de vídeos até desenvolvimento de aplicativos e plataformas. No sentido empregado, a ênfase está na ideia de digital como recurso que viabiliza ou incrementa alguma ação ou atividade.
- Digital como arena ou fonte de dados: estudos entendem o digital como espaço da análise, seja porque se constitui na arena onde algum debate ocorre ou, ainda, porque se trata de um espaço que armazena, organiza e entrega dados importantes para a pesquisa. Neste sentido, a ênfase está nas informações disponíveis no ambiente digital, que servem ao propósito da atividade ou ação.
- Digital como objeto de estudo em si: estes estudos abordam plataformas, mídias, buscadores, enfim, os elementos que compõem o mundo digital como o objeto da investigação, entendendo suas características próprias e abordando-as na interface com a saúde e a sociedade. Neste caso, é comum que sejam discutidas também outras visões do digital em conjunto (como arena/fonte ou ferramenta/recurso).

Na leitura dos 17 artigos, foi possível aplicar esta classificação por abordagem, conforme mostrada no Quadro 2. Notem que é possível que um mesmo estudo tenha adotado mais de uma acepção para o digital:

| Abordagem do “digital” | Estudos (ID) |
|------------------------|------------------------------|
| Ferramenta/Recurso | B, C, E, F, J, M, O, P, Q, R |
| Arena/Fonte | A, D, G, H, I, L, N |
| Objeto | A, D, F, G, H, L |

QUADRO 2 – estudos classificados por abordagem.

Em síntese, quando os estudos mencionam a expressão “método digital”, em geral, fazem referência 1) a adoção, uso ou desenvolvimento de alguma tecnologia, ferramenta, plataforma ou dispositivo para registro de informação ou comunicação; ou 2) ao online como espaço que abriga debates e dados e, acessar este espaço, é em si adotar um método digital; ou 3) a estudos mais recentes, onde objectos nativos digitais passam a ser vistos como objetos de estudo.

Os artigos em nosso levantamento que colocaram o digital como objeto de estudo aparecem a partir de 2016. Os anteriores consideram como *digital methods* o uso de alguma ferramenta online ou dispositivo. Depois de 2016, encontramos ainda esta visão atrelada ao conceito de storytelling (B, E). É também perceptível que é nas revistas do campo da saúde que estão os artigos mais ferramentais, enquanto outras áreas escoam produções com a acepção de digital como objeto. A exceção do estudo F, a apropriação do digital como arena/fonte de estudo ocorre em conjunto com a de objeto em si.

É importante ressaltar que nenhuma das perspectivas acima mencionadas para a expressão “métodos digitais” nos estudos analisados estão erradas ou são limitadas. A partir de cada perspectiva em particular, os autores fazem a apropriação de determinado tipo de método, metodologia ou técnica para tentar ler aquilo que eles enxergam: um objeto em si, uma arena de debate ou fonte, uma ferramenta ou recurso. No entanto, ao transcender a separação entre elas, estudos como o A, D, F, G, H e L ilustram melhor a perspectiva a que os métodos digitais se propõem, ao considerarem que entender os mecanismos e subjetividades imiscuídas nas plataformas e os objetos nativamente digitais (como links, threads, hashtags, reactions, etc.) é central para o objetivo da pesquisa.

As abordagens de métodos digitais utilizadas nos estudos de saúde, segundo a exploração detalhada anteriormente, remetem à tríade objeto-local-instrumento de pesquisa explorada por Fragoso et al. (2011) para métodos para Internet, ainda que, à época, o panorama de possibilidades das pesquisas fosse outro.

Porém, como diferenciar arena e fonte? No Twitter, por exemplo, quando coletamos material desta plataforma, ela é uma arena (onde está ocorrendo o debate) ao mesmo tempo em que é fonte de dados (o próprio Twitter fornece os dados, e somente utilizamos alguma ferramenta no processo de extração). Ou ainda em uma netnografia, por exemplo, em que entramos em um grupo de Facebook, observamos e eventualmente coletamos algo, sendo também uma arena ao mesmo tempo em que é fonte desses dados.

No estudo A, analisado neste trabalho, temos o exemplo do uso de uma ferramenta para fazer coleta e análise de dados de saúde da população tal como informado por ela mesma; ou seja, se cria um ambiente que será alimentado para,

então, servir de fonte de dados para a análise de determinada situação. Este tipo de situação mostra o quanto a separação entre a ideia de arena e fonte é delicada e por isso mantida em conjunto na nossa proposição.

Disto vemos que tanto a divisão que Fragoso et al. (2011) faz, quanto a que nos pareceu dar conta dos estudos de saúde, não são divisões simples à primeira vista, pois trazem problemas de sobreposição na sua origem. No entanto, ao mesmo tempo, levantam debates importantes para melhor definir o escopo dos métodos digitais e do que eles podem aportar de novo para a área. Que novas perguntas a perspectiva dos métodos digitais poderiam trazer para esses estudos em saúde? E ainda: o que, na saúde, já é feito que coaduna com a perspectiva dos métodos digitais, ainda que usando outro nome? O que é ou pode ser feito não apenas sobre a web, mas com a web?

POTENCIAIS CONTRIBUIÇÕES DOS MÉTODOS DIGITAIS AO CAMPO DA SAÚDE: PROBLEMATIZANDO AS PERSPECTIVAS, DEFENDENDO AS SUAS AFFORDANCES

Em se considerando a tríade proposta, de acordo com esses 17 estudos, primeiramente assumimos a fusão apresentada entre fonte e arena, para o caso da saúde; além disso, vamos sugerir um condicionante para o entendimento do digital como *objeto de estudo*. Ao seguirmos no debate do que melhor caracterizaria esse terceiro tipo de abordagem, presente em 6 dos 17 estudos, exemplificaremos as potencialidades dos métodos digitais para a saúde. A chave interpretativa que propomos está na consideração das *affordances* das plataformas e suas características.

Se, ao fazermos uma coleta de material no Twitter, seguida da análise do que é postado, não entrarmos na percepção das características dessa plataforma que viabilizam, interferem ou até constroem o debate como ele é (e por conseguinte leva a uma série de interpretações), estamos simplesmente fazendo análise de conteúdo, olhando o que as pessoas estão falando, tratando como arena. Assim sendo, faria pouca diferença se o material viesse do Twitter, do Facebook, se são as notícias de jornal, se era um grupo de bate papo por telefone, ou artigos da BVS; afinal, o meio não é levado em consideração como tão protagonista quanto o humano.

Há diversos trabalhos sobre questões do meio digital que usam – mas naturalizam – plataformas como o Facebook, por exemplo, no sentido de acessar o grupo para ver o que as pessoas discutem, mas não entram na mecânica e na discussão sobre o que é um grupo de Facebook. Ou seja, são deixadas de lado na

discussão as características e possibilidades que um grupo de Facebook aporta (por exemplo, de curtir, compartilhar, ou fazer outras coisas), quais são as características que os define (por exemplo, quais as questões de um grupo fechado, um grupo aberto, um grupo secreto), entre outras. Na discussão sobre dados de um grupo secreto, por exemplo, já tenho embutido todo o entendimento do que é um grupo secreto, autocontido e protegido por uma certa sensação de “sigilo”, não por indivíduo, mas pelo coletivo.

Em cada plataforma existem possibilidades de diferentes tipos de ações e reações, que apontam para a lógica subjacente ao sistema relacional e que não deve ser ignorado. Um exemplo: no mundo “real”, esse conceito de curtir e compartilhar não existe, ou seja, isso é uma *affordance* de algumas plataformas digitais. O compartilhamento, nelas, é a reprodução *verbatim* de algo que outra pessoa disse, o que num processo de história oral, comentando sobre o que o outro disse, é mais interpretativo, e não *ipsis litteris* como no compartilhamento. Isso também é uma *affordance* do sistema que, por ser digital, permite uma reprodução das coisas. Quando compartilhamos, reproduzimos digitalmente algo. Podemos até fazer um uma interpretação ou um comentário em cima daquilo, podemos talvez editar aquele conteúdo e, nesse sentido, gerar uma mensagem diferente em alguns casos (memes, por exemplo). As coisas acontecem assim, nesta plataforma, porque ela dá essas possibilidades. Ou seja, a forma de poder gostar de algo lá é curtindo ou botando um coração no texto ou fazendo um comentário, ou postando um meme. Mas são essas as opções, um leque pré-determinado.

Se não tomadas em consideração essas questões, perde-se uma boa via crítica de debates para os estudos, que podem ir além das perspectivas ferramentais, descritivas e conteudistas ao entender esses espaços e suas características, e é isso que traz a riqueza dos métodos digitais para além da lógica arena-ferramenta-fonte. Se não levarmos em conta as *affordances* do meio, perdemos a ideia de “seguir o meio”.

A discussão crítica das *affordances* do meio é o que, na nossa visão, diferencia os métodos digitais de abordagens como a netnografia⁶. Não estamos falando na

⁶ Na realidade, em que pese a profusão de estudos que se dizem netnográficos, o que de fato os caracteriza é a sua transposição para o meio digital da proposição de Malinowski de entrar na comunidade e não fazer um mero relato secundário (como coletar coisas e analisar sem participar daquele lugar). Há muitos estudos que dizem fazer netnografia mas, na verdade, simplesmente coletam os dados que foram discutidos numa determinada página e fazem interpretação. Eles não veem o processo do desenvolvimento do debate. Então isso não seria netnografia, é, sim, uma análise de conteúdo. Coletar, por exemplo, pelo falecido Netvizz todo o debate que teve num determinado lugar e fazer uma análise, mas não ver debate acontecer passo a passo, impede de se perceber, o quão intenso ou rápido foram as conversas, o quanto as pessoas tocaram umas às outras, o captar a emoção e entender esse processo. Isto não seria possível numa simples coleta de dados.

necessidade de se integrar ao meio e ao grupo apenas, porque neste caso o meio não é o objeto de estudo. Então, na ideia de “seguir o meio”, dos métodos digitais, o meio serve de protagonista. A plataforma em si é um ator não-humano, e suas *affordances* são chave para o entendimento do que acontece ali. Não é apenas o contexto ou o cenário onde as coisas estão acontecendo; é por causa dela que as coisas estão acontecendo; ela faz acontecer assim.

De fato, o meio não é o objeto no caso da netnografia, mas ele de certa forma acaba tendo seu entendimento, porém subjetivo. A sua participação leva a percepção e consideração indireta das *affordances*. É importante perceber, neste sentido, que não há no método digital uma proposta ou visão etnográfica em si, pois não necessariamente acompanhamos o desenrolar do debate; mas, ao mesmo tempo, trazemos essa questão da *affordance* do meio, dando centralidade ao seu entendimento. E por isso ele é um método com características muito próprias, porque não é um método de vivência, como a netnografia, mas ainda assim considera a vivência que se tem de uma determinada plataforma.

Em alguns dos estudos em saúde analisados qualitativamente, há, sim, uma certa percepção das *affordances* e características, ainda que incipiente. Todos os estudos que entendem o digital como objeto vão explorar estas questões (estudos A, D, F, G, H, L), aportando, com isso, um debate mais crítico em relação aos temas pesquisados. Característica comum entre esses estudos é a perspectiva integradora dos métodos digitais como referencial, abordando o papel das mídias sociais e diversas plataformas não apenas como difusoras, mas também como produtoras de informação (e conhecimento) em saúde para diversos públicos por conta justamente de elementos que, aqui, seriam traduzidos como suas *affordances*.

O primeiro movimento que sugerimos, nesse sentido de integração dos métodos digitais aos estudos de saúde, está na colocação das próprias perguntas do estudo para que a pesquisa seja, em si, digital desde seu desenho. Como defendido por Rogers (2019), a formulação de questões de pesquisa e a narrativa dos achados e resultados são as bases dos métodos digitais, e suas duas principais ações são a construção de uma boa *query* e de uma boa lista de fontes. Um exemplo, afora os artigos mencionados, é o estudo de caso desenvolvido no Smart Data Sprint de 2018, no qual se investigou como o público em geral se envolveu com o tema da epidemia de zika nas redes sociais⁷. Os pesquisadores abordaram a questão a partir de panorama mais amplo (abordando a Wikipédia e o Google, pois são fontes de informações em vários idiomas) e de uma visão específica,

⁷ <https://smart.inovamedialab.org/smart-2018/project-reports/visualising-engagement-on-zika-epidemic/>

analisando focalmente as discussões no Instagram e no Facebook em língua portuguesa. Para cada plataforma, diferentes abordagens e métodos foram especificamente desenhados.

Um segundo movimento seria aproveitar a linguagem da própria web, dos objetos nativos digitais e dos métodos que a própria web dá (ex. sistemas de recomendação, crawling, as gramáticas das redes sociais), a serviço da investigação em saúde. Nesse sentido, ilustramos com o caso estudado no SMART Data Sprint 2019, que adotou a perspectiva dos métodos digitais para, a partir da informação que circula e a partir de objetos e métodos digitais nativos, rastrear informações sobre terapias detox e movimento antivacina na web, buscando não apenas atores, mas lemas de engajamento e interesses de diferentes pessoas neste tipo de causa⁸.

A partir de pesquisas preliminares, sabíamos que tais controvérsias mobilizaram diferentes interesses, atores, estratégias de comunicação e discursos. Também tínhamos clareza de que neste processo é fundamental o entendimento das *affordances* dos meios digitais em estudo. Parte-se do pressuposto de sua não neutralidade e influência nas construções discursivas em rede a partir dos recursos que disponibilizam para tal, e pela forma de apresentação e priorização algorítmica dos resultados apresentados aos usuários. Assim, enquanto pesquisadores, mantemos o olhar também no entendimento da plataforma onde os debates se estabelecem e suas modificações no tempo. Durante o sprint, foi possível identificar onde, na web, esses debates aconteceram; mapear os problemas em diferentes plataformas, visando a análise crítica; identificar as ideias, argumentos e representações de publicações e posts subjacentes à saúde; e inferir possíveis repercussões, encaminhamentos e impactos sociais do que circula na rede sobre comportamentos de saúde em relação aos temas escolhidos como casos.

Em ambos os estudos exemplificados acima, não necessariamente conseguimos dar conta por completo das várias questões das *affordances*. No entanto, neles já entendemos a importância e atentamos às características que o meio tem; levamos elas em consideração na hora de uma coleta, por exemplo, quando tentamos perceber o perfil de uma dada comunidade através dos emojis usados. Começamos a entender que cada plataforma permite uma forma de expressão específica e que a coleta reflete isto, em especial e principalmente no *query design*. Certamente ainda há o que se fazer para ampliar o potencial dos métodos digitais nesses experimentos, mas já são sinalizações do que queremos para o campo dos métodos digitais na interface com a saúde.

⁸ <https://smart.inovamedialab.org/smart-2019/project-reports/how-is-health-being-depicted-in-online-debates-the-cases-of-detox-practices-and-vaccines/>

REFERÊNCIAS

- Camargo, K., & Grant, R. (2015). Public health, science, and policy debate: Being right is not enough. *American Journal of Public Health*, 105(2), 232-235. <https://doi.org/10.2105/AJPH.2014.302241>
- Fragoso, S., Recuero, R., Amaral, A. (2011). *Métodos de pesquisa para internet*. Porto Alegre, RS: Sulina.
- Marres, N., & Weltevrede, E. (2013). Scraping the social?: Issues in live social research. *Journal of Cultural Economy*, 6(3), 313–335. <https://doi.org/10.1080/17530350.2013.772070>
- Rabello, E. T., Carneiro, B., Ribeiro, D., Gouveia, F.C., Prado, J. A., Brademann, M., & França, T. (2019). How is health being depicted in online debates? The cases of 'detox practices' and 'vaccines'. Recuperado de *Smart Data Sprint Blog – Inova Media Lab – Universidade Nova de Lisboa* website: <https://smart.inovamedialab.org/smart-2019/project-reports/how-is-health-being-depicted-in-online-debates-the-cases-of-detox-practices-and-vaccines/>
- Rabello, E. T., Matta, G. C., Costa, A. R., Teixeira, A., Barbosa, C., Flaim, G., ... Silva, T. (2018). Visualising engagement on Zika epidemic. Recuperado de *Smart Data Sprint Blog – Inova Media Lab – Universidade Nova de Lisboa* website: <https://smart.inovamedialab.org/smart-2018/project-reports/visualising-engagement-on-zika-epidemic/>
- Rogers, R. (2013). *Digital methods*. Cambridge, Massachusetts: The MIT Press.
- Rogers, R. (2015). Digital methods for web research. In R. A. Scott & S. M. Kosslyn (Orgs.), *Emerging Trends in the Social and Behavioral Sciences* (p. 1-22). <https://doi.org/10.1002/9781118900772.etrds0076>
- Rogers, R. (2019). *Doing digital methods* (1st edition). Thousand Oaks, CA: SAGE Publications.



O PAPEL DO *DESIGNER* NO CONTEXTO DO *DATA SPRINT*

MICHELE MAURI
BEATRICE GOBBO
GABRIELE COLOMBO

CONTEXTO TEÓRICO

Os *sprints* nascem no contexto informático, principalmente na produção de *software* de código aberto, para resolver os diversos níveis de incertezas: quem vai participar, que características terá, que soluções técnicas serão adotadas.

Tendo em conta a incapacidade de estruturas rígidas e complexas gerirem tais incertezas, o *sprint* propõe-se como um evento, limitado no tempo, para chegar ao melhor resultado possível com um objetivo específico. O *sprint* foi afinado, tornando-se um dos elementos centrais de algumas metodologias Agile (Schwaber, 2017). Ao longo dos anos o termo foi utilizado como referência a períodos de trabalho intensivo, sendo utilizado em várias áreas, por exemplo na escrita de livros (Barker, 2013), na projeção de serviços (Knapp, 2017) e, mais recentemente, em áreas académicas, na recolha e análise de dados (Venturini, 2018). Neste capítulo vamos concentrar-nos neste último ponto, o *data sprint*.

CARACTERÍSTICAS DO *DATA SPRINT*

Apesar de existirem muitas variações nas diversas interpretações e nas várias áreas, é possível identificar alguns elementos comuns como os que são especificados a seguir.

Tempo limitado

Uma das poucas certezas de um *sprint* é o tempo: o prazo não pode ser alterado. Esta limitação transforma-se num desafio para os participantes: é preciso produzir algo no tempo-limite.

Dirigido ao produto

Um *sprint* tem o objetivo de produzir alguma coisa: um texto, um código, um *dataset*, uma apresentação. Além do tipo de *output*, procura-se produzir um resultado mais pormenorizado e certo possível.

Orientado pelos casos de estudo

Ao contrário do que o nome sugere, um *data sprint* nunca é guiado pelo objetivo de criar um *dataset*. O objetivo é sempre definido através de questões de investigação, às quais é possível dar uma resposta a partir dos dados recolhidos *online*.

Em copresença

Os participantes trabalham em simultâneo e de forma colaborativa no projeto, mesmo que não presencialmente. Isto também define o número de participantes: deve ser equilibrado de modo a cobrir as competências necessárias sem, no entanto, anular as vantagens relativamente aos grupos demasiado grandes. Esta característica também torna evidente a necessidade de definir uma série de ferramentas tanto digitais como físicas, que permitam a partilha de resultados, a colaboração e a rastreabilidade das ações executadas.

Os *data sprints* funcionam particularmente bem em alguns contextos, podendo ser contraproducentes noutros (Munk, Madsen & Jacomy, a publicar). Por um lado, e de acordo com Munk, Madsen & Jacomy, os *sprints* funcionam bastante bem quando participantes de diferentes nacionalidades e *backgrounds* precisam de entender que tipo de resultados podem alcançar juntos. São de igual modo promissores, quando há falta de métodos estruturados de análise, promovendo, conseqüentemente, um ambiente favorável para a criação de novos méto-

dos. Por outro lado, e quando as premissas do projeto são estáveis, os *data sprints* podem inclusive ser contraproducentes (Venturini, Munk & Meunier, 2018).

Como se irá ver, os *sprints* podem ter diversos objetivos, mas em geral são úteis quando se está a gerir problemas que não foram bem definidos, quando é difícil definir o resumo do projeto inicial, e quando as diversas categorias profissionais têm de organizar o tipo de resultados que querem alcançar.

TIPOS DE DATA SPRINTS

Na literatura, as características dos *data sprints* são normalmente referidas procurando descrever o modelo independentemente do resultado. Nos últimos anos, o DensityDesign Research Lab¹ organizou e participou em vários *sprints* e identificou empiricamente três tipos principais de *data sprints*. Os de *ideação*, de *produção* e de *afinação*.

Os *sprints* de *ideação* funcionam quando há uma oportunidade de investigação ou uma hipótese, ainda que vaga, e nenhum *dataset* foi ainda definido nem há uma ideia clara de como produzi-lo. O objetivo do *sprint* de *ideação* é entender como recolher, filtrar e estruturar dados funcionais para responder à questão inicial de investigação. Este tipo de *sprint* é aquele em que, mais frequentemente, uma multiplicidade de produtos semiacabados é criada, sendo o uso de visualização pedagógico para mostrar as características dos dados recolhidos e para testar as suas potencialidades em relação às questões iniciais de investigação.

Os *sprints* de *produção*, pelo contrário, visam explorar em profundidade uma linha de investigação que já foi formalizada e testada anteriormente. Tipicamente, neste tipo de *sprint*, já se tem uma ideia da fonte de dados que se quer usar e o tipo de exploração que se deseja fazer, e as operações são mais sobre recolha, limpeza e enriquecimento de dados. As visualizações neste caso têm uma função de validação da consistência do *dataset* com as questões de investigação iniciais.

Os *sprints* de *afinação*, por último, são *sprints* em que o objetivo é entender como devolver o *corpus* de uma pesquisa a audiências mais amplas (Mauri, Colombo, Angeles & Ciuccarelli, 2019). Nestes casos, temos vários produtos semiacabados, em termos de *dataset* e possíveis visualizações, e trata-se de entender que informações comunicar, a que públicos, e de que forma o fazer. Normalmente, neste

¹ <https://densitydesign.org/>

tipo de *sprint*, a visualização é usada para criar protótipos com a maior fidelidade possível para comunicar os dados.

Resultados: os semiacabados

O *data sprint* é um momento vulcânico, no qual são produzidos numerosos artefactos que se sobrepõem uns aos outros. De uma questão de investigação, surgem várias possibilidades de criação de *datasets*. Cada um destes *datasets* pode responder à questão ou motivar novas questões. Esses artefactos podem ser chamados produtos semiacabados, cuja utilidade é entender como proceder na análise, e não tanto usá-los como produtos acabados.

Como já observado na literatura (Mauri & Ciuccarelli, 2016), a maior parte destes resultados no final do projeto será descartada, mas, graças a eles, foi possível entender se são um elemento final, se levantam novas hipóteses, se é necessário rever a última passagem, ou até mesmo rever todo o processo.

Os artefactos trazidos como apresentação final podem, portanto, ser vistos como “rascunhos” dos resultados mais promissores: a partir destes resultados é possível rever o processo pelo qual os rascunhos foram produzidos para que possam ser estruturados num processo de *design*.

ORGANIZAÇÃO DO *DATA SPRINT*

Um *sprint* normalmente é organizado ao longo de um período de quatro, cinco dias. Esta é a dimensão certa para que haja um momento de *divergência* (exploração de possíveis soluções) e de *convergência* (identificar os mais promissores e explorá-los em profundidade).

No primeiro dia, é feito o resumo do projeto, ou melhor a questão inicial de investigação (ou questões). O material inicial também é fornecido e ilustrado, de modo a permitir que todos os participantes tenham um ponto de partida comum e saibam os objetivos claros do *sprint*.

Ao longo da *data sprint* é prática comum envolver especialistas de domínio e/ou das metodologias para dar *feedback* aos participantes, de modo a dar-lhes uma visão externa sobre o trabalho que estão a desenvolver, e identificar erros e incoerências.

No último dia, são feitas as apresentações. Cada grupo de participantes apresenta aos demais o trabalho produzido durante o *sprint*.

O tempo restante é totalmente utilizado pelos participantes para trabalhar em grupo nos seus projetos.

Pre-data sprint

O *data sprint*, para ter um tempo de execução limitado, requer uma preparação cuidadosa antes da sua realização. Em primeiro lugar, é necessário definir as principais questões de investigação, para que os grupos de trabalho já saibam em que trabalhar, quando o *sprint* começar. Também é aconselhável partilhar o programa do *sprint* para que todos os participantes estejam cientes do prazo final. Finalmente, é necessário recolher todo o material útil e partilhá-lo antes do evento. Em algumas ocasiões, como durante as escolas de verão/inverno da Digital Methods Initiative, é proposta uma lista de referências bibliográficas. Noutras ocasiões, os conjuntos de dados foram pré-selecionados, recolhidos e documentados ou fornecidos dados úteis para desempenhar o trabalho. Como o *data sprint* é focado na resolução de problemas “mal definidos”, é aconselhável realizar todas as operações mecânicas, repetitivas ou que geralmente não exigem uma abordagem multidisciplinar antes do *sprint* em si.

Pós-data sprint

Como vimos, o resultado de um *data sprint* é uma multiplicidade de produtos semiacabados. É clara a necessidade de deixar vestígios das ações executadas: retomando a metáfora da erupção, é necessário documentar o material antes que “esfrie”, caso contrário, será impossível reconstruir, *a posteriori*, que escolhas foram feitas e que significado tiveram os diversos produtos semiacabados, tornando-os de facto inutilizáveis.

O PAPEL DO DESIGNER NO DATA SPRINT

Uma metáfora usada para descrever o papel do *designer* (Mauri & Ciuccarelli, 2016) é a cartográfica: o *sprint* é uma exploração de um território desconhecido, o *designer* é o cartógrafo de bordo que traça o que vê na viagem, mas acima de tudo é aquele que reúne as notas e os esboços no regresso, para poder reconstruir uma visão global, um mapa, que outros poderão usar.

Em cada etapa do *sprint*, o *designer* desempenha um papel diferente.

Na fase pré-*data sprint* é importante definir o formato e os “recipientes” que serão adotados durante o *sprint*. Ter restrições ajuda a focar o esforço do *design* (por exemplo, planearei de forma diferente se souber que a visualização será projetada e não impressa num poster).

Na fase do *data sprint*, as tarefas do *designer* são três: informar, orientar e registrar.

DATA SPRINTS E DESIGNERS, NA PRÁTICA

As já históricas Escolas de Verão e Inverno de Métodos Digitais – organizadas pela Universidade de Amsterdão – e a recente iniciativa SMART Data Sprint – coordenada pela Faculdade de Ciências Sociais e Humanas da Universidade Nova de Lisboa – são encontros sazonais em que os investigadores do grupo DensityDesign, especialistas em *design* da informação, estão envolvidos. Essas iniciativas reúnem, principalmente, investigadores especialistas em diferentes setores disciplinares (ciências sociais, sociologia, *future studies*, tecnoantropologia, ciências computacionais), que mergulham em dias intensos de investigação dirigida ao aprofundamento e à apresentação de temas contemporâneos relacionados com os novos *media* e os métodos digitais.

O grupo e os papéis

O grupo de investigadores em *design* que apoia os participantes ao longo da semana de *sprint* é composto por pós-doutores, estudantes de doutoramento, bolsiros de investigação e estudantes com diferentes níveis de especialização e experiência na área de *design* da informação e visualização de dados.

Técnicas básicas de processamento de dados, conhecimento da teoria de mapeamento de controvérsias, manipulação de conteúdos gráficos e prototipação de artefactos “semiacabados” (Mauri & Ciuccarelli, 2016) estáticos e interativos são as competências exigidas aos investigadores envolvidos em *design*.

De facto, durante os dias de *data sprint*, os investigadores de *design* – individualmente ou em pares – nos grupos de projeto, dão apoio aos participantes desde a fase de exploração de dados aproximados e instáveis (fase de divergência) até à fase de convergência, que requer uma apresentação visual e concreta dos conteúdos analisados.

No entanto, como antecipado acima, o papel do investigador em *design* durante os dias de *sprint* é múltiplo e pode ser dividido em três ações principais:

- (a) a ação de informar: o *designer* realiza sessões de tutorial promovendo ferramentas ou abordagens utilizadas e testadas no campo do *design* da informação;
- (b) a ação de guiar: o *designer* acompanha e ajuda os participantes como facilitador durante o desenvolvimento da investigação;
- (c) a ação de registrar: o *designer* toma nota e apresenta visualmente todas as fases da investigação – diagrama do protocolo de pesquisa ou “protocolo visual” (Niederer & Colombo, 2019), a fim de tornar o processo transparente, facilmente partilhável e replicável.

INFORMAR

O objetivo das primeiras horas ou dias de *um data sprint*, dedicado a sessões de tutorial, é iniciar os participantes – que tenham algum conhecimento em gestão e processamento de dados – no uso dos métodos digitais, criando um glossário comum de termos, técnicas e ferramentas fundamentais para o desenvolvimento da investigação nos dias seguintes.



IMAGEM 1 – Uma fotografia tirada durante um tutorial em janeiro de 2019 no SMART Data Sprint. Digital Media Winter Institute 2019, SMART Data Sprint: Beyond Visible Engagement, 28 January – 1 February 2019 – Universidade Nova de Lisboa.

Do ponto de vista do *design* da informação, é necessário fornecer aos participantes, provenientes dos diferentes setores disciplinares, ferramentas essenciais para visualizar rapidamente as informações, que permitem e ajudam a processar a leitura crítica dos fenômenos complexos analisados e estimulam a coprodução de conhecimento (Venturini, Munk & Meunier, 2018).

Os momentos dos tutoriais são ocasiões informais com duração variável de 75' a 90' em que os *designers* apresentam ferramentas ou abordagens aos participantes, aperfeiçoando e atualizando, ano após ano, a sua estrutura e critérios de apresentação, num processo iterativo submetido a modificações contínuas. Cada tutorial é realizado por dois ou três investigadores: o primeiro, mais experiente, cuida da condução da aula, os outros apoiam e auxiliam os participantes. Atualmente, cada tutorial é composto por quatro fases:

FASE 1

A ferramenta (ou abordagem) é introduzida e apresentada, usando *slides*.

FASE 2

Dedicada a um exercício guiado que os participantes são convidados a seguir passo a passo.

FASE 3

Os participantes, livres para “sujar as mãos” com os dados, repetem o processo da fase anterior a partir de novos *datasets*. Ao longo do exercício, é dada atenção especial à construção de visualizações capazes de responder a uma questão de investigação, e funcionais, de modo a destacar um aspeto específico surgido da exploração e da observação dos dados.

FASE 4

Momento em que as visualizações produzidas pelos participantes são projetadas e comentadas publicamente.

Informar

introdução → passo a passo → exercício → discussão

IMAGEM 2 – A estrutura de um tutorial no contexto de um *data sprint*.

Durante um *data sprint*, com base no tema geral dos projetos e nos conteúdos das sessões paralelas de tutorial, são apresentadas duas ou três das quatro ferramentas/abordagens listadas nas linhas a seguir.

Alguns tutoriais mostram uma seleção de funcionalidades de uma aplicação, como no caso do tutorial sobre RawGraphs (1), enquanto outros combinam metodologias específicas com várias ferramentas, como no caso do tutorial para visualizações de redes de imagens (3).

1. Da folha de cálculo à visualização

Uma ferramenta intuitiva que permite visualizar rapidamente os dados é o RawGraphs², uma aplicação de código aberto para a visualização de dados.

A partir de uma folha de cálculo, o RawGraphs permite criar, personalizar e fazer o *download* de visualizações estáticas.

Uma das principais dificuldades que surgiram durante as sessões de tutorial diz respeito à reorganização de um *dataset* funcional para a construção da visualização; por essa razão ao longo dos anos, a fase de introdução foi ampliada e acompanhada de uma secção dedicada ao projeto da estrutura de linhas e colunas de base dados.

2. Da pasta de imagens para a visualização

As ferramentas de código aberto ImageSorter³ e ImagePlot⁴ permitem que os investigadores analisem grupos de imagens, ordenando-as cromaticamente ou organizando-as no espaço cartesiano de acordo com os parâmetros numéricos derivados de metadados das imagens – ou seja, número de comentários ou reações à imagem numa página do Facebook, data e hora da publicação, saturação, etc.

Como nem todos os participantes têm uma pasta de imagens para visualizar, o tutorial também inclui uma fase dedicada ao uso de ferramentas que permitem baixar imagens e respetivas URLs da *web* (i.e. os plug-ins: DownThemAll e GrabThemAll).

3. Da pasta de imagens para a rede

Quando é necessário mostrar as relações entre imagens, com base no seu conteúdo, ferramentas gratuitas de classificação de imagens baseadas em interfaces

2 <https://rawgraphs.io/>

3 <https://imagesorter.software.informer.com/>

4 <http://lab.softwarestudies.com/p/imageplot.html>

de programação de aplicações (APIs)⁵, como o Clarifai⁶ ou o Imagga⁷, juntamente com o Gephi⁸, ajudam a criar redes para visualizar tendências, *clusters* e fenómenos visuais.

Como no caso anterior, uma etapa do exercício passo a passo é dedicada à recolha de imagens e à organização dos URLs em *datasets*.

4. Da visualização à *messa in scena*

Por fim, elaborar uma visualização de dados também significa planear a sua encenação (Colombo, 2018) ou organizá-la num contexto de elementos visuais como título, legenda e etiqueta que garantam a sua legibilidade por um público específico.

Neste caso, um dos aspetos críticos é a capacidade de organizar os conteúdos de forma hierárquica, para guiar e orientar a leitura.

GUIAR

A partir do segundo dia do *data sprint*, os *designers* são distribuídos de forma equitativa pelos grupos de projeto, de acordo com as suas competências, interesses e experiência.

Uma comunicação intergrupo entre os membros do Density Design Lab permite gerir problemas, partilhar informações e coordenar as questões de interesse transversal.

Em termos de comunicação intragrupo, é essencial que o *designer* participe nas primeiras fases de discussão dos projetos em que as questões iniciais de investigação são definidas e aperfeiçoadas, a fim de esboçar coletivamente uma abordagem inicial de investigação que será otimizada durante o trabalho.

De facto, durante a fase central do *data sprint*, dirigida à investigação, o *designer* tem a tarefa de colaborar com os outros participantes num contexto igualitário.

Ao contrário do papel de formador característico das sessões tutoriais, o papel do facilitador durante o trabalho de grupo não é adequado a uma esquematização, pois é variável, adaptável e sujeito a estímulos contínuos, dependendo das competências do grupo e do tema de investigação.

⁵ Application Programming Interface. Ferramentas que permitem integrar funções específicas, como Image Content Recognition, no interior das aplicações.

⁶ <https://clarifai.com/>

⁷ <https://imagga.com/>

⁸ <https://gephi.org/>

Desempenhar o papel de facilitador implica apoiar os participantes no uso das ferramentas de visualização apresentadas durante as sessões de tutorial; tanto para aconselhar os modelos visuais existentes, funcionais para a investigação – e, às vezes, para experimentar novos –, como para guiar o grupo na *messa in scena* e apresentação das visualizações semiacabadas.

Como mostrado na Figura 3, o processo pode ser representado por um padrão circular no qual as ações não se seguem de maneira ordenada. A leitura crítica das visualizações exploratórias preliminares do *dataset*, através de anotações (Niederer, 2019), sobreposições e comparações de múltiplos resultados, gera estímulos e questões de investigação às quais novas visualizações tentarão responder.

Além disso, a tarefa do *designer* de informação é também elaborar e reelaborar visualizações gerais, sempre “semiprocessadas”, úteis para a compreensão dos resultados, que serão incluídos numa única apresentação.

Este processo é muitas vezes difícil e lento: é necessário selecionar as conclusões mais relevantes em detrimento de outras, encontrar os modelos visuais apropriados para mostrar os resultados de uma maneira abrangente e encenar tudo numa apresentação breve e concisa.

Guiar

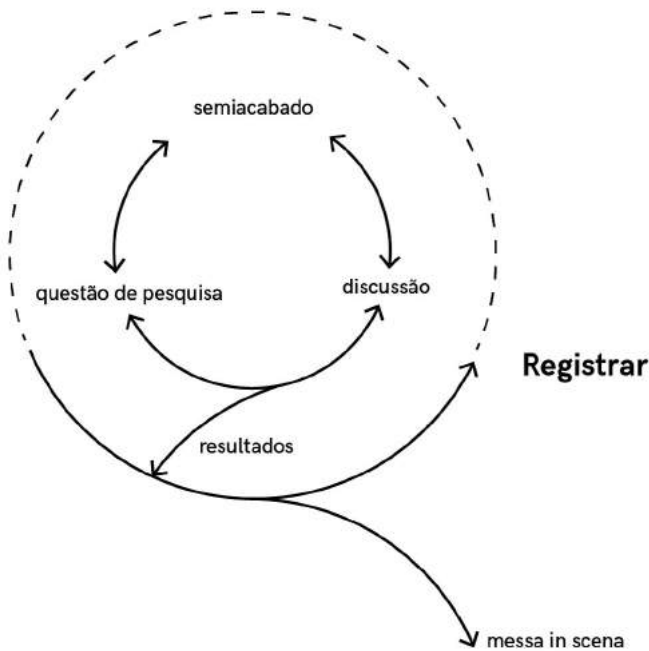


IMAGEM 3 – Representação do papel do *designer*, como facilitador, durante o *data sprint*.

REGISTRAR

Por fim, o *designer* tem o papel de registrar as ações da investigação, a fim de tornar o processo compreensível e reproduzível.

Muitas vezes, na verdade, é preciso lidar com processos complexos de investigação, que exigem uma pesquisa paralela sobre diferentes fontes e o uso de múltiplos instrumentos.

A produção do diagrama do protocolo de investigação capaz de seguir o processo de *design* de pesquisa, por um lado, permite ilustrar cada etapa durante a apresentação de fim de semana, mas acima de tudo representa um elemento indispensável para partilhar o trabalho e permitir que outros investigadores repitam as mesmas ações e obtenham o mesmo resultado.

Por esta razão, o papel do protocolo visual está associado ao da receita culinária: uma progressão de ações a serem realizadas usando vários ingredientes e ferramentas.

No entanto, o objetivo não é apenas apoiar o leitor na reprodução das receitas apresentadas, mas também mostrar uma certa abordagem, assim como acontece na gastronomia, esperando que os leitores se sintam inspirados a adaptar, modificar e experimentar (Bounegru, Gray, Venturini & Mauri, 2017).

Cada protocolo, geralmente, contém informações sobre: o processo de recolha de informações (*queries design*, tipo de *scraping*) de uma ou mais fontes; as ferramentas (*scripts*, *softwares*, codificação manual) usadas para transformar e analisar os dados; e as plataformas e *softwares* usados para visualizá-los.

A construção do protocolo visual reflete o processo iterativo da pesquisa, que se adapta, dia após dia, ao trabalho do grupo do qual até os fracassos e as frustrações podem fazer parte. De facto, é boa prática começar a projetar o artefacto desde o início e observar as mudanças durante a semana, a fim de obter um mapeamento completo das escolhas feitas pelo grupo de trabalho.

Essa ação continua na fase pós-*data sprint*, pois muitas vezes apenas no final da análise é possível reconstruir o que aconteceu, que resultados foram obtidos e os caminhos seguidos.

Diferenciar visualmente os elementos e organizá-los hierarquicamente de forma diagramática auxilia na descodificação e leitura pelo utilizador (ou um investigador hipoteticamente interessado no tema que deseja repetir o mesmo processo).

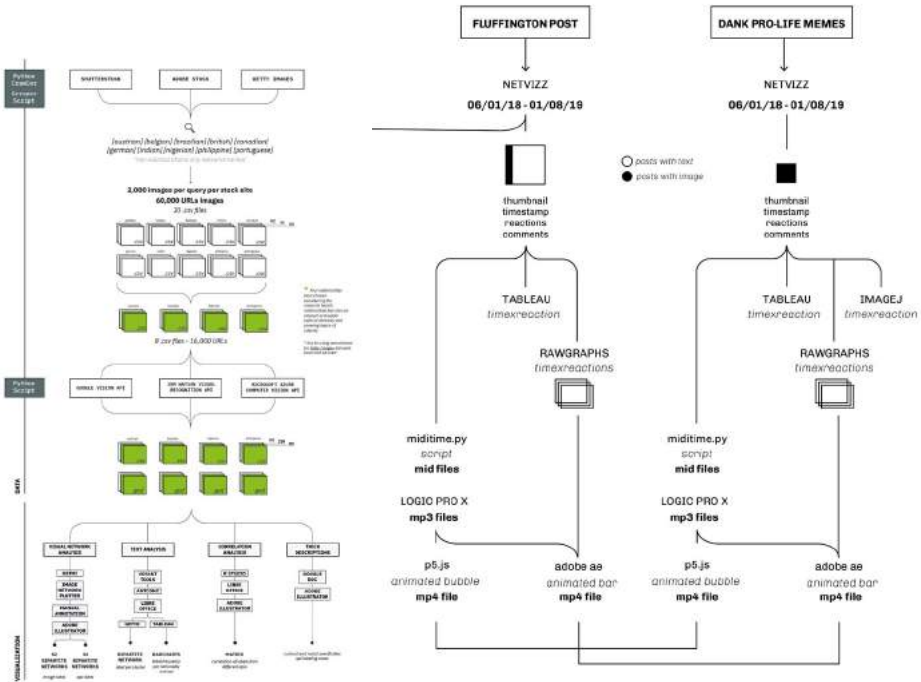


IMAGEM 4 – Um exemplo de protocolos visuais relacionados com dois projetos. Em alguns casos, o conjunto de dados inicial é único (imagem à esquerda) e, a partir daí, são projetadas mais visualizações. Noutros casos, os datasets iniciais podem ser diferentes (imagem à direita) e levar ao processamento de várias visualizações.

A reprodutibilidade do processo depende da clareza do artefacto e da precisão com que cada passo é descrito; pode ser repetível à letra, mas também, à semelhança da receita, modificado e adaptado às situações de investigação semelhantes, para as quais são fundamentais informações sobre falhas e dificuldades sem resolução, para não se deparar com situações semelhantes.

Acontece, na verdade, que protocolos já testados em edições passadas de Escolas de Verão/Inverno da iniciativa dos Métodos Digitais ou *Smart Data Sprints* são repropostos e adaptados a outros contextos.

CASOS DE ESTUDO

As linhas a seguir apresentam uma seleção dos resultados dos projetos desenvolvidos durante o mês de janeiro de 2019, por ocasião de dois *data sprints*.

A apresentação dos casos é baseada nos materiais presentes no relatório final, um documento importante que cada grupo de trabalho elabora durante o mês seguinte ao *sprint*. O relatório tem o papel fundamental de descrever o processo e os resultados da semana do *data sprint*, que pode representar o ponto de partida para futuros desenvolvimentos do projeto.

A análise dos estudos de caso será realizada desmontando as visualizações mais relevantes e comentando o processo da sua construção.

Em ambos os casos, o grupo era formado por um *designer* e por cinco a dez investigadores em ciências sociais, sociologia, tecnoantropologia, ciências computacionais e métodos digitais.

O primeiro estudo de caso “Understanding Normification. A Cross-Platform Analysis of the QAnon Conspiracy Theory” refere-se a um projeto desenvolvido durante a Escola de Inverno de Digital Methods Initiative (DMI Winter School 2019). Neste caso, é investigado como o conceito de *normification* – processo de normalização no qual os conteúdos *underground* de comunidades *online* de subculturas marginais viajam e são difundidas nas principais plataformas de redes sociais e notícias – podem ser aplicadas à teoria de conspiração QA da extrema direita, analisando a maneira como o conteúdo se espalhou desde as plataformas *underground* como a 4Chan e a 8Chan até às plataformas como YouTube ou Facebook.

O segundo estudo de caso o “Interrogating Vision APIs” propõe, em vez, usar os serviços de Vision APIs – ferramentas que também permitem integrar as funções Image Content Recognition em aplicações – como dispositivos de investigação e aplicá-los experimentalmente ao estudo da representação visual dos diferentes países, através dos resultados de *website* de banco de imagens (realizando a busca por “brasileiros”, “portugueses”, “americanos”, etc.).

UNDERSTADING NORMIEFICATION E O USO DAS VISUALIZAÇÕES METAFÓRICAS

Grupo de trabalho: Angeles Briones, Sal Hagen, Stijn Peeters, Birgitte Haanshuus, Emilija Jokubauskaite, Rachel Blennerhassett, Esther Blokbergen, Carmen Ferri, Flora Woudstra Hablé, Marlou Poncin, Willem Hilhorst, Ryan Tsapatsaris, Daniel de Zeeuw. Escola de Inverno da Iniciativa dos Métodos Digitais, 2019, Amesterdão.

Este estudo de caso foi escolhido como exemplo de uma colaboração efetiva entre *designers* e outros participantes, e também porque a visualização principal

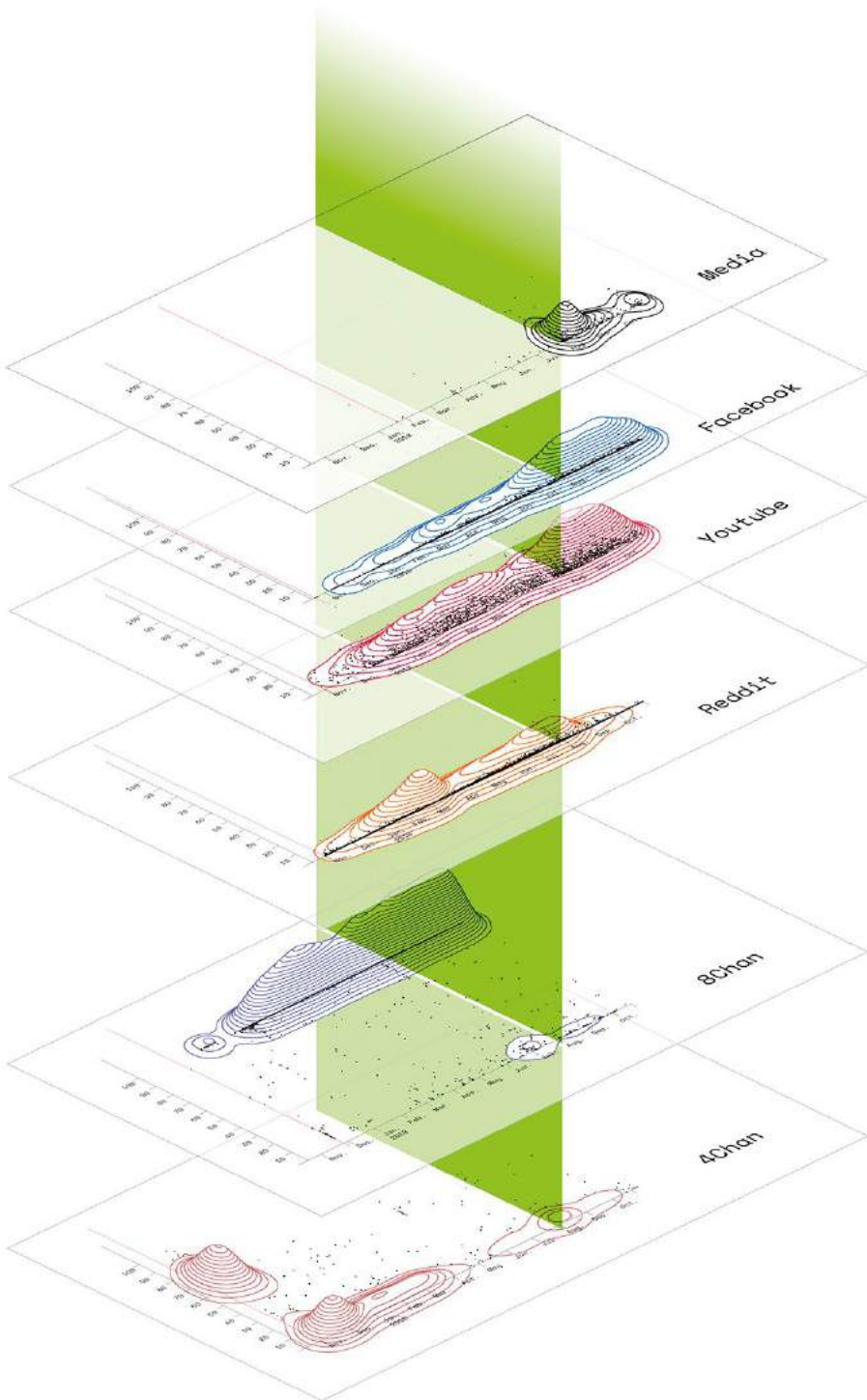


IMAGEM 5 – A difusão de conteúdo sobre a teoria da conspiração QAnon, entre 28 de outubro de 2017 e outubro de 2018, através do 4chan, 8chan, Reddit, YouTube, Facebook e *media online*. A barra verde indica o início da atenção dos *media*.

é o resultado da experimentação, na qual o *designer* adaptou um modelo visual conhecido às necessidades do projeto, utilizando uma linguagem metafórica apropriada ao tópico de investigação.

O processo de normalização do conteúdo é um processo *bottom up* no qual notícias e informações que vêm dos canais *underground* emergem e se espalham nos canais comuns, até chegarem aos *media* tradicionais (CNN, BBC).

O grupo de trabalho analisou e visualizou conteúdos relacionado com a teoria da conspiração do *QAnon* ao longo do tempo, em seis plataformas diferentes, desde os canais de notícias mais subterrâneos até os canais de notícias *online*, tentando identificar e mapear, se fosse possível, os métodos e os tempos do processo de normalização.

No relatório do *data sprint* (Hagen *et al.*, janeiro de 2019), a visualização final é o resultado de uma montagem composta por visualizações únicas, organizadas verticalmente e em axonometria, a fim de mostrar não apenas as relações entre elas, mas também a deslocação temporal de conteúdo de plataforma para plataforma.

Usando a metáfora da estratificação e do território, de facto, queríamos destacar a natureza emergente da normalização que, dos canais *underground*, muda ao longo do tempo para os canais *mainstream*.

A visualização é o resultado de uma reelaboração e disposição de vários elementos.

Na base da estrutura existem seis *contorn plot* – desenhados com *RawGraphs* –, modelos de dispersão visual, que mostram a difusão do termo *QAnon* ao longo do tempo nas plataformas analisadas.

Através de um programa de edição de imagens vetoriais, as visualizações foram deslocadas isometricamente e tornadas tridimensionais: a altura da montanha é proporcional à frequência do termo naquele período de tempo específico.

Cada plataforma representa um nível e a ordem de disposição é relativa à profundidade do canal. O eixo verde que penetra em cada nível indica o início da atenção dos *media* à teoria da conspiração.

O uso da metáfora da estratificação territorial para mostrar como os graus subterrâneos das diferentes plataformas são funcionais e exemplos de uma colaboração eficaz entre o *designer* e os outros membros do grupo. Além disso, a configuração da visualização, pensada para ser mostrada usando um *scrolling*, enfatiza o conceito de profundidade.

Como mostrado no relatório, neste caso, os membros do grupo de trabalho tiveram espaço para experimentar diversos modelos visuais produzindo outras visões, algumas das quais foram usadas para mostrar aspetos específicos dos *dataset* analisados, em apoio à visão axonométrica principal.

INTERROGANDO AS APIS DE VISÃO COMPUTACIONAL E AS REPRESENTAÇÕES DOS SEUS ESPAÇOS SEMÂNTICOS

Grupo de trabalho: André Mintz, Tarcizio Silva, Beatrice Gobbo, Elena Pili-pets, Hamdan Azhar, Helen Takamitsu, Janna Joceli Omena, Taís Oliveira. SMART Data Sprint 2019, Lisboa

Este estudo de caso foi escolhido porque fica claro, a partir do relatório, como a anotação manual dos produtos semiacabados levou à elaboração de uma visualização holística e à associação de arquétipos/*personas* com os serviços analisados através de Vision APIs.

Por espaços semânticos, entende-se a topologia que é formada pelas relações entre as etiquetas atribuídas às imagens pelas APIs de visão computacional. Com métodos de atribuição variáveis, estas etiquetas respondem às técnicas de aprendizagem automática das respetivas APIs de visão computacional de Google, IBM e Microsoft. O espaço semântico das relações formado pelas coocorrências dos

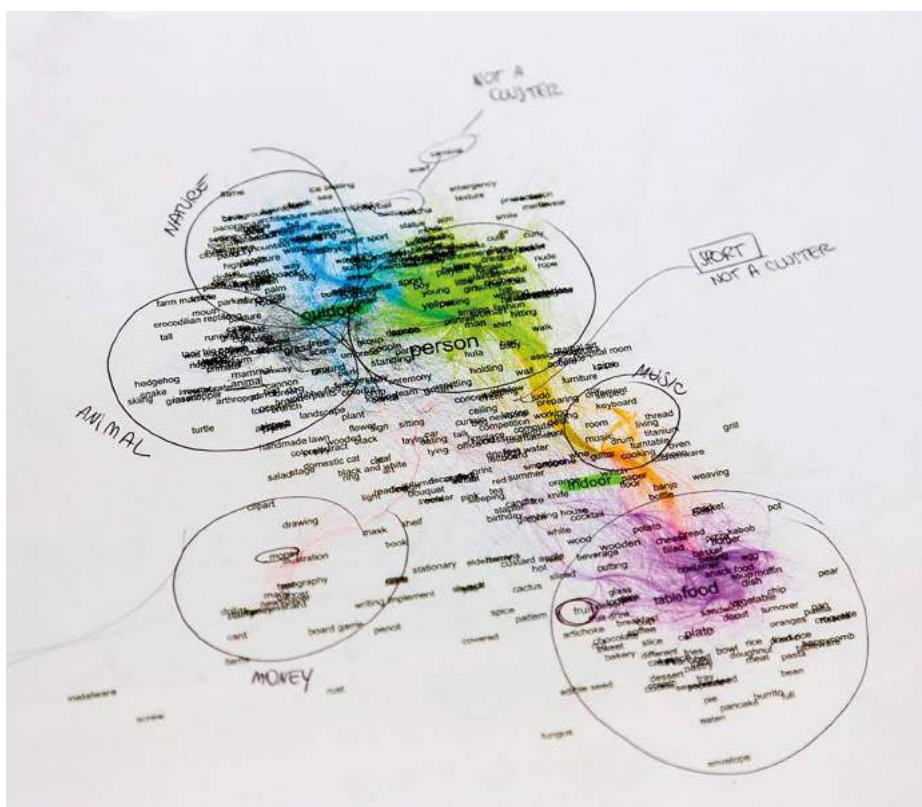


IMAGEM 6 – Exemplo de rede anotada. (Registo fotográfico: Ana Figueiras)

rótulos pode ser comparado. Depois de observar as diferenças e considerar o nível de detalhe oferecido por cada Vision API, chegou-se a descrevê-las com três arquétipos de *personas*.

Conforme descrito no parágrafo anterior, a anotação e leitura crítica dos semiacabados – utilizando, por exemplo, técnicas de análise de *visual network analysis* – estimulam a coprodução de conhecimento e levam a projetar outras visualizações e a elaborar novas questões de investigação.

A leitura crítica e anotação das primeiras redes semiacabadas de imagens e palavras revelaram diferenças de linguagem dos três serviços de visão computacional testados (Google, IBM e Microsoft). Embora reconhecendo *clusters* semelhantes nos mesmos grupos de imagens, os termos usados foram diferentes, em alguns casos, específicos, e noutros, muito genéricos.

Foi necessário encontrar uma maneira de comparar os espaços semânticos usando uma única visualização que pudesse destacar as diferenças e semelhanças nas línguas adotadas. Por esta razão, após uma série de tentativas e falhas, optou-se por utilizar a técnica de *small multiples* aplicada a uma rede (Gephi + Adobe Illustrator), destacando como o espaço semântico formado por todas as palavras e conexões descritivas do mesmo grupo de imagens, varia dependendo do serviço de Vision API adotado.

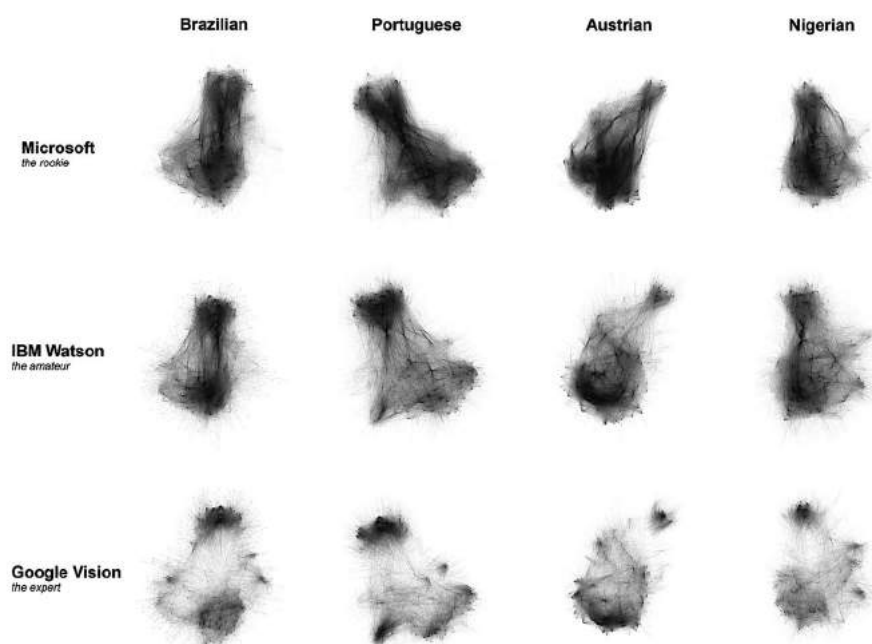


IMAGEM 7. – Matriz de comparação de espaços semânticos de cada Vision API.

Comparando cada elemento dentro da matriz, fica claro que, por um lado, as conexões criadas pelo Microsoft Azure ocupam sempre a parte central de cada grafo, pois usa uma linguagem genérica para classificar as imagens, enquanto as da API do Google Vision, mais detalhadas e específicas, se concentram nas áreas marginais da rede.

O uso de *small multiples*, dispostos numa matriz, mostra vários grafos, destacando ao mesmo tempo especificidades e diferenças, dando ao leitor a oportunidade de comparar rapidamente os comportamentos dos vários serviços de Vision API.

CONCLUSÕES

Neste capítulo foi descrito o papel do investigador em *design* no contexto de *data sprint*.

Foram feitas considerações sobre o papel do *designer* tanto na fase de pré-*sprint*, como de *sprint* e pós-*sprint*.

Em geral, o papel do *designer* pode ser resumido em três ações principais: informar, guiar e registrar.

Finalmente, foram apresentados dois estudos de caso como exemplos recentes de colaboração efetiva entre *designers* e investigadores com diferentes *expertises* e *backgrounds*.

REFERENCES

- Barker, P., Campbell, L. M., & Hawksey, M. (2013). *Writing in Book Sprints*, 1-7.
- Berry, D., Borra, E., Helmond, A., Plantin, J., & Walker Rettberg, J. (2015). *The data sprint approach: exploring the field of digital humanities through Amazon's Application Programming Interface (API)*. *Digital Humanities Quarterly*, 9(4).
- Colombo G., (2018) *The design of composite images. Displaying digital visual content for social research* (Tese de doutoramento) Instituto Politécnico de Milão, Itália..
- Bounegru, L., Gray, J., Venturini, T., & Mauri, M. (2017) *A Field Guide to 'Fake News' and Other Information Disorders*, Public Data Lab, Amsterdam. Retirado de <https://fakenews.publicdatalab.org/>
- Knapp, J. (2017). *SPRINT: Solve big problems and test new ideas in just five days*. Retirado de https://books.google.no/books?hl=no&lr=&id=ICSpCwAAQBAJ&oi=fnd&pg=PA1&ots=s4Vga1urQ4&sig=8Zos7e4cw202kvhqrnBoyUcPHaM&redir_esc=y#v=onepage&q&f=false
- Mauri, M., & Ciuccarelli, P. (2016). Designing diagrams for social issues. In *Proceedings of DRS2016: Design + Research + Society - Future-Focused Thinking* (Vol. 3, pp. 1-17). <https://doi.org/10.21606/drs.2016.185>

- Mauri, M., Colombo, G., Angeles, B., & Ciuccarelli, P. (2019). Teaching the Critical Role of Designers in the Data Society: The DensityDesign Approach. In N. A. G. Z. Börekçi, D. Ö. Koçyıldırım, F. Korkut, & D. Jones (Eds.), *Proceedings of DRS Learn X Design 2019: Insider Knowledge* (pp. 183-195). Ankara, Turkey: METU Department of Industrial Design. <https://doi.org/10.21606/LEARNXDESIGN.2019.12046>
- Munk, A. K., Madsen, A. K., & Jacomy, M. (n.d.). *Thinking Through the Databody: Sprints as Experimental Situations. Designs for Experimentation and Inquiry: Approaching Learning and Knowing in Digital Transformation*.
- Niederer, S., & Colombo, G. (2019). *Visual Methodologies for Networked Images: Designing Visualizations for Collaborative Research, Cross-platform Analysis, and Public Participation*. *Diseña*, 14, 40-67. <https://doi.org/10.7764/disena.14.40-67>
- Venturini, T., Munk, A., & Meunier, A. (2018). *Data-sprinting*. In & E. U. C. Lury, P. Clough, M. Michael, R. Fensham, S. Lammes, A. Last (Ed.), *Routledge Handbook of Interdisciplinary Research Methods* (1st ed., pp. 158-163). London: Routledge. <https://doi.org/10.4324/9781315714523-24>
- Hagen S., de Zeeuw D., Peeters S., Jokubauskaite E., Briones Á., Blennerhassett R., Ferri C., Woudstra Hablé F., Blokbergen E., Haanshuus B., Poncin M., Hilhorst W., Tsapatsaris R. (January 2019), *Understanding Normification*, <https://wiki.digitalmethods.net/Dmi/WinterSchool2019Normification>
- Mintz A., Silva T., Gobbo B., Pilipets E., Azhar H., Takamitsu H., Omena J.J., Taís Oliveira, (March 2019), *Interrogating Vision API*, <https://smart.inovamedialab.org/smart-2019/project-reports/interrogating-vision-apis/>
<http://lab.softwarestudies.com/p/imageplot.html>

*Tradução por Alda Rocha e Rosanna Carretta.



CINCO

**EXPLORAR
PRATICAMENTE,
NAVEGAR CRITICAMENTE**

PRODUÇÃO E CONSUMO DE VÍDEOS EM 360° — TENDÊNCIAS PARA O JORNALISMO BRASILEIRO NO YOUTUBE¹

ANA MARTA M. FLORES

O cenário contemporâneo do jornalismo vem passando por variadas e constantes mudanças que, para além das tecnologias aplicadas, impactam também as práticas jornalísticas. Em especial, destaca-se a diversificação nas formas de produzir e distribuir conteúdo noticioso nas décadas recentes. Há uma preocupação dos veículos jornalísticos em fidelizar e conquistar novos leitores e usuários frente à miríade de possibilidades — que inclui diversificação de produtos jornalísticos, novos modelos de negócio, disputa de espaço com as plataformas de mídias sociais, entre muitas outras. Com isso, é importante perceber os comportamentos das audiências para desenvolver estratégias que contribuam para uma relação relevante do jornalismo com seus públicos. Entender que o público não é mais apenas um mero leitor ou espectador, mas, sim, um usuário e um editor, alavanca mudanças mais profundas. No encontro desta atual conjuntura das audiências no jornalismo — mais participativas e aliadas às novas tecnologias que permitem maior interatividade — é onde situamos nosso estudo.

A inovação no jornalismo pode ser considerada um dos movimentos mais contudentes para a área na atualidade. Inovar surge em diferentes frentes dentro do jornalismo como uma resposta à dispersão de leitores de notícias. Dentre

¹ A aplicação e condução dos resultados da pesquisa foram realizadas pelo grupo de pesquisa SMART (Social Media Research Techniques), do iNOVA Media Lab, da Universidade Nova de Lisboa — também com o fundamental trabalho das investigadoras Janna Joceli Omena e Elsa Caetano, e do estudante de mestrado, Cristian Ruiz.

variadas expressões da inovação, o jornalismo imersivo vem sendo considerado um dos exemplos de maior impacto no setor (Marconi, Nakagawa, 2017; Newman et al., 2017, 2018; Doyle, Gelman, Gill, 2016; Google News Lab, 2018). Entende-se por jornalismo imersivo “a produção de notícias em uma forma em que os públicos podem obter experiências de eventos ou situações descritas nas notícias em primeira pessoa” (De la Peña et al., 2010, p. 291). Entre suas peculiaridades, ressaltase a ruptura do enquadramento tradicional e um ponto de vista itinerante que coloca o espectador em um cenário de 360° (Gracia & Damas, 2018).

Para observar e entender esse fenômeno propomos um recorte específico com o intuito de identificar os usos por veículos jornalísticos brasileiros e o consumo do vídeo 360° no YouTube. O YouTube é a plataforma mais popular de hospedagem de vídeos em nível global. Esta plataforma também permite o upload de conteúdos em 360° desde 2015. Os objetivos deste estudo são o de desenhar um panorama geral dessa tecnologia de vídeo na plataforma e o de identificar sinais que possam pontuar tendências emergentes no jornalismo. A relevância da pesquisa está em verificar se as projeções dadas por dossiês e estudos acadêmicos que correspondem às práticas aplicadas no campo, além da compreensão mais extensiva dos usos e consumos reais deste tipo de conteúdo. Com isso, em um outro momento, será possível pontuar indicadores que poderão ser úteis para a criação de estratégias aliadas à inovação no jornalismo contemporâneo. A perspectiva interdisciplinar adotada neste estudo procura entender como as tecnologias e seu uso no jornalismo podem inferir novos desejos e comportamentos de público, sinalizando tendências a serem exploradas em um futuro próximo.

A abordagem pretendida para este artigo entrelaça aos já mencionados Estudos de Tendências com os Digital Methods (Rogers, 2013) utilizando o jornalismo brasileiro como universo de estudo. Para atingir os objetivos propostos, observamos (i) a produção jornalística e (ii) o consumo de vídeos 360° durante o período de 1 de maio a 25 de junho de 2018. O primeiro (i) considera o vídeo em 360° de propriedade de canais jornalísticos, como Folha de S. Paulo, Valor Econômico, O Estado de S. Paulo, BBC Brasil, Nexo, Diário de Pernambuco e mais. O segundo (ii) reflete a função de pesquisa do YouTube e a cultura de classificação ou ranking culture (Rieder et al., 2018) — isto é, o que as pessoas encontram, assistem e consomem ao pesquisar conteúdo de vídeo na plataforma. Neste caso, o consumo não diz respeito a exclusivamente uma transação financeira, mas também ao acesso a conteúdos noticiosos. Com os resultados nestes dois âmbitos, propomos identificar quais conteúdos são de fato produções em vídeo 360° e classificar categorias temáticas. O propósito é compreender quais editorias fazem mais uso dessa tecnologia e quais resultados chegam aos usuários do YouTube.

BASES TEÓRICAS

Para fundamentar a proposta iremos traçar um percurso teórico que parte de conceituações interdisciplinares e importantes para o estudo. Iniciaremos delimitando o entendimento sobre tendência a partir dos Trends Studies (Gomes, 2016, 2017, 2018; Dragt, 2017; Rech, 2017, 2018; Gomes, Cohen & Flores, 2018). Assim como o termo “inovação”, “tendência” também tem um uso exaustivo e popularizado no cotidiano. Por isso, é importante definir qual a concepção de tendência para este estudo: uma tendência é, primeiramente, um reflexo da mentalidade que impacta as dinâmicas de consumo (Vejlgaard, 2008; Higham, 2009; Raymond, 2010; Gomes, 2016). Logo, circunscrevemos a ideia de inovação apropriada para o jornalismo na classificação do jornalismo de inovação. Em seguida, pontuaremos as peculiaridades da tecnologia de vídeo em 360° na expressão do jornalismo imersivo, considerando a plataforma do YouTube como principal canal de publicação e consumo nos dias de hoje². Por fim, o design interrogativo, tratamento e análise de dados extraídos através da API da plataforma serão guiados pela abordagem dos Digital Methods (Rogers, 2013).

Mais do que um produto ou um objeto, uma tendência é antes um padrão que se movimenta em um determinado sentido por meio de motivações específicas. De acordo com Dragt (2017), “os elementos-chave para a definição de uma tendência são entender a direção da mudança, os valores e necessidades associados a ela, as forças que afetam a tendência, as manifestações que podem ser observadas e a identificação dos grupos de pessoas onde a tendência surge e se desenvolve” (Dragt, 2017, pp. 36-37). Nesse sentido, os objetos e produtos são manifestações concretas das tendências. Um exemplo prático pode clarificar essa lógica: é possível inferir pelos pontos de mudança do consumo de jornalismo em uma experiência crescentemente mais participativa que, para além das tecnologias disponíveis, a motivação destes sinais – inquéritos online, participação pelas plataformas de mídias sociais, jornalismo colaborativo, newsgames, entre outros – são estratégias que envolvem a audiência em um contexto mais ativo e não de apenas leitor passivo. Entende-se que a audiência modifica a sua visão acerca do jornalismo quando participa da notícia ou reproduz uma experiência mais envolvente no consumo do noticiário. Consumir jornalismo sugere uma camada extra: a experiência da notícia.

² O YouTube é o segundo sítio com maior engajamento global e também no Brasil. Fonte: <<https://www.alexa.com/siteinfo/youtube.com>> (acesso em agosto/2019)

Idealmente, de acordo com as metodologias propostas nos Estudos de Tendências, a identificação de novos padrões em um determinado setor deve passar por uma série de etapas metodológicas. Entre elas, técnicas como i) desk research – pesquisa robusta de dados secundários com ênfase em fontes de mídia e estudos culturais, ii) coolhunting³ – observação direta de sinais emergentes de mudança, novos comportamentos e mentalidades que se mostram de maneira criativa e inovadora na sociedade e iii) painel de validação de tendências com especialistas. Neste capítulo, no entanto, buscamos uma avaliação exploratória do cenário contemporâneo do jornalismo com foco em sinais de mudança, a fim de identificar movimentos emergentes no jornalismo. A partir desses indícios no jornalismo aliados à resposta do público é possível propormos inferências e reconhecer padrões e, assim, definir tendências válidas para o setor jornalístico.

Para entender melhor a partir de qual perspectiva compreendemos o jornalismo contemporâneo tratamos de propor uma visada sobre o conceito de jornalismo de inovação. A proposta aqui sugerida não tem a pretensão de se sobrepor às tantas outras tipologias consideradas para o jornalismo nos dias de hoje, mas, sim, dar conta da abordagem aqui empregada. Pensar a atividade jornalística para além de suas formas tradicionais é uma prática recorrente no jornalismo. A pluralidade de audiências, comportamentos e tecnologias desafia os modelos clássicos de jornalismo de forma a incentivar a inovação em diferentes aspectos da atividade. Nesse sentido, expomos três instâncias em que o jornalismo contemporâneo já apresenta mudanças, seja como processo ou como produto: 1) conteúdo & narrativa, 2) tecnologia & formato e 3) modelos de negócio (Flores, 2017).

Nesse sentido, o jornalismo de inovação é aquele que necessariamente passa por modificações em técnicas, tecnologias, processos, linguagens, formatos e dispositivos destinados a potencializar a produção e o consumo da informação jornalística (Machado, 2010). Outro ponto importante para delimitar o jornalismo de inovação é entender que os processos básicos da prática jornalística – verificação, checagem, edição e circulação, por exemplo – são mantidos. No entanto, cada etapa desse processo pode sofrer alterações relacionadas à inovação. Quando identificamos o sinal emergente do jornalismo ao apropriar a tecnologia 360° no YouTube, entendemos que esta prática é focada na categoria 2) tecnologia & formato do jornalismo de inovação. Essa categoria destaca exemplos em que tecnologia é, de fato, a centralidade da inovação, perpassando os processos do conteúdo jornalístico e sua formatação para o consumo e divulgação.

³ O coolhunting é uma técnica baseada em observação, registro e previsão como parte da busca por tendências de mentalidades em uma abordagem coletiva (Gloor & Cooper, 2007).

Concordamos com Mark Briggs (2013) quando o autor considera a tecnologia e a cultura da inovação como não mais um diferencial, mas, sim, uma necessidade das organizações jornalísticas.

Entre as inovações mais aparentes no setor, destacamos o crescimento do jornalismo imersivo. Conforme pontuamos na introdução, o jornalismo imersivo é aquele que através da tecnologia de vídeo promove uma experiência de consumo inovadora, possibilitando uma espécie de vivência em primeira pessoa (De la Peña et al., 2010), muito além da leitura de um texto ou mesmo de uma reportagem em vídeo tradicional. No entanto, a imersão apresenta diferentes níveis para tecnologias específicas (Figura 1). Nesse sentido, é necessário apresentar uma breve distinção sobre o vídeo 360° e a realidade virtual (RV). O vídeo de 360° é uma tecnologia mais popular e não exige necessariamente dispositivos sofisticados e pode ser consumido sem óculos especiais, por exemplo.

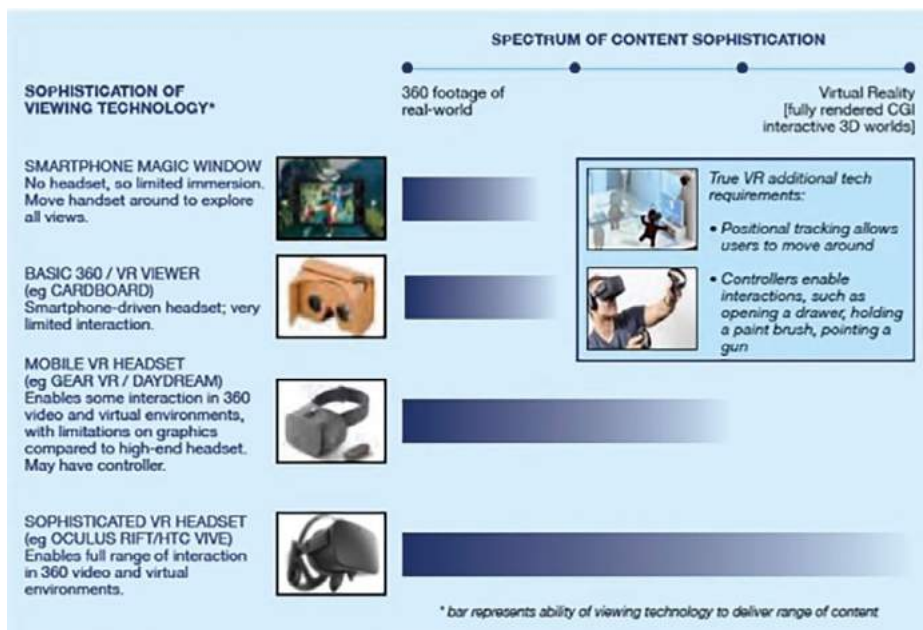


FIGURA 1 – Espectro de imersão para diferentes dispositivos tecnológicos (Watson, 2017, p. 4).

A introdução das plataformas Facebook 360 e YouTube 360 permitiu que as organizações jornalísticas publicassem vídeos com essa tecnologia sem investir em seus próprios reprodutores de vídeo (Watson, 2017). Por outro lado, a experiência completa em realidade virtual exige o uso de equipamentos próprios (high-end), como o Oculus Rift, o primeiro protótipo comercializado. As

primeiras produções jornalísticas com tecnologias relacionadas à RV começaram a ganhar destaque em periódicos já a partir de 2012. Alguns exemplos são o *The Des Moines Register* (*Harvest of Change*)⁴, *The Wall Street Journal* (*Is the Nasdaq in Another Bubble?*)⁵ e na BBC (*Calais “Jungle”*)⁶. Um marco dessa apropriação veio em novembro de 2015, quando o *The New York Times* distribuiu na edição do jornal impresso mais de um milhão de óculos de Google Cardboards⁷, na mesma data em que disponibilizou o aplicativo gratuito NYT VR⁸ e publicou o relatório *Displaced*, com vídeos esféricos produzidos com essa tecnologia. No Brasil, desde 2016, veículos como o jornal *Folha de S. Paulo*, a TV Record e a revista *Veja* também se aventuraram nessa estratégia.

Relatar eventos dando a impressão de que o leitor está no lugar exato onde o fato aconteceu é uma inovação nunca antes proporcionada na história do jornalismo. A própria ideia de “repórter jornalístico” vem da ação de reportar um acontecimento para um público que não estava lá presente. Com isso, o jornalismo imersivo vem sendo apontado como o principal investimento em jornalismo de inovação. Jake Silverstein, editor-chefe da *The New York Times Magazine*, comenta sobre o potencial da realidade virtual: “Nós nos interessamos pela realidade virtual quando vimos um filme em um campo de refugiados feito para a ONU. Mostramos para algumas pessoas na redação e todos ficaram impressionados. Editores endurecidos na seção internacional retiravam o fone de ouvido e diziam: ‘Ouça, eu editei centenas de histórias sobre refugiados e nunca tive uma experiência como essa’” (Silverstein, 2016, p. 20)⁹.

Como o jornalismo é uma atividade que tem como característica a popularização (Groth, 2011), o acesso à realidade virtual ou aos vídeos esféricos acaba por ter no YouTube um alcance maior e mais simplificado. O YouTube é um sítio de geração de conteúdo pelo utilizador (UGC, *user generated content*, em inglês), isto quer dizer que o conteúdo é criado pela iniciativa dos mesmos utilizadores que socializam na rede (boyd, 2008). Com isso, compreendemos o YouTube

4 Disponível em: <<https://www.desmoinesregister.com/pages/interactives/harvest-of-change>>

5 Especial em que realidade virtual guia uma viagem de 21 anos do pregão eletrônico Nasdaq Disponível em: <<http://graphics.wsj.com/3d-nasdaq>>.

6 Reportagens capturadas em 360° que mostram a rotina de Calais, um dos maiores campos de imigrantes da Síria na França. Disponível em: <<https://www.youtube.com/watch?v=9McdcF3CglE>>.

7 Google cardboards são óculos confeccionados em papelão, ímãs e lentes específicas, mas a custo acessível. O dispositivo é adaptável a diferentes smartphones que ao rodar aplicações compatíveis criam um efeito 3D pela distância das lentes.

8 Disponível em: <<http://www.nytimes.com/marketing/nytvr>>.

9 No original: “Listen, I’ve edited hundreds of stories about refugees, and I’ve never had an experience like this one.” (Silverstein, 2016, p. 20)

não como sinônimo de jornalismo imersivo, mas sim como uma ferramenta altamente válida para o jornalismo no sentido de que facilita a aproximação do público a conteúdos imersivos. Nesse ponto também não iremos problematizar as limitações decorrentes do efeito de “plataformização da sociedade”¹⁰ (Van Dijk, Poell & De Waal, 2018), e as relações cada vez mais imbricadas do jornalismo com as plataformas sociais. Google (aí inserido o YouTube), Amazon, Facebook, Apple e Microsoft (GAFAM) não exercem apenas um canal de distribuição, como se acreditava na primeira fase do contato do jornalismo associado a plataformas de mídias sociais. Já é muito claro que as plataformas determinam o que o público vê, quem é remunerado pela audiência, qual formato e gênero de jornalismo emplacam (Bell & Owen, 2017).

Para explorar as tendências de inovação no jornalismo, relacionadas a instâncias tecnológicas de imersão, consideramos relevante identificar qual tipo de conteúdo é apresentado ao público. Definimos que uma das formas mais diretas e visíveis é a própria caixa do motor de busca, em que o usuário procura a(s) palavra(s)-chave desejada(s) rastreando a plataforma. O retorno são centenas de opções, incluindo vídeos, canais, listas de vídeos, etc.¹¹ Algoritmos são capazes de filtrar, organizar e hierarquizar informações. Uma operação particularmente importante no YouTube, pois este processo se estabelece por um algoritmo dinâmico que define um conjunto de “variáveis de popularidade”, como quantidade de views, likes, busca por tema, entre outros critérios que determinam o ranking e hierarquia da informação desejada (neste caso vídeos ou canais), sendo alguns ranqueamentos mais dinâmicos do que outros a depender da temática procurada ou combinação de palavras-chave¹². Finalmente, os Digital Methods (Rogers, 2013) oferecem uma abordagem para entender como objetos nativos digitais podem ser utilizados como fonte e ferramenta de investigação ao mesmo tempo, não focado no desenvolvimento de suas funções, mas em como contribui para moldar a realidade, cultura e, no nosso caso, tendências de consumo.

10 Recomenda-se aprofundar neste conceito para ter uma visão mais macro do fenômeno estudado: Van Dijk, Poell, and de Wall, *The Platform Society: Public Values in a Connective World* (2018).

11 Mesmo não sendo foco da nossa pesquisa, vale mencionar que há linhas de investigação que não apenas estudam o objeto como tecnologia em relação ao indivíduo, mas também questionando o impacto dessa interação na sociedade. Hargittai (2004) estuda o impacto das empresas da indústria dos web engines na forma em que o conteúdo é organizado apresentado e distribuído aos utilizadores e Tavani (2005), Hinman (2005) e Zimmer (2006) exploram as conseqüências dessas ferramentas a nível de privacidade. Na abordagem de Estudos Culturais é sugerido que os motores de busca poderiam ter um impacto na construção do conhecimento (Wouters et al., 2004; Hellsten et al., 2006) tendo influências no mundo real (Edelman & Luca, 2014).

12 Ver mais sobre a caracterização destas morfologias em Rieder et al., 2018.

PROCEDIMENTOS METODOLÓGICOS: COMO PENSAR A PRODUÇÃO E CONSUMO DE VÍDEOS EM 360° NO YOUTUBE

Organizamos essa seção a partir das perspectivas do processo aplicado na presente pesquisa. Inicialmente, conforme explicitado, unimos propostas e visões interdisciplinares oriundas dos Estudos de Tendências e dos Métodos Digitais para embasar nossos questionamentos. Em seguida, analisamos um corpus de conteúdo do YouTube por meio da abordagem de pesquisa proposta pelos Digital Methods, utilizando o jornalismo brasileiro como nosso objeto empírico. Para cumprir o objetivo proposto, analisamos (i) a produção jornalística e (ii) o consumo de vídeos 360°. O primeiro (i) considera apenas os vídeos em 360° de propriedade de canais jornalísticos. O segundo (ii) reflete a função de pesquisa do YouTube e sua cultura de classificação.

Quanto à (i) produção jornalística, consultamos a API do YouTube V3 por meio do aplicativo YouTube Data Tools (Rieder, 2015), especialmente o módulo de rede de vídeos (video network) (Figura 2) – que proporciona a leitura do sistema de recomendações de vídeos da plataforma a partir de palavras-chave ou ids de vídeo.

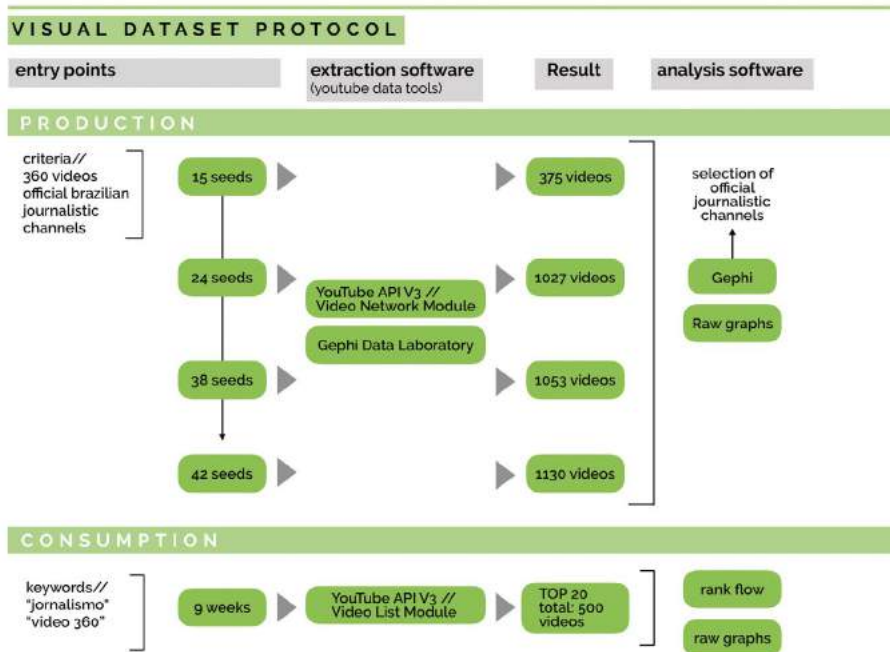


FIGURA 2 – Processo para coleta de dados em (i) produção e (ii) consumo (Flores, Omena, Caetano & Ruiz, 2018).

A seleção dos vídeos foi baseada em dois critérios: vídeos produzidos no formato 360° necessariamente por canais oficiais de veículos jornalísticos brasileiros. Com isso, primeiro geramos uma lista com 15 vídeos publicados pelos jornais (a) Folha de S. Paulo e (b) Estadão e pela (c) revista Veja. Depois de explorar a rede de vídeo no Gephi (Bastian, Heymann & Jacomy, 2009), detectamos novos vídeos que correspondem aos critérios de seleção – por meio de uma pesquisa exploratória no Laboratório de Dados Gephi. Consultamos novamente a API do YouTube V3 e obtivemos um total de 24 seeds. Em um terceiro momento, geramos uma lista adicional com oito vídeos de veículos jornalísticos em diferentes regiões brasileiras, a fim de tornar a amostra mais diversificada e reunimos vídeos 360° dos jornais (d) Valor Econômico, (e) O Dia, (f) Diário de Pernambuco e (g) Folha de Londrina, redes (h) BBC Brasil e (i) Globo, e dos nativos digitais, (j) Nexo e (k) UOL Tab. A terceira lista teve um total de 38 vídeos, contabilizando também seis seeds detectadas através do Laboratório de Dados do Gephi. A quarta lista reuniu 42 vídeos e, pela última vez, aplicamos a API YouTube V3 encerrando o processo de mapeamento de vídeos relacionados. Ao final, tínhamos quatro arquivos no formato .gdf – legíveis em software de visualização e análise de rede, como o Gephi. Após esse processo, reunimos todos os quatro arquivos de saída em Gephi, gerando uma rede de vídeos correlacionados (Figura 3). Considerando a quantidade de produção jornalística, a pesquisa mostra que os jornais que mais produziram conteúdo em 360°¹³ foram os jornais Estado de São Paulo, Folha de S. Paulo e revista Veja.

Considerando nosso estudo de caso focado em tendências para o jornalismo brasileiro, e tentando mapear a produção de vídeo no país de propriedade de veículos jornalísticos também brasileiros, removemos todos os nós criados por canais estrangeiros no YouTube ou não relacionados ao Brasil. Por exemplo, vídeos criados pelo canal SPORT TV – da televisão portuguesa e vídeos em outros idiomas, por exemplo, em inglês ou em espanhol.

Do ponto de vista do (ii) consumo de vídeos 360°, tomamos o mesmo software usado no estágio de produção, o YouTube DataTools (Rieder, 2015). Consultamos a API V3, porém desta vez para explorar como o consumo está a se manifestar através do mecanismo de busca da plataforma (Figura 2). Para alcançar esse objetivo lançamos o módulo videolist que tem a propriedade, dentre outras, de criar uma lista de vídeos recuperados a partir de termos específicos – como “jornalismo”. Os dados foram extraídos semanalmente durante os meses de maio e junho de

13 Vale ressaltar, no entanto, que a perspectiva avaliada neste trabalho é somente de vídeos publicados na plataforma do YouTube, ou seja, é possível que esses mesmos veículos usem domínios próprios para a publicação de conteúdos afins.

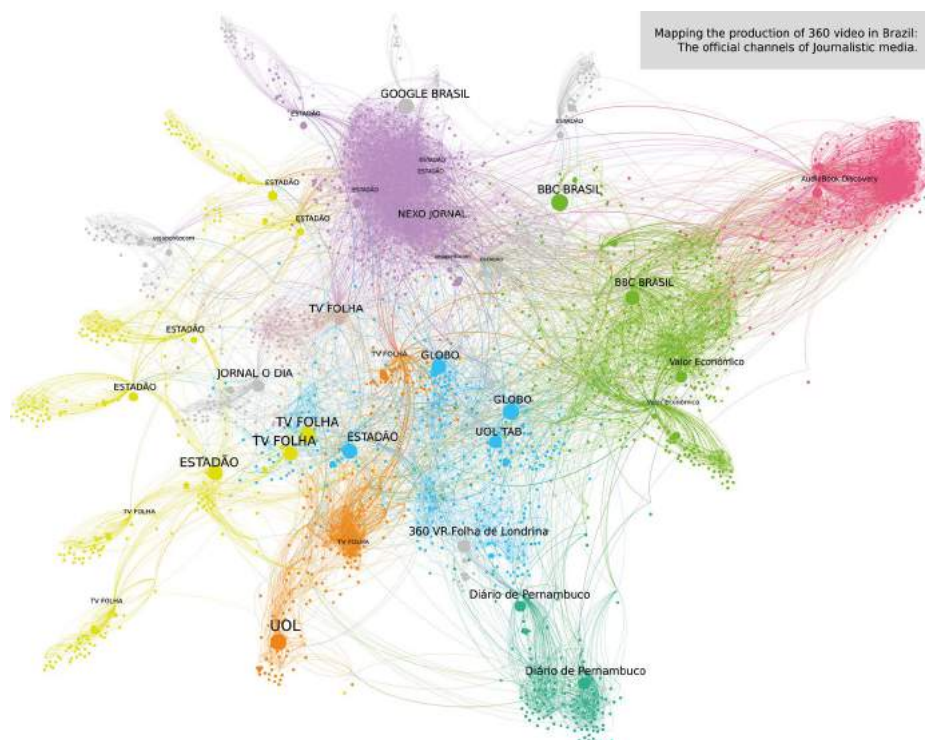


FIGURA 3 — Rede de produção de vídeos em 360° por canais oficiais jornalísticos brasileiros (Flores et al., 2018).

2018, utilizando duas palavras-chave: “Jornalismo” + “Vídeo 360”¹⁴. Restringimos a pesquisa ao Brasil usando o código de região (BR) com a intenção de reduzir as possíveis interferências de conteúdo estrangeiro. Para classificar os resultados, escolhemos o parâmetro “relevância” para a consulta de pesquisa do YouTube, uma vez que espelha mais eficientemente os resultados da plataforma do que data, classificação, título ou contagem de visualizações.

Das 500 entradas de vídeo (10 iterações) que foram recuperadas a cada semana, selecionamos os primeiros 20 resultados para análise. Assistimos a cada um destes 20 primeiros vídeos indexados e selecionamos o conteúdo que obedece aos critérios de análise: em 360° e de veículos jornalísticos brasileiros. Após essa primeira análise, reunimos os vídeos analisados em uma nova planilha usando o RawGraph (Mauri, Elli, Caviglia, Uboldi & Azzi, 2017) e também visualizamos

¹⁴ Exploramos outras combinações de palavras-chave, tentamos usar os mesmos termos em uma ordem diferente e também adicionamos “imersivo” e “VR” em nossos experimentos, mas optamos por não usar esses dados pois são termos mais técnicos e provavelmente pouco utilizados por pesquisadores-padrão.

o fluxo de classificação do conjunto de dados usando a ferramenta RankFlow (Rieder, 2016).

RESULTADOS

Os resultados da pesquisa apontam para um uso e apropriação do vídeo 360° ainda experimental nas redações brasileiras, além de sugerir uma relação distante entre os produtores e os potenciais consumidores da tecnologia. Isso se traduz, por exemplo, na ausência de canais jornalísticos de média na pesquisa da plataforma por relevância com os termos “jornalismo” e “vídeo 360”. Em vez disso, o retorno de busca do YouTube sugere uma série de vídeos acadêmicos criados por estudantes de jornalismo, vídeos experimentais ou de entretenimento publicado por youtubers. Também é válido mencionar que a maioria dos vídeos sugeridos como resultado da busca não é captada com a tecnologia 360°, conforme ilustrado abaixo (Figura 4). Ao visualizar o dataset dos resultados de busca durante oito semanas no RankFlow, confirmamos a baixa variação nos 20 principais resultados da pesquisa do YouTube com as palavras-chave: “vídeo 360” e “jornalismo”. Além disso, a maioria dos vídeos não é nem captado em 360°, nem jornalístico.

A fim de entender do ponto de vista da produção quais temáticas o jornalismo tem explorado a tecnologia 360° propomos uma classificação “jornalística” para os vídeos selecionados em (i) produção jornalística. Assim como os canais no YouTube, os vídeos lá publicados podem ser classificados separadamente pelo proprietário em categorias¹⁵ como Filmes & Animação, Notícias & Política, Entretenimento, Educação ou Pessoas & Blogs, por exemplo. Em nossa pesquisa exploratória, delimitamos esses mesmos vídeos em diferentes categorias, com ênfase no tema de cada vídeo em uma espécie de analogia às editorias de um veículo jornalístico. Em destaque surgem temas como Turismo, Saúde, Entretenimento, Política, Desastres Naturais, Esportes Radicais, Educação ou Cultura (Figura 5). Ao analisar os vídeos no período entre 2015 e 2018, observa-se que a categoria Cultura apresenta o maior crescimento independentemente do veículo jornalístico – de 0 em 2015, para 10 vídeos em 2018. Este dado é bastante consistente, já que os vídeos em 360° exigem uma forma de produção e edição mais específica, com planejamento mais elaborado, ideal para conteúdos não factuais.

¹⁵ Todas as categorias disponíveis pelo YouTube são: Film & Animation, Autos & Vehicles, Music, Pets & Animals, Sports, Travel & Eventos, Gaming, People & Blogs, Comedy, Entertainment, News & Politics, How to & Style, Science & Technology e Nonprofits & Activism.

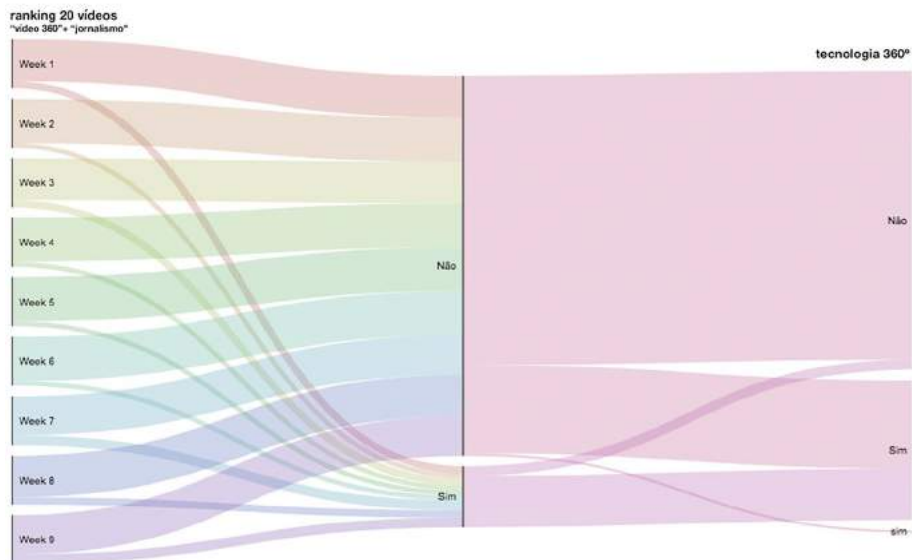


FIGURA 4 — Ao pesquisar “jornalismo” e “vídeo 360”, quais são os resultados obtidos? O que as pessoas assistem? Diagrama aluvial sobre a variação dos 20 primeiros resultados de busca no YouTube em maio e junho de 2018 e classificação dos vídeos com tecnologia 360° (Flores et al., 2018).

Do ponto de vista dos vídeos em 360° oriundos de veículos jornalísticos brasileiros, o pioneiro — dentro do escopo cronológico selecionado, é o jornal digital Nexo, iniciativa independente fundada pelas jornalistas Paula Miraglia, Renata Rizzi e Conrado Corsalette. O primeiro e único vídeo publicado pelo Nexo com essa tecnologia (até o fim da coleta de dados, em 2018) é classificado como Educativo pelos critérios qualitativos da nossa pesquisa (People & Blogs pelo canal) e representa uma espécie de meta-vídeo, em que um cenário digital é simulado para explicar exatamente como vídeos em 360° funcionam. É interessante notar que o primeiro conteúdo disponibilizado com essa tecnologia é introdutório, com a intenção de apresentar ao usuário essa nova possibilidade digital. Nos próximos três anos analisados, outros temas como Esportes Radicais, Turismo e Natureza ganham destaque e sugerem um uso mais significativo pelos veículos. Entendemos que tópicos como esses são altamente aderentes às possibilidades exploratórias da tecnologia 360°. A partir dos níveis de imersão, embora limitada pela navegação pelo browser, a experiência de consumir vídeos sobre natureza, turismo ou esportes extremos indica ser entendida pela mídia como mais atraente para a sua audiência.

Por outro lado, quando comparamos os temas com as métricas de visualização, os resultados são surpreendentes: a categoria política tem o maior crescimento

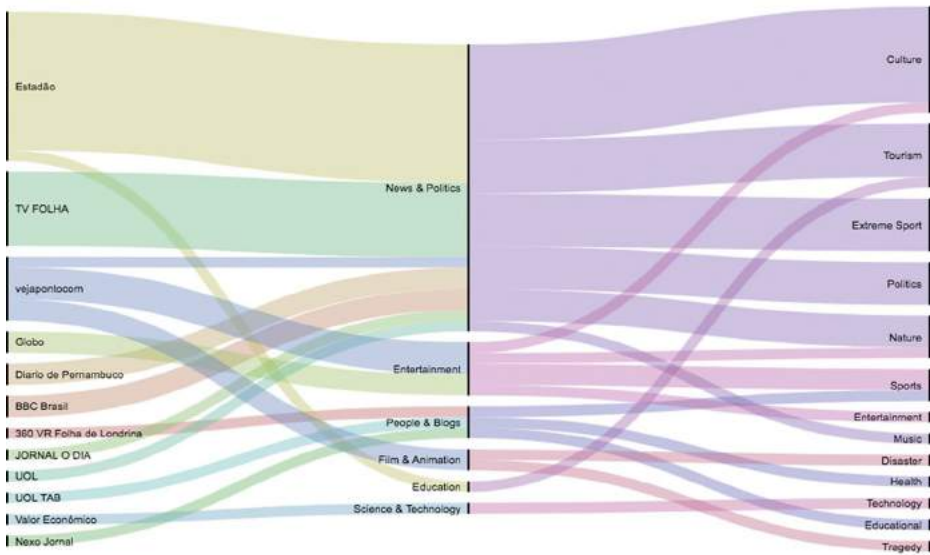


FIGURA 5 — Quais são os media jornalísticos brasileiros que utilizam a tecnologia de vídeo em 360°? Quais categorias foram mais exploradas nos últimos três anos? Gráfico com relação entre veículos jornalísticos e as temáticas que mais publicam conteúdos em 360° (Flores et al., 2018).

entre todas as já mencionadas (Figura 6). Em 2015, a categoria política sequer pontua, uma vez que identificamos apenas um vídeo educacional. No entanto, ao considerar conteúdos entre 2016 e 2018, a categoria política acumula um total de 269 259 visualizações para apenas cinco vídeos, enquanto Cultura, soma apenas 16 516 visualizações para o dobro de vídeos publicados no período. Ainda mais interessante é o fato de que o ano de 2018, avaliado até a data da coleta, em 22 de maio, corresponde a um maior número de visualizações que 2016 e 2017, que trazem dados do ano inteiro. O contexto brasileiro de efervescência política certamente contribui para uma busca maior de vídeos sobre o assunto, embora a produção jornalística com essa tecnologia não seja proporcional ao aparente interesse do público.

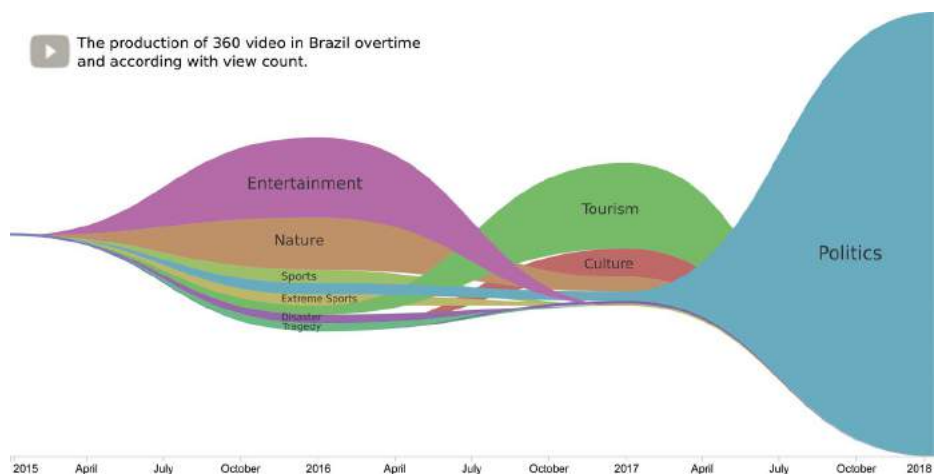


FIGURA 6 — O que, de fato, as pessoas assistem? Treemap proporcional à visualização de temáticas dos vídeos em 360° produzidos por veículos jornalísticos entre 2015 e 2018 (Flores et al., 2018).

Quanto ao retorno da busca, em (ii) consumo, ao selecionar somente os vídeos com tecnologia 360°, identificamos uma série de critérios que podem contribuir para uma noção de gramática do vídeo jornalístico captado em 360°. A ideia é perceber melhor parâmetros para avaliar os usos desta tecnologia que variam do experimental ao especialista. Entre os critérios, destacamos quatro: 1) roteiro adaptado para imersão; 2) roteiro guiado por texto ou áudio; 3) vídeo oferece experiência exploratória e 4) vídeo explora o senso de presença do consumidor. Com exceção do critério 2) em que a resposta alterna entre “texto”, “áudio” ou “nenhum”, esses parâmetros, embora certamente com variáveis, foram limitados a “sim” e “não”. As respostas positivas apontam para conteúdos jornalísticos que souberam melhor explorar as potencialidades das técnicas imersivas (Figura 7).

Com isso, chegamos ao resultado de que a maioria dos vídeos produzidos pela mídia jornalística no Brasil, entre dezembro de 2015 e abril de 2018, não se apropria de estratégias narrativas específicas para imersão (Figura 8).

Em cinza assinalamos os vídeos que não exploram as possibilidades narrativas do vídeo 360°, enquanto em verde os veículos em que identificamos características narrativas mais adaptadas. Interessante perceber que esses mesmos vídeos que se apropriam das características identificadas correspondem a um maior índice de likes e views na plataforma. A visualização demonstra a produção dos vídeos 360° ao longo dos anos (2015-2018), percebemos que apesar da larga produção no ano de 2016, a minoria dos vídeos responde à gramática imersiva de um vídeo 360° (veja.com). Sugere, de certa forma, um caráter mais experimental

de apropriação da tecnologia imersiva pelos canais jornalísticos. A partir de 2017, cresce as ofertas que respondem às características narrativas dos vídeos 360°.

CONSIDERAÇÕES FINAIS

Nossa motivação inicial ao conduzir os experimentos surgiu principalmente da impressão sutil de discordância entre os relatórios que avaliam o futuro próximo do jornalismo contemporâneo e o mercado do jornalismo brasileiro. Dossiês internacionais publicados a partir de 2016 (Associated Press, 2017; Newman et al., 2017, 2018; Knight Foundation, 2016; Google News Lab, 2018) afirmam que a imersão no jornalismo através de tecnologias como vídeos esféricos e em 360°, realidade aumentada e realidade virtual deveriam ser o grande investimento no setor. Com isso, procuramos entender por meio do YouTube qual o panorama geral dessa tecnologia. Mais além, também buscamos identificar sinais de tendências emergentes no jornalismo. Os resultados confirmam a suposição inicial de que há, de fato, uma certa divergência entre as pesquisas e a esfera prática do jornalismo no cenário brasileiro¹⁶. Ainda assim, observamos pontos muito interessantes acerca dos usos da tecnologia de vídeo em 360° sob ambos os aspectos avaliados: seja na produção, seja no consumo.

A partir da perspectiva da produção, entendemos que o estágio atual de apropriação do vídeo 360° é ainda experimental no jornalismo brasileiro durante o período avaliado, entre 2015 e 2018. Grandes corporações jornalísticas do Brasil fazem um uso aplicado ainda tímido e restrito, e quando o fazem pouco usam uma narrativa adaptada para a exploração esférica em que o utilizador está no centro. Um exemplo pontual é o vídeo de uma manifestação popular publicado pela Folha de S. Paulo¹⁷ em que o usuário apenas é “colocado” no centro de uma das avenidas mais importantes do país, entre manifestantes em som ambiente, sem nenhuma história contextual a ser contada. É possível que em temas mais polarizados como a política, ao deixar que o próprio usuário tome suas conclusões, o posicionamento do jornal seja tentar ser “neutro”. No entanto, em outro exemplo de narrativa pouco adaptada às potencialidades do vídeo em 360° e que não tem essa abordagem política se dá com o conteúdo¹⁸ publicado pelo jornal

16 Vale ressaltar, todavia, que os relatórios são focados no mercado internacional, especialmente Estados Unidos e Europa, embora também incluam o Brasil.

17 “Veja vídeo 360° da manifestação na av. Paulista”. Disponível em: <<https://www.youtube.com/watch?v=Cpoor-Wlk8w>>.

18 “Estadão 360: MASP”. Disponível em: <<https://www.youtube.com/watch?v=ijtBZRk-kCg>>.

O Estado de São Paulo, em uma visita ao Museu de Arte MASP. Nele, somente um passeio entre as obras (que mal podem ser apreciadas) e o público é proporcionado com uma trilha sonora instrumental. Embora haja liberdade para a audiência transitar e explorar os ambientes de uma história, entendemos que o jornalismo ainda deve ter o papel de guiar o utilizador para que ele compreenda o contexto no qual está imerso. Neste ponto observamos também um nível introdutório do que é considerado Tecnologia & Formato das categorias de jornalismo de inovação (Flores, 2018).

Complementarmente, em vídeos mais recentes publicados a partir de 2018, notamos uma adaptação mais apropriada ao que entendemos como uma espécie de gramática do vídeo imersivo. É o caso da peça jornalística publicada pela revista *Veja*¹⁹ sobre uma vila que deixou de existir depois do desastre ambiental de uma mina de rejeitos na cidade de Mariana, em Minas Gerais. Nesse vídeo, há uma preparação para o consumidor de uma proposta de trajeto, o uso indicado dos auriculares, um contexto claro de onde ele está – as ruínas de uma escola – e como o utilizador pode perceber os depoimentos, a narração e o som ambiente como seus guias. Outra aplicação adequada está no vídeo publicado pelo UOL TAB²⁰, que trata de um tema delicado como a depressão e suas sensações em uma peça que flerta também com a ficção, na seleção de atores e roteiro ao mesmo tempo que cumpre seu papel informativo. Nesse ponto, então, destacamos que os usos, conforme já identificados na análise, se dão em editoriais e em pautas mais específicas. Entendemos que a tendência não se dá necessariamente no jornalismo diário ou nas peças jornalísticas de todo e qualquer tema. Observamos pela perspectiva da produção, que a tendência sugere ser melhor adaptada em temáticas como desporto, meio ambiente, saúde, viagens e entretenimento, com o objetivo de promover uma experiência de mais impacto para um tema de interesse pessoal dos utilizadores. Vale considerar sempre a linha editorial do veículo para que ela seja coerente aos temas tratados em uma audiência já segmentada.

A partir da perspectiva do consumo, os resultados mostram que existe ainda um desconhecimento (ou desinteresse) por parte do público brasileiro desta tecnologia e o seu uso jornalístico, pois a compilação de dados extraída durante as oito semanas mostrou uma variação mínima top 20 do ranking das buscas, sendo que quase todos os vídeos da amostra sequer eram vídeos em 360° ou jornalísticos. Além disso, surpreende como dentro do dataset, não existir a presença de

¹⁹ “Bento Rodrigues – A vila que deixou de existir”. Disponível em: <<https://www.youtube.com/watch?v=fVrWFzdVKyI>>.

²⁰ “[DEPRESSÃO EM 360°] Tudo dói e ninguém entende : o mundo pelos olhos de um depressivo”. Disponível em: <<https://www.youtube.com/watch?v=eebTZ9vs5Aw>>

conteúdo produzido pelas mídias aqui mesmo estudadas, fortalecendo a sugestão de que ainda há uma desconexão proporcional entre o produtor e o consumidor de jornalismo perante aos formatos inovadores.

Ao examinar os poucos vídeos 360° da amostra descobrimos que a maioria é de produção de estudantes de jornalismo e instituições de ensino. O que novamente permite-nos reforçar a ideia de que o público geral ainda não tem abertura ao consumo deste tipo de formato. Para superar tal desconexão, sugerimos procurar formas de acercar ao público com estratégias de promoção por parte dos produtores para gerar uma nova cultura do consumo de formatos de inovação.

REFERÊNCIAS

- Alexa. (2018). *YouTube traffic statistics*. Disponível em: <www.alexa.com/siteinfo/youtube.com>. Acesso: 05/11/2018.
- Bell, E.; Owen, T. (2017). *The platform press: how Silicon Valley reengineered journalism*. Tow Center for Digital Journalism. Disponível em: <<https://newsinitiative.withgoogle.com/google-news-lab>>. Acesso: 10/05/2018.
- Briggs, M. (2013). *What makes journalism 'innovative'? Lessons from this year's Scripps Howard Awards*. Disponível em: <<http://goo.gl/G5yLqf>>. Acesso: 12/01/2017.
- Bastian, M.; Heymann, S. & Jacomy, M. (2009). Gephi: an open source software for exploring and manipulating networks. *International AAAI Conference on Weblogs and Social Media*, 8 (2009), 361-362.
- boyd, D. & Ellison, N. (2008). Social Network Sites Definition, History, and Scholarship. *Journal of Computer-Mediated Communication*, 13, 210-230.
- De la Peña, N., Weil, P., Lobera, J., Giannopoulos, E., Pomés, A., Spanlang, B., Friedman, D., Sanchez-Vives, M., Slater, M. (2010). Immersive Journalism: Immersive Virtual Reality for the First-Person Experience of News. *Presence. Cambridge. Massachusetts Institute of Technology*. Vol. 19, N.º 4, p. 291-301.
- Domínguez, E. (2014). *Periodismo inmersivo: La influencia de la realidad virtual y del videojuego en los contenidos informativos*. Editorial UOC, Barcelona.
- Doyle P., Gelman, M. (2016). *Viewing the future? Virtual reality in journalism – key challenges*. Online: Knight Foundation.
- Dragt, E. (2017). *How to research trends – move beyond trend watching to kickstart innovation*. BIS Publishers: Amsterdam.
- Flores, A.M.M. (2017). Innovation Journalism: a multiple concept. In *Brazilian Journalism Research*. v. 13, p. 156-179. Disponível em: <https://bjr.sbpjor.org.br/bjr/article/view/970/914>. Acesso: 08/08/2019.
- Flores, A.M.; Omena J.J.; Caetano, E.; Ruiz, C. (2018). Exploring Trends for Journalism on YouTube. The production and consumption of 360° videos in Brazil. *IV Encontro Jovens Investigadores – A investigação em Ciências da Comunicação na era Digital*, 11 e 12 de outubro 2018, Universidade do Porto, Portugal.
- Gloor, P.; Cooper, S. (2007). *Coolhunting: chasing down the next big thing*. New York: Amacon.
- Google News Lab. (2018). *About*. Disponível em: <<https://newsinitiative.withgoogle.com/google-news-lab>>. Acesso: 18/08/2018.
- Gomes, N. P; Francisco, A. F. (2013). *Introdução aos Estudos de Tendências: Conceitos e Modelos*. Estoril: TRC – Trends Research Center/Beyond Universe.

- Gomes, N. P. (2016). Trends Management applied to Branding and Cultural Management. *E-Revista Logo*, v.5, n.1.
- Gracia, M. J. B., Damas, S. H. (2018). *Realidad virtual y periodismo inmersivo con vídeo 360°*. Análisis del uso de los medios de comunicación españoles durante el período 2015-2017. Universidad de Málaga Sociedad Española de Periodística.
- Lee, U., Liu, Z. & Cho, J. (2005). Automatic identification of user goals in web search. In *Proceedings of WWW '05*, p.391-400.
- Machado, E. (2010). Creatividad e innovación en el periodismo digital. In *Actas II Congreso Internacional de Ciberperiodismo y Web 2.0*. Bilbao: Universidad del País Vasco, p. 64-72.
- Marconi, F.; Nakagawa, T. (2017). *The Age of Dynamic Storytelling – A guide for journalists in a world of immersive 3-D content*. Associated Press: online.
- Maschio, T. (2017). *Storyliving: an Ethnographic Study of How Audiences Experience VR and What that means for Journalists*. Google News Lab: online.
- Mauri, M., Elli, T., Caviglia, G., Uboldi, G., & Azzi, M. (2017). RAWGraphs. *Proceedings of the 12th Biannual Conference on Italian SIGCHI Chapter – CHIItaly '17*. doi:10.1145/3125571.3125585
- Newman, N. et al. (2017) *Digital News Project 2017 – Journalism, Media, and Technology Trends And Predictions 2017*. Reuters Institute for the Study of Journalism: online.
- Rieder, B. (2015). *YouTube Data Tools*. Computer software. Vers. 1.08. N.p., 5 December 2017. Web. <<https://tools.digitalmethods.net/netvizz/youtube/>>.
- Rieder, B. (2016). RankFlow. Computer software. Disponível em: <<http://labs.polsys.net/tools/rankflow>>. Acesso: 10/07.2018.
- Rieder et al. (2018). From ranking algorithms to 'ranking cultures': Investigating the modulation of visibility in YouTube search results. *Convergence: The International Journal of Research into New Media Technologies* 2018, Vol. 24(1) 50-68 DOI:10.1177/1354856517736982
- Rogers, R. (2013). *Digital Methods*. Cambridge, MA: MIT Press.
- Rose, D. E., & Levinson, D. (2004). Understanding user goals in web search. *Proceedings of the 13th Conference on World Wide Web – WWW '04*. doi:10.1145/988672.988675
- Roy, M., & Chi, M. T. (2003). Gender differences in patterns of searching the web. *Journal of educational computing research*, 29(3), 335-348.
- Sergey Brin Lawrence Page (2019) The anatomy of a large-scale hypertextual Web search engine. *Computer Networks and ISDN Systems* Volume 30, Issues 1-7, April 1998, Pages 107-117
- van Dijck, J., Poell, T. & de Wall, M. (2018). *The Platform Society: Public Values in a Connective World*. Oxford University Press.
- Van Couvering, E. (2008) The History of the Internet Search Engine: Navigational Media and the Traffic Commodity. In: Spink A., Zimmer M. (eds) *Web Search. Information Science and Knowledge Management*, vol 14. Springer, Berlin, Heidelberg.
- Watson, Z. (2017). *VR for News: The New Reality?* Disponível em: <<http://www.digitalnews-report.org/publications/2017/vr-news-new-reality>>. Acesso: 03/10/2017.

CIRCULAÇÃO DE IMAGENS ENTRE PLATAFORMAS: DESAFIOS E EXPERIÊNCIAS COM UMA API DE VISÃO COMPUTACIONAL

ANDRÉ MINTZ
CARLOS D'ANDRÉA

O estudo da circulação entre plataformas (cross-platform) é um dos desafios contemporâneos das pesquisas vinculadas aos métodos digitais (Rogers, 2015). Caracterizados pela articulação entre a materialidade sociotécnica dos meios digitais e as práticas sociais que eles abrigam, os estudos baseados nos métodos digitais tendem a estudar, de forma isolada, cada plataforma de mídia social, o que em parte decorre da sua dependência das interfaces e dos dados disponibilizados pelas próprias plataformas (Venturini & Rogers, 2019). Neste sentido, evidencia-se nos últimos anos a necessidade de elaborar abordagens capazes de compreender, em uma perspectiva mais abrangente, as práticas sociais online, por exemplo observando aspectos pertinentes a mais de uma plataforma, como a circulação de conteúdos entre diferentes delas (Rogers, 2018).

Um dos modos de enfrentar esse desafio envolve a identificação de “marcadores de tráfego” – no inglês, traffic tags (Elmer & Langlois, 2013) – que atuem como indicadores da atividade entre as plataformas. Exemplos desse tipo de conteúdo identificável incluem hashtags, URLs e formações verbais recorrentes (expressões, slogans etc.). Embora esses marcadores possuam implicações específicas para diferentes plataformas, eles podem ser identificados e coletados em mais de uma delas e, portanto, servem como operadores para compreender práticas sociais que interligam plataformas. Neste capítulo, buscamos contribuir a esses esforços apresentando uma abordagem experimental que toma imagens

estáticas como um marcador possível de rastreamento da circulação de conteúdos entre plataformas.

Trata-se de um passo importante em vista da crescente relevância dos conteúdos visuais nas plataformas de mídia social. A bibliografia recente tem indicado essa tendência como uma “virada visual” (Faulkner, Vis, & D’Orazio, 2018; Gibbs, Meese, Arnold, Nansen, & Carter, 2015), uma “virada pictórica” (Niederer, 2018), ou como uma crescente “ubiquidade do visual” (Highfield & Leaver, 2016). Além disso, as imagens constituem marcadores significativos para o estudo de circulação entre plataformas porque suas particularidades lhes conferem uma identidade reconhecível em contextos e discursos distintos. Mais do que uma formação verbal, por exemplo, uma fotografia possui características informacionais que as tornam relativamente estáveis e rastreáveis em sua circulação.

Propomos, nesse sentido, uma abordagem experimental para descrever a circulação de imagens através de dados fornecidos por uma interface de programação de aplicações (API na sigla em inglês) de visão computacional. Mais especificamente, nos apropriamos do módulo de busca reversa de imagens (Web Detection) de uma das ferramentas comerciais lançadas nos últimos anos (API Cloud Vision da Google) para, ao mesmo tempo, melhor conhecer o conjunto de páginas web e de plataformas em que imagens foram publicadas e discutir as implicações epistemológicas do uso do serviço de visão computacional. Inspirando-nos na abordagem do “trabalho de campo técnico” proposto por Bernhard Rieder e colaboradores (Rieder, Abdulla, Poell, Woltering, & Zack, 2015), os experimentos metodológicos empíricos são abordados como pontos de partida para refletir sobre as oportunidades analíticas e limitações metodológicas derivadas da API empregada no estudo. Em consonância com o campo transdisciplinar dos Estudos de Ciência e Tecnologia (Science and Technology Studies, STS), alinhamo-nos ao princípio da simetria segundo sua atualização por Noortje Marres e David Moats (2015). Uma abordagem simétrica, no sentido proposto pelos autores, consideraria tanto a circulação entre múltiplas plataformas quanto as especificidades sociotécnicas das ferramentas e aplicações mobilizadas para seu estudo. Entre as implicações que emergem desse esforço, o tema da plataformação da web (Helmond, 2015) assume uma posição importante tendo em vista os desafios postos ao estudo pela arquitetura reclusa das plataformas, em especial quando comparadas aos princípios universalistas que guiavam a “web aberta” (Plantin, Lagoze, Edwards, & Sandvig, 2016).

Trazemos, aqui, uma versão adaptada de estudo que publicamos em outra oportunidade e que apresenta com mais detalhes as análises que serão referenciadas (d’Andrea & Mintz, 2019). Buscamos, na presente ocasião, sintetizar

alguns dos aspectos metodológicos mais importantes e apontar desdobramentos potenciais para estudos derivados. Com esse intuito, o capítulo estrutura-se da seguinte forma. A segunda seção busca discutir algumas das bases teórico-metodológicas do que desenvolvemos – principalmente, a abordagem “entre plataformas” dos métodos digitais (Rogers, 2018) e os estudos críticos de APIs (Bucher, 2013). Na sequência, a terceira seção discute aspectos das tecnologias de visão computacional, com foco na recuperação de imagens baseada em conteúdo, e as implicações de se utilizar no estudo um serviço comercial da Google. A quarta seção descreve brevemente desafios para a coleta inicial de imagens enfrentados na pesquisa anterior, como o processo de identificação de imagens publicadas no Twitter durante uma transmissão audiovisual ao vivo. A quinta seção volta-se para algumas das possibilidades analíticas propiciadas pela abordagem proposta, considerando a decomposição das URLs mapeadas em partes significativas para os estudos de circulação. A análise dos dados referentes a duas imagens sobre o evento midiático estudado nos permite apontar algumas especificidades de sua circulação em diferentes plataformas e em sites da “web aberta”. A sexta seção discute algumas das principais implicações da metodologia empregada, com destaque para os diferentes modos como as imagens nas plataformas de mídia social são indexadas pela API Cloud Vision. Apresentamos a noção de “horizonte de recuperabilidade”, proposta para nomear os limites difusos e contingentes do que pode ser observado por meio de uma determinada ferramenta de busca ou por dados gerados por APIs. Por fim, a conclusão sintetiza os pontos principais do capítulo e os possíveis desdobramentos.

UMA ABORDAGEM ‘ENTRE PLATAFORMAS’ BASEADA EM IMAGENS

Uma das contribuições principais dos métodos digitais (Rogers, 2015) é a proposição de estudar a vida social por meio de procedimentos e técnicas baseadas em especificidades dos meios digitais. Nos últimos anos, uma grande diversidade de métodos e ferramentas foram desenvolvidos para coletar e analisar dados online baseados nesse princípio. Contudo, a maior parte desses estudos tende a enfatizar as singularidades de plataformas individuais, isolando-as de uma ecologia mais ampla. De acordo com Rogers (2018), esta tendência resultou em uma predominância de “estudos guiados por APIs” que, condicionados à dependência dos dados fornecidos por cada plataforma, resultaram na consolidação de uma perspectiva “uniplataforma”. Portanto, há uma demanda por métodos e procedimentos capazes de “desdobrar as mídias sociais”, indo além do estudo isolado de

uma plataforma e “desenvolvendo técnicas para a análise de múltiplas plataformas que consigam ter sensibilidade ao meio” (Rogers, 2018, p. 105).

Rogers faz referência à proposição anterior de Elmer e Langlois (2013), que afirmaram que a transição ao paradigma da Web 2.0 e das plataformas demandava a superação do que chamam de “métodos HTTP”. Estes seriam baseados em entidades universais, como os hyperlinks, que até então eram tomadas como referência para estudar a conexão entre diferentes meios e contextos. Alternativamente, os autores notaram a importância de se identificar e rastrear “marcadores de tráfego” (traffic tags) definidos como “operadores que permitem a junção de múltiplos modos de organização e conexão de diferentes atores” (Elmer & Langlois, 2013, p. 50). Trata-se de uma noção abrangente que considera qualquer forma de conteúdo identificável que possa ser rastreado em diferentes plataformas. Marcadores de tráfego podem ser frases verbais, endereços IP, hyperlinks e números de identificação, entre outros objetos únicos identificáveis que possam ser transportados e traduzidos a diferentes plataformas e formatos.

Estudos empíricos recentes e proposições metodológicas que compartilham desta perspectiva têm concentrado seus esforços na análise de conteúdos “colinkados, intercurtidos e/ou intertagueados” (Rogers, 2018, p. 101) por meio de objetos verbais – ou alfabéticos – como palavras-chave, hashtags ou URLs. Burgess e Matamoros-Fernández (2016), por exemplo, incluem em seu protocolo de pesquisa a identificação de URLs do YouTube compartilhadas no Twitter com a hashtag #gamergate para mapear a “rede de relações” entre as duas plataformas. De modo similar, uma das análises conduzidas por Driscoll e Thorson (2015) enfocaram a circulação de vídeos relacionados com o movimento Occupy usando “URLs fixas” como chaves para integrar bases de dados “multiplataforma”.

Estudos entre plataformas centrados em imagens, contudo, são ainda recentes¹ e raros. Warren Pearce et al. (2018) trazem uma exceção ao se proporem a descrever as representações visuais da mudança climática segundo os vernaculares visuais de diferentes plataformas. Já a pesquisa coletiva liderada por Farida Vis e Olga Guriunova (2015) constitui outra referência fundamental ao considerar como a circulação da fotografia trágica de Alan Kurdi transformou o debate da guerra na Síria e da crise de refugiados no continente europeu. Nossa proposição é devedora destes esforços, mas é distinta em seu método e abordagem, pois se concentra no rastreamento da ampla circulação de imagens específicas entre diferentes plataformas e pela “web aberta”. Em um esforço similar, estudo

¹ Para um estudo anterior, veja Issue Animals: <https://web.archive.org/web/20190202224944/https://wiki.digitalmethods.net/Dmi/IssueImageAnalysis>

de Tarcízio Silva, Pedro Barciela e Pedro Meirelles (2018), utiliza-se de alguns dos mesmos procedimentos e ferramentas que mobilizamos aqui, em uma aplicação da metodologia para fins de monitoramento de fake news e desinformação.

Concentrar-se na circulação das imagens constitui uma importante demanda metodológica dos estudos entre plataformas, principalmente se considerarmos a crescente centralidade do visual nas práticas online contemporâneas. Por exemplo, Highfield e Leaver (2016, p. 49) indicam como essa “ubiquidade do visual” tem sido acompanhada e encorajada pelas affordances das plataformas. Plataformas centradas em imagens como YouTube, Instagram e Snapchat são importantes, neste aspecto, bem como a crescente priorização de conteúdos visuais no Twitter, Facebook e Whatsapp. Igualmente, montagens visuais e fotográficas (Shifman, 2014) hoje confundem-se com a própria noção de meme e tornam-se centrais para a compreensão do espalhamento de conteúdos pelas mídias sociais. Vale ressaltar também que, dada a qualidade não-verbal de fotografias e ilustrações, usá-las como marcadores de tráfego permite à pesquisa superar barreiras linguísticas, que constituem um desafio perene das pesquisas em internet. Nesse sentido, o conteúdo visual desponta como um possível marcador de tráfego para estudos entre plataformas, embora com notáveis desafios e implicações metodológicas.

VISÃO COMPUTACIONAL PARA A PESQUISA EM IMAGENS

Uma das barreiras iniciais das abordagens centradas na imagem para a pesquisa em internet é o desafio técnico do seu processamento em larga escala. Neste estudo, este esforço se faz necessário na “busca reversa” de figuras publicadas na web, o que demanda não apenas a indexação massiva de conteúdos online mas também técnicas eficientes de detecção de diferentes reproduções (idênticas ou quase idênticas) de uma imagem. Um problema fundamental deste desafio é a sobreposição de incertezas derivadas dos métodos computacionais em sua tradução entre a representação de dados de ‘baixo nível’ e suas interpretações, de ‘alto nível’. Por serem intransponíveis, as lacunas deixadas por essas incertezas fazem do desafio do reconhecimento de imagens um “problema mal-posto”, pois não há nunca uma resposta única e definitiva, mas apenas aproximações (Smeulders, Worring, Santini, Gupta, & Jain, 2000)². Golan Levin (2006, p. 468) chamou de

² Tipicamente, na computação indica-se por baixo nível formas de representação de dados que mais se aproximam de sua forma “bruta” ou concreta de registro – ilustrativamente: zeros e uns. Enquanto por “alto nível” refere-se a representações com maior grau de abstração – como as descrições verbais do que encontra-se na imagem, por exemplo.

opacidade do visual esta barreira ao processamento computacional de imagens, pois são necessárias várias camadas de inferência sobre os dados dos pixels para se viabilizar o tratamento de atributos visuais por algoritmos.

O desenvolvimento de tais camadas de inferência é o que objetiva a ampla subdisciplina das Ciências da Computação denominada visão computacional. Dado o volume de dados que tipicamente compõem suas bases empíricas, os desenvolvimentos da visão computacional são importantes recursos para a pesquisa em internet que toma imagens como seu tema principal de investigação. Neste capítulo, o desafio abrangente de processar imagens de forma computacional se restringe à tarefa de recuperação de imagens baseada em conteúdo (content-based image retrieval, CBIR). Em geral, trata-se de qualquer busca por imagens baseada em seu aspecto visual (ao invés de metadados), como acontece, por exemplo, na identificação de diferentes cópias idênticas de uma imagem em uma base de dados (Smeulders et al., 2000). Embora trate-se de uma tarefa mais específica do que outras desempenhadas hoje pela visão computacional, ainda assim se trata de um processo complexo. Em linhas gerais, a CBIR demanda a geração de descrições computacionais das imagens buscadas, enquanto se prioriza aspectos invariantes a modificações que, a depender das necessidades da aplicação, devem ser desconsideradas — tais como defeitos de compressão e algum nível de edição da imagem (balanço de cor, cortes pequenos, sobreposição de conteúdo etc.). Embora muitas dessas tarefas tenham sido realizadas, no passado, por meio do desenvolvimento de sistemas especialistas (Smeulders et al., 2000), as tendências da última década apontam para uma aplicação crescente de sistemas probabilísticos de base conexionista (Cardon, Cointet, & Mazières, 2018) de aprendizado de máquinas (Wan et al., 2014).

Contudo, a aplicação de CBIR à web, como se busca fazer aqui, implica desafios adicionais. Se, tipicamente, a recuperação de imagens é restrita a bases de dados mais estáveis, buscar uma imagem na web significa trabalhar com uma base de dados que é um índice dinâmico e crescente de páginas e imagens publicadas constantemente. Esse tipo de busca tem sido realizada, ao menos nos últimos dez anos, por ferramentas como TinEye ou a função de “busca por imagem” da Google³. Ambas empresas oferecem interfaces gráficas para buscar ocorrências de uma imagem (ou outras similares). Embora boa parte da abordagem proposta aqui possa ser desenvolvida a partir destas interfaces, no estudo que realizamos

³ TinEye foi lançada pela empresa canadense Idée em 2008 (cf. https://web.archive.org/web/20160914171724/https://tineye.com/press/archived_news). A função de “busca por imagem” da Google foi lançada em 2011 (cf. <https://web.archive.org/web/20180129140614/https://techcrunch.com/2011/06/14/google-search-by-image-use-a-snapshot-as-your-search-query/>).

utilizamos a API Cloud Vision da Google. Em comparação com a função oferecida via interface gráfica de usuário (GUI, na sigla em inglês), o acesso via API possibilita automatizar parte do processo, realizando consultas periódicas programadas, bem como buscar por várias imagens de uma só vez, o que facilita a criação de protocolos de coleta e monitoramento. Lançada em 2015, a API da Google é parte de um conjunto de produtos similares lançados no período e que, em comum, oferecem recursos de descrição automatizada de conteúdos visuais⁴. O módulo ‘Web Detection’, que integra o produto da Google, realiza a tarefa de busca reversa das imagens em páginas da web e não possui tantas alternativas em outras empresas, embora TinEye, Microsoft e Pixsy ofereçam funcionalidades similares.

Na API da Google, a busca é realizada por meio da máquina de indexação massiva (Rogers, 2015) que constitui a Google como um agente-chave na organização e recuperação de informação na web. Além disso, a busca é um dos alicerces dos métodos digitais, tendo em vista a premissa do campo de estudar reflexivamente objetos nativos do digital através de recursos oferecidos por eles. Embora a empresa não ofereça documentação detalhada acerca do funcionamento da sua “busca por imagem” ou do módulo Web Detection, pode-se presumir que são ferramentas baseadas em uma aplicação das técnicas de CBIR a uma base de dados dinâmica construída pelos mecanismos de crawling da empresa. Portanto, o que é alcançado por meio da API Cloud Vision da Google é provavelmente ao menos tão restrito quanto a ferramenta tradicional de busca da empresa, com a complexidade extra dos mecanismos de correspondência entre diferentes reproduções de uma mesma imagem.

Um mecanismo de busca reversa tipicamente oferece como resultado uma lista de endereços URL para as páginas na web em que a imagem buscada foi indexada. Ou seja, recuperam-se múltiplas páginas que, em comum, abrigam uma cópia idêntica ou quase idêntica daquela imagem. Em cada momento em que tal informação é recuperada, temos algo como um instantâneo circunstancial da presença online da imagem, o que certamente varia com o tempo não apenas em função da efetiva circulação da imagem como, também, pela dinâmica de indexação da ferramenta. A API Cloud Vision em sua versão atual, distingue entre imagens idênticas e quase idênticas em sua indicação das páginas em que elas aparecem e, também, dos endereços URL das imagens correspondentes em

4 Entre as principais encontram-se as das empresas Microsoft, IBM, Amazon e Clarifai.

si⁵. Estas URLs, então, servem de substrato para a análise de processos de circulação, podendo ser abordadas de modo sincrônico ou diacrônico.

É crucial, contudo, não tomar a centralidade da Google como totalizante: seus resultados são indícios, mas não são a web. De outro modo, deve-se reconhecer sua condição mediadora. Parafraseando a caracterização proposta por Taina Bucher (2013) para a API do Twitter, a Google não é simplesmente um intermediário entre os usuários e a web, mas, sim, um agente de transformação e moldagem de ambos. A indexação da Google não é apenas incompleta, mas é também orientada a fins comerciais (Rieder & Sire, 2013), o que resulta em um viés particular aos resultados que não pode ser ignorado. Nesse sentido, os dados obtidos pela API devem servir a um olhar duplamente interessado. Por um lado, eles servem como indicadores, embora incompletos e enviesados, do rastreamento possível das imagens em circulação. Por outro, esses dados também oferecem uma janela para compreendermos ou, ao menos, nos indagarmos acerca das affordances epistêmicas da API da Google, que modulam as nossas possibilidades de acesso a esses dados. Voltaremos nesse ponto adiante.

COLETA DE DADOS E AS IMAGENS ANALISADAS

Antes de discutir as possibilidades analíticas em si, é importante fazer um breve comentário acerca de estratégias de identificação e coleta inicial de imagens cuja circulação entre plataformas será posteriormente estudada. Evidentemente, há muitos casos em que as imagens são já conhecidas dos pesquisadores devido à sua relevância em determinado contexto de investigação. Contudo, quando este não é o caso, a coleta precisará lidar com a centralidade do verbal enquanto forma inicial de busca nas diferentes plataformas. Tipicamente, as possibilidades de acesso mediadas pelas APIs e mesmo pelas GUIs são dependentes de termos de busca verbais (palavras, expressões ou hashtags) ou de identificadores alfanuméricos (relativos a contas de usuário, páginas ou localização geográfica). Portanto, a já complexa tarefa de escolha das chaves de busca (Rogers, 2017) complexifica-se ainda mais quando o objetivo central é centrar-se em imagens. O isolamento entre distintas plataformas é também

5 A API também retorna dois outros dados não utilizados aqui: URLs de imagens 'visualmente semelhantes', ou seja, que não têm relação com a imagem buscada senão por sua semelhança aparente; e a vinculação da imagem a entidades semânticas mapeadas pela Google em sua *rede semântica da web*, ou *Google Knowledge Graph*, que organiza uma espécie de ontologia da web a partir da rede de relações entre suas entidades.

reforçado pela própria configuração dos dispositivos de coleta, que dificultam integrações entre múltiplas plataformas. Seja via API ou via raspagem direta da GUI, cada plataforma constitui protocolos bastante distintos que demandam que uma investigação desenvolva tantos métodos de coleta quantas sejam as plataformas que se deseja investigar.

Os modos de enfrentamento desses desafios irão variar conforme o contexto da investigação. O objeto que investigamos em nosso estudo (d'Andrea & Mintz, 2019) foi a transmissão ao vivo do sorteio final das chaves da Copa do Mundo FIFA de 2018, que se realizou em 1.º de dezembro de 2017 em Moscou. O caráter internacional do evento sugeria o potencial de observação de dinâmicas transnacionais de circulação das imagens, enquanto o caráter 'ao vivo' da transmissão permitia o acompanhamento da circulação de imagens desde suas primeiras publicações. Como as imagens não eram ainda conhecidas por nós, era inevitável que nossa pesquisa enfrentasse os desafios pertinentes à coleta inicial de imagens.

Portanto, orientamos nossa estratégia tomando o Twitter como plataforma para a coleta inicial, tendo em vista a sua reconhecida utilização por fãs de esportes e o seu uso por audiências televisivas de modo articulado com transmissões audiovisuais. Foi considerada também a relativa permissibilidade da API Streaming do Twitter, a qual pudemos acessar por meio da ferramenta DMI-TCAT (Borra & Rieder, 2014). Para lidar com a definição dos termos de busca, fizemos um amplo levantamento anterior e selecionamos 47 termos e hashtags distribuídos entre 17 idiomas distintos, com prioridade para aqueles dos países com representação no torneio. Com base nos tuítes obtidos, utilizamos a função "Media frequency", integrada ao DMI-TCAT para extrair uma lista URLs de imagens mais retuitadas durante o evento e nas duas horas subsequentes. Estas imagens foram então submetidas simultaneamente à API Cloud Vision da Google em diferentes momentos de modo a recuperar rastros da circulação ao longo do tempo⁶. Desse modo, obtivemos dados de circulação de cinco momentos para cada imagem ao longo de três semanas após a realização do sorteio da Copa.

⁶ Utilizamos para este processamento o script Memespector (Mintz, 2018; Rieder, Den Tex, & Mintz, 2018).



FIGURA 1 – Duas das imagens consideradas para a análise de circulação.

Do conjunto de 219 imagens submetidas ao processamento, destacamos neste capítulo duas sobre as quais nos voltamos em nossa análise (Figura 1)⁷. Uma é uma fotografia de bastidores publicada pelo ex-jogador brasileiro Cafu em sua conta no Twitter, em que ele aparece ao lado do também ex-jogador Pelé, entre outros atletas e autoridades. Sua escolha se deveu à grande circulação da fotografia, que obteve o maior número de republicações entre imagens fotográficas coletadas no Twitter no período observado. Além disso, trata-se de imagem produzida no evento, portanto sem circulação anterior. A segunda imagem é um meme pertinente ao contexto latino-americano e que faz piada com a desclassificação da equipe chilena nas eliminatórias do torneio. O meme baseia-se em uma fotografia de 2015 do ex-jogador italiano Cannavaro. Dois aspectos singularizam esta imagem. Primeiro, a pertinência regional é dada não apenas pelo tema como também pelo idioma do conteúdo verbal (espanhol), configurando um potencial de circulação em regiões específicas do mundo. Segundo, trata-se de um meme fotográfico construído sobre uma imagem que já circulava antes da realização do evento. Esta circulação anterior traz aspectos particulares para o mapeamento da circulação.

⁷ Em d'Andrea & Mintz, 2019, duas outras imagens foram analisadas: uma outra foto de bastidores, desta vez produzida e divulgada pela FIFA, e uma foto-montagem com textos em árabe e que remetem a um dos jogos definidos no sorteio (Egito X Arábia Saudita)

OPORTUNIDADES ANALÍTICAS

Por meio do protocolo de coleta e processamento descrito acima, obtivemos, para cada imagem, listas de URLs das páginas em que a imagem foi encontrada em diferentes momentos das três semanas analisadas. Uma primeira abordagem possível dos dados seria a análise detida destas múltiplas páginas em que a imagem apareceu. Nessa abordagem teríamos, plenamente, a imagem operando como um “marcador de tráfego” como sugerimos inicialmente. Enquanto elemento comum a estas diferentes páginas web e postagens em plataformas de mídia social, a imagem surgiria como elemento de interconexão entre diferentes conteúdos e diferentes plataformas. Os casos particulares de sua aparição poderiam ser, assim, cotejados, revelando diferentes contextos de circulação e diferentes formações de sentido para uma mesma imagem. Contudo o volume de URLs identificadas configura um desafio: ainda que se enfoque uma “única” imagem, milhares de URLs foram recuperadas nas consultas à API de visão computacional. Além disso, em um contexto de circulação internacional como o que observamos em nosso estudo de caso, encontramos páginas em múltiplos idiomas e contextos impermeáveis às nossas capacidades linguísticas e culturais localizadas.

Devido a estas barreiras e ao interesse mais marcadamente metodológico de nossa investigação, não realizamos uma análise caso a caso das páginas identificadas. Voltamo-nos para um trabalho com os dados agregados, isto é, subsumindo a individualidade das ocorrências em favor de formas de considerá-las conjuntamente. Em nosso estudo, desenvolvemos dois modos de fazê-lo, tomando por base a estrutura protocolar das URLs⁸. Primeiramente, em vista de nosso interesse na perspectiva da circulação entre plataformas, identificamos os domínios pertinentes às plataformas de mídia social conhecidas. Em um segundo momento, tendo em vista principalmente o conteúdo de páginas da web (sites de notícia, portais, blogs), voltamo-nos a indícios de sua distribuição geográfica pelos “domínios de topo” (TLD na sigla em inglês).



FIGURA 2 — Decomposição da estrutura das URLs.

⁸ Para operacionalizar esta análise, utilizamos um script de elaboração própria com recursos do módulo de Python *Tldextract* (Kurkowski, 2019).

Para compreender esse modo de análise, primeiro é preciso compreender as partes componentes de uma URL e os sentidos que podemos atribuir a elas. A Figura 2 ilustra a decomposição destas partes. Como forma de identificação das plataformas em que as imagens circularam, observamos especificamente o componente B, relativo ao “domínio”. Esta é a parte mais fundamental de uma URL e depende de registro pelo proprietário da página no órgão competente. Um mesmo domínio pode conter múltiplos subdomínios a critério do proprietário do registro. O domínio também é a entidade que concentra os diferentes caminhos (paths) de páginas a ele pertencentes. Contudo, os domínios são sempre específicos aos TLDs, segundo a categoria do site e sua subsunção a uma localidade geográfica.

Em nosso primeiro esforço, tomamos o domínio como identificador das plataformas de mídia social em que as imagens foram encontradas. O componente C, por sua vez, relativo ao TLD, foi tomado como um indicador possível da relevância geográfica da página. Importante dizer, trata-se de um indicador impreciso dado ser comum que páginas se registrem em TLDs pertinentes a outras localidades, seja como forma de simplificar o endereço (.com ao invés de .com.br, por exemplo); para escapar da regulação de órgãos de governança de certas localidades; ou, ainda, para aproveitar uma conjunção temática advinda do TLD (como o .fm, relativo à Micronésia, mas utilizado por rádios).

Para identificar as ocorrências em plataformas específicas, portanto, agrupamos as URLs obtidas segundo seu domínio e buscamos aquelas relativas a uma lista de 17 plataformas compiladas a partir de levantamentos que encontramos online. Para as duas imagens referidas aqui, encontramos ocorrências em 11 plataformas, conforme indicado na Tabela 1.

A presença ou ausência das imagens em plataformas específicas sugere especificidades do tipo de circulação que elas tiveram. Enquanto ambas têm presença registrada no Twitter, Facebook, YouTube e Pinterest, nas demais plataformas cada imagem tem presença exclusiva. Apenas a imagem A (postada pelo ex-jogador Cafu) aparece nas plataformas VKontakte, Odnoklassniki e Sina Weibo, que são populares em regiões específicas do hemisfério leste do globo — as duas primeiras na região da Rússia e a última, na China. Por sua vez, apenas a imagem B (que se refere à Seleção Chilena) aparece na plataforma Taringa, de relevância particular para o público latino-americano, com enfoque em esportes. Também apenas ela aparece nas plataformas Tumblr e Reddit, que têm relevância no contexto de circulação de memes e na qual atores institucionais não costumam ter presença explícita. A rede de plataformas articulada pela circulação das imagens sugere, então, dinâmicas vinculadas às características dos públicos que se apropriam de cada mídia social.

TABELA 1 — Plataformas de mídia social com ocorrências das imagens rastreadas.

| Plataforma | Domínio | Imagem com ocorrência | | Repres. no gráfico |
|---------------|--------------------|-----------------------|---|--------------------|
| Facebook | facebook [.com] | A | B | [Cyan] |
| YouTube | youtube [.com] | A | B | [Blue] |
| Twitter | twitter [.com] | A | B | [Red] |
| Reddit | reddit [.com] | | B | [Purple] |
| Pinterest | pinterest [vários] | A | B | [Green] |
| Tumblr | tumblr [.com] | | B | [Orange] |
| Google+ | plus.google [.com] | | B | [Blue] |
| Vkontakte | vk [.com] | A | | [Green] |
| Sina Weibo | weibo [.com/.cn] | A | | [Yellow] |
| Odnoklassniki | ok [.ru] | A | | [Brown] |
| Taringa! | taringa [.net] | | B | [Orange] |

Para compreender a distribuição entre plataformas ao longo do tempo, nos valem de visualizações Stream graph (Mauri, Elli, Caviglia, Uboldi, & Azzi, 2017), que permitem representar séries temporais categorizadas. A Figura 3 apresenta visualização de todas as URLs obtidas para cada uma das imagens.

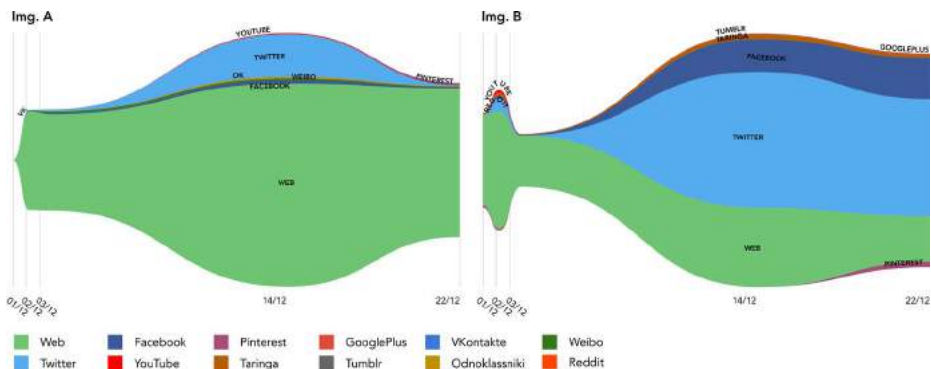


FIGURA 3 — Distribuição das URLs de páginas com imagens correspondentes, categorizadas segundo o domínio. Para cada gráfico, a escala vertical foi normalizada considerando o número máximo de URLs obtidas em cada caso. Os gráficos representam cinco momentos no tempo, conforme indicado no eixo horizontal. Curvas de transição foram extrapoladas com a finalidade única de legibilidade visual, sem indicar dados reais entre cada momento.

Um primeiro aspecto da comparação entre os gráficos é a diferença na silhueta geral, indicando uma dinâmica de crescimento distinto em cada caso. Enquanto para a imagem A o gráfico assume uma forma de cone, iniciando de registros nulos e então crescendo gradativamente, a imagem B começa com uma base já mais larga para então diminuir e voltar a crescer. Isto se deve ao fato de que a imagem B é um meme baseado em imagem já em circulação. Portanto, desde o primeiro momento, a API da Google recuperou versões indexadas da imagem na web. Contudo, a aparição de sua releitura na forma de um meme fotográfico parece ter reconfigurado a indexação, ocasionando a oscilação posterior na quantidade de ocorrências identificadas.

Quanto à diferença nas ocorrências das imagens claramente há uma diferença grande na proporção de URLs de plataformas de mídia social, em comparação a URLs da web, isto é, que não pertencem a nenhuma plataforma conhecida. A imagem A tem ocorrências registradas majoritariamente na web, enquanto para a imagem B a situação se inverte. Um fator que possivelmente contribui para esta diferença é o tipo de enunciação que cada uma compreende: enquanto a imagem A é um registro de bastidores, a imagem B é um meme que faz uma piada sobre a não classificação de um time. A imagem A pareceria ser, assim, mais aderente aos interesses de sites jornalísticos e blogs, enquanto a imagem B não teria tanto apelo a atores institucionais e, portanto, sua circulação majoritária em plataformas de mídia social seria compreensível.

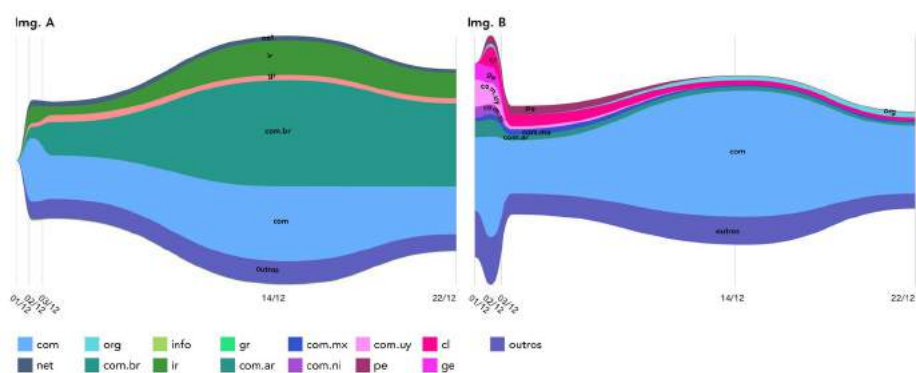


FIGURA 4 — Distribuição das URLs de páginas com imagens correspondentes, categorizadas segundo o TLD. Para cada gráfico, a escala vertical foi normalizada considerando o número máximo de URLs obtidas em cada caso. Os gráficos representam 5 momentos no tempo, conforme indicado no eixo horizontal. Curvas de transição foram extrapoladas com a finalidade única de legibilidade visual, sem indicar dados reais entre cada momento.

Para a análise de indicadores geográficos, baseado nos TLDs, voltamo-nos especificamente às URLs que não pertenciam às plataformas de mídia social mapeadas inicialmente – ou seja a domínios a que referimos como a *web aberta*. Isto porque no âmbito das plataformas, ainda que publicações sejam oriundas de localidades diferentes, elas são geralmente reunidas sob o TLD da plataforma, que na maioria dos casos é o “.com”, que seria sobrevalorizado na análise.

A Figura 4 exibe a distribuição observada entre TLDs para as URLs em cada imagem. O domínio não-regional “.com” permanece um dos mais proeminentes, mas os domínios regionais apontam para alguns aspectos relevantes para análise proposta. A suposição inicial de um apelo mais regional para a imagem B é confirmada por uma forte ocorrência em TLDs relativos a países latino-americanos, de países como México (‘.com.mx’), Argentina (‘.com.ar’), Chile (‘.cl’), Peru (‘.pe’), Uruguai (‘.com.uy’) e Brasil (‘.com.br’), entre outros. A imagem A, por sua vez, apresenta TLDs globalmente mais dispersos, muitos dos quais, inclusive, subsumidos pela categoria “outros” na visualização. Além da já esperada proeminência do domínio brasileiro (‘.com.br’), também encontramos domínios do Irã (‘.ir’), China (‘.cn’), Bulgária (‘.bg’), Rússia (‘.ru’), República Tcheca (‘.cz’), Vietnã (‘.vn’), Coreia do Sul (‘.kr’) e Reino Unido (‘.uk’), entre outros. Verifica-se, desse modo, nossa hipótese de que a imagem A possuiria um apelo global mais amplo do que a imagem B.

POTENCIALIDADES E LIMITAÇÕES

Em vista destes experimentos analíticos, compreendemos haver potencial para o estudo da circulação de imagens entre plataformas e na web a partir das estratégias de coleta e análise aqui delineadas. Como indicamos, seria possível ir ao nível das páginas individuais em que cada imagem foi rastreada na Internet e analisar as ressignificações em suas ocorrências individuais. No caminho que seguimos, de análise agregada dos dados, apresenta-se o potencial de identificar certas dinâmicas amplas de circulação para cada imagem, com aspectos que ganham saliência em uma perspectiva comparada. As ocorrências em plataformas sugerem aspectos relativos aos públicos engajados pelas imagens, bem como algumas indicações geográficas. As ocorrências na web, por sua vez, segmentadas pelos domínios de topo, apontam para ocorrências localizadas geograficamente. Um dos pontos fortes da metodologia, portanto, é a possibilidade de acessar instâncias regionais ou locais da web, que estariam, de outro modo, além das capacidades linguísticas ou do conhecimento cultural dos pesquisadores.

Avançando na perspectiva simétrica que propusemos ao início, é também preciso tecer observações acerca dos limites e inflexões epistemológicas introduzidas pela metodologia adotada. Um dos pontos principais, evidentemente, é o papel mediador desempenhado pela API Cloud Vision da Google, sobre o qual pouco se dá a conhecer devido à opacidade de seu processo de indexação. Para além desta instância, também será preciso observar os limites interpostos pelo modo como a web se organiza contemporaneamente, no entorno das plataformas, e também os limites que cada site coloca para a própria indexação da Google.

Um primeiro aspecto por se observar é o caráter relativamente limitado das listas de URLs que pudemos obter em nossos experimentos. Embora tenhamos definido, na etapa de processamento, um limite mais amplo para o número de resultados para cada imagem⁹, em nenhum momento obtivemos um número superior a 951 URLs para uma dada imagem. Consideramos este número baixo por ser menor do que as ocorrências conhecidas da imagem no Twitter, por exemplo. Também se nota, contraintuitivamente, que o número de URLs não cresce de forma cumulativa ao longo do período observado, mas chega a diminuir entre 14 e 22 de dezembro. Na imagem A, em particular, esta redução se deu com o quase desaparecimento de URLs do Twitter, embora saibamos que não houve este radical apagamento das postagens contendo a imagem. Destaca-se, também, a indexação relativamente tardia de URLs desta plataforma de mídia social, que apenas surgem nos dados na coleta do dia 14 de dezembro. Outro aspecto importante diz respeito à distribuição dos resultados entre as plataformas conhecidas, com um grande número de URLs pertinentes ao Twitter e poucas de outras plataformas. Em particular, destaca-se a total ausência de URLs do Instagram, embora seja altamente provável que versões idênticas ou quase idênticas destas imagens tenham aparecido nesta plataforma.

A mediação da API Cloud Vision, da Google, com seu processo de indexação da *web*, é certamente salientada por estes indícios, o que demanda que pesquisas como a aqui resumida observem criticamente o papel das APIs na produção dos dados e como condicionante da nossa capacidade de conhecer a circulação das imagens. Como antecipado, não podemos tomar os dados de forma positivista. Devemos, de outro modo, reconhecer que eles são produto da conjunção entre a efetiva circulação das imagens entre plataformas e na web e o modo de operação do mecanismo de indexação da Google. Em linha com a perspectiva dos métodos digitais, a possibilidade de fundamentação da análise com dados obtidos online precisa ser, então, considerada criticamente em uma abordagem reflexiva das

⁹ Atribuímos o valor 10.000 para o atributo do limite de número de resultados da API.

instâncias mediadoras da *web* e das plataformas, inclusive enquanto condicionantes de nossas capacidades epistêmicas.

Outro aspecto por salientar diante das limitações observadas nos dados diz respeito ao processo de plataformização da *web* (Helmond, 2015), levando a um progressivo afastamento do modelo universalista idealizado na fundação da WWW no início dos anos 1990. Este processo incide nos dados obtidos quando observamos a ausência do Instagram e a pequena quantidade de resultados obtidos em plataformas além do Twitter. Embora, inicialmente, pudéssemos supor que a maior incidência no Twitter seria devido a termos colhido as imagens nesta plataforma, não parece ser este o motivo¹⁰. A diferença parece ser resultado das políticas adotadas por cada plataforma com relação à indexação do conteúdo por agentes externos, como o mecanismo de busca da Google. Embora parte destas medidas sejam justificáveis em vista da privacidade dos usuários, destaca-se a tendência a certo sectarismo das plataformas em relação aos padrões da *web* aberta, estabelecendo seus próprios modos para organizar e recuperar informações. Esta situação tende a dificultar o intercâmbio de conteúdos entre plataformas e limitar o alcance de pesquisas que visam compreender a circulação de imagens entre elas.

Para que seja passível de indexação, um conteúdo precisa estar disponível através do protocolo HTTP, em uma URL de acesso público (sem a necessidade de *login*), e, também, a URL não deve ser bloqueada pelo site no arquivo padrão de exclusão de robôs (*robots.txt*). Tuítes são tipicamente mais acessíveis a buscadores, entre outros motivos, pela maior frequência do caráter público das contas de usuário, tanto por uma definição padrão da plataforma quanto por sua cultura de uso. A plataforma também é comparativamente mais permissiva à indexação externa do conteúdo¹¹. O Facebook e o Instagram, por sua vez, são bem mais restritivos. No Facebook, quando da realização de nosso estudo apenas a raiz das páginas públicas eram acessíveis por URLs públicas e por crawlers de indexação. Posts individuais das páginas não são acessíveis e as URLs das imagens em si, tampouco¹². No caso do Instagram, embora os perfis de usuário sejam indexáveis, o código-fonte das páginas mascara as URLs das imagens e dificultam o seu

10 Na publicação original deste estudo (d'Andrea & Mintz, 2019), uma das imagens consideradas era um meme que embora colhido no Twitter havia sido inicialmente publicado no Facebook, até onde pudemos rastreá-lo.

11 O arquivo *robots.txt* do Twitter permite o acesso a tuítes e contas de usuário públicas (<https://web.archive.org/web/20180815000157/http://twitter.com/robots.txt>).

12 Veja o arquivo *robots.txt* do Facebook (<https://web.archive.org/web/20180910120036/https://facebook.com/robots.txt>), que limita o acesso a alguns caminhos relacionados a imagens ("*photo.php*," "*photos.php*," and "*album.php*").

acesso por indexadores. A plataforma é também bastante restrita a publicações entre plataformas, como pelo impedimento de reprodução das imagens quando o usuário posta conteúdo da plataforma simultaneamente no Twitter.

Este conjunto de fatores coloca em destaque a necessidade de compreender o caráter circunstancial dos esforços de investigação e dos métodos mobilizados. As limitações devem ser consideradas reflexivamente enquanto condicionantes da investigação e, também, enquanto indícios para uma compreensão das dinâmicas que são simultaneamente constituintes e constitutivas da realidade observada. Propomos, então, a noção relativística de *horizonte de recuperabilidade* para nomear os limites difusos que circunscrevem as capacidades dos métodos de investigação empregados. Há um limite contingente, constituído pelo tema, pelas ferramentas e pela circunstância da investigação, e este limite precisa ser abordado reflexivamente como forma de habilitar a navegação em meio às incertezas do objeto e dos dados utilizados para descrevê-lo. Mesmo que o conteúdo esteja efetivamente online, ele pode não ser recuperado por determinada ferramenta ou conjunto de métodos empregados. Entretanto, deve-se ressaltar, trata-se de limite interposto também à experiência de uso da web quando ela é mediada por tais operadores.

No caso da circulação entre plataformas, a estratégia analítica que utiliza mecanismos de busca de imagem baseada em conteúdo sugere um horizonte de recuperabilidade complexamente definido, envolvendo os processos de reconhecimento de imagem, as dinâmicas de indexação e os limites transitórios estabelecidos por cada plataforma aos mecanismos externos de indexação. Como apontamos nesta seção, a dinâmica de plataformização da web sugere ainda uma tendência ao estreitamento desse horizonte. São limites que condicionam os dados que podem ser obtidos, mas não impedem, em absoluto, o desenvolvimento de estudos baseados nestes métodos — desde que as limitações sejam reconhecidas e incorporadas como elementos da análise.

CONCLUSÕES

Neste capítulo, discutimos métodos potenciais para o estudo da circulação de imagens entre plataformas baseados em uma API de visão computacional para a busca de imagens baseada em conteúdo na web. Por meio da abordagem proposta, as imagens são tomadas como marcadores de tráfego (Elmer & Langlois, 2013) que articulam diferentes publicações online e apontam para dinâmicas de circulação transnacional de conteúdos entre diferentes plataformas e a “web

aberta”. Para cada imagem, a ferramenta permite a obtenção de listas de URLs de páginas em que foram encontradas imagens idênticas ou quase idênticas. Estas páginas podem ser analisadas individualmente na pesquisa ou, como sugerimos, analisadas em forma agregada através da decomposição dos endereços em partes protocolares significativas, tais como o domínio e o domínio de topo. Apresentamos possibilidades de interpretação dos dados gerados por tal análise, com a inferência de aspectos relativos aos públicos e às localidades geográficas em que cada imagem repercutiu.

Baseando-nos no princípio de simetria dos STS, em sua releitura por Marres e Moats (2015), bem como na reflexividade dos métodos digitais (Rogers, 2015), voltamo-nos em nossa discussão tanto às oportunidades analíticas quanto às limitações (Rieder et al., 2015) dos métodos empregados. Entre as oportunidades identificadas, ressaltamos o potencial de descrição de dinâmicas de circulação baseadas em imagens, com a vantagem de parcialmente superar limitações linguísticas e culturais das pessoas pesquisadoras. Entre as limitações, indicamos aspectos relativos à opacidade das mediações analíticas, que trazem diversas incertezas para as descrições. Indicamos, também, como os dados obtidos apontam para o impacto do processo de plataformização da web (Helmond, 2015) no contexto da pesquisa em métodos digitais. Ressaltamos a necessidade de incorporar as incertezas decorrentes destes processos à análise, inclusive ressaltando como os indícios das limitações oferecem-se enquanto dados para compreender aspectos constitutivos da realidade observada. A noção relativística de *horizonte de recuperabilidade* foi proposta como forma de nomear a circunstancialidade dos dados obtidos no estudo enquanto um dos produtos dos métodos empregados.

REFERÊNCIAS

- Borra, E., & Rieder, B. (2014). Programmed method: Developing a toolset for capturing and analyzing tweets. *Aslib Journal of Information Management*. <https://doi.org/10/f6grpr>
- Bucher, T. (2013). Objects of Intense Feeling: The Case of the Twitter API. *Computational Culture*, (3). Recuperado de <http://computationalculture.net/objects-of-intense-feeling-the-case-of-the-twitter-api/>
- Burgess, J., & Matamoros-Fernández, A. (2016). Mapping sociocultural controversies across digital media platforms: One week of #gamergate on Twitter, YouTube, and Tumblr. *Communication Research and Practice*, 2(1), 79-96. <https://doi.org/10.1080/22041451.2016.1155338>
- Cardon, D., Cointet, J.-P., & Mazières, A. (2018). Neurons spike back. The invention of inductive machines and the artificial intelligence controversy. *Réseaux*, n° 211(5), 173. <https://doi.org/10/gfzpgs>

- d'Andrea, C., & Mintz, A. (2019). Studying the Live Cross-Platform Circulation of Images With Computer Vision API: An Experiment Based on a Sports Media Event. *International Journal of Communication*, 13(0), 21.
- Driscoll, K., & Thorson, K. (2015). Searching and Clustering Methodologies: Connecting Political Communication Content across Platforms. *The ANNALS of the American Academy of Political and Social Science*, 659(1), 134-148. <https://doi.org/10.1177/0002716215570570>
- Elmer, G., & Langlois, G. (2013). Networked campaigns: Traffic tags and cross platform analysis on the web. *Information Polity*, 18(1), 43-56. <https://doi.org/10.3233/IP-2011-0244>
- Faulkner, S., Vis, F., & D'Orazio, F. (2018). Analysing Social Media Images. In J. Burgess, A. Marwick, & T. Poell (Orgs.), *The SAGE Handbook of Social Media* (p. 160-178). <https://doi.org/10.4135/9781473984066.n10>
- Gibbs, M., Meese, J., Arnold, M., Nansen, B., & Carter, M. (2015). #Funeral and Instagram: Death, social media, and platform vernacular. *Information, Communication & Society*, 18(3), 255–268. <https://doi.org/10.1080/1369118X.2014.987152>
- Helmond, A. (2015). The Platformization of the Web: Making Web Data Platform Ready. *Social Media + Society*, 1(2). <https://doi.org/10.1177/2056305115603080>
- Highfield, T., & Leaver, T. (2016). Instagrammatics and digital methods: Studying visual social media, from selfies and GIFs to memes and emoji. *Communication Research and Practice*, 2(1), 47-62. <https://doi.org/10.1080/22041451.2016.1155332>
- Kurkowski, J. (2019). *Tldextract*. Recuperado de <https://pypi.org/project/tldextract/>
- Levin, G. (2006). Computer vision for artists and designers: Pedagogic tools and techniques for novice programmers. *AI & SOCIETY*, 20(4), 462-482. <https://doi.org/10.1007/s00146-006-0049-2>
- Marres, N., & Moats, D. (2015). Mapping Controversies with Social Media: The Case for Symmetry. *Social Media + Society*, 1(2). <https://doi.org/10.1177/2056305115604176>
- Mauri, M., Elli, T., Caviglia, G., Uboldi, G., & Azzi, M. (2017). RAWGraphs: A Visualisation Platform to Create Open Outputs. *Proceedings of the 12th Biannual Conference on Italian SIGCHI Chapter – CHIItaly '17*, 1-5. <https://doi.org/10.1145/3125571.3125585>
- Mintz, A. (2018). *Memespector Python*. Recuperado de <https://github.com/amintz/memespector-python>
- Niederer, S. (2018). *Networked images: Visual methodologies for the digital age*. Amsterdam: Amsterdam University of Applied Sciences.
- Pearce, W., Özkula, S. M., Greene, A. K., Teeling, L., Bansard, J. S., Omena, J. J., & Rabello, E. T. (2018). Visual cross-platform analysis: Digital methods to research social media images. *Information, Communication & Society*, 1-20. <https://doi.org/10.1080/1369118X.2018.1486871>
- Plantin, J.-C., Lagoze, C., Edwards, P. N., & Sandvig, C. (2016). Infrastructure studies meet platform studies in the age of Google and Facebook. *New Media & Society*. <https://doi.org/10.1177/1461444816661553>
- Rieder, B., Abdulla, R., Poell, T., Woltering, R., & Zack, L. (2015). Data critique and analytical opportunities for very large Facebook Pages: Lessons learned from exploring “We are all Khaled Said”. *Big Data & Society*, 2(2), 205395171561498. <https://doi.org/10.1177/2053951715614980>
- Rieder, B., Den Tex, E., & Mintz, A. (2018). *Memespector*. Recuperado de <https://github.com/bernorieder/memespector>
- Rieder, B., & Sire, G. (2013). Conflicts of interest and incentives to bias: A microeconomic critique of Google's tangled position on the Web. *New Media & Society*, 16(2), 195–211. <https://doi.org/10.1177/1461444813481195>
- Rogers, R. (2015). Digital methods for web research. *Emerging Trends in the Social and Behavioral Sciences: An Interdisciplinary, Searchable, and Linkable Resource*.

- Rogers, R. (2017). Foundations of digital methods: Query design. In M. T. Schäfer & K. Van Es (Orgs.), *The datafied society: Studying culture through data* (p. 75-94). Amsterdam: Amsterdam University Press.
- Rogers, R. (2018). *Digital methods for cross-platform analysis* (Por pages 91-108). <https://doi.org/10.4135/9781473984066>
- Shifman, L. (2014). The Cultural Logic of Photo-Based Meme Genres. *Journal of Visual Culture*, 13(3), 340-358. <https://doi.org/10.1177/1470412914546577>
- Silva, T., Barciela, P., & Meirelles, P. (2018). Mapeando Imagens de Desinformação e Fake News Político-Eleitorais com Inteligência Artificial. 3º CONEC: Congresso Nacional de Estudos Comunicacionais Da PUC Minas Poços de Caldas – Convergência e Monitoramento, 413-427. Recuperado de <https://conec.pucpcaldas.br/wp-content/uploads/2019/06/anais2018.pdf>
- Smeulders, A. W. M., Worring, M., Santini, S., Gupta, A., & Jain, R. (2000). Content-based image retrieval at the end of the early years. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(12), 1349–1380. <https://doi.org/10.1109/34.895972>
- Venturini, T., & Rogers, R. (2019). “API-based research” or how can digital sociology and journalism studies learn from the Cambridge Analytica affair.
- Vis, E., & Goriunova, O. (Orgs.). (2015). *The Iconic Image on Social Media: A rapid research response to the death of Aylan Kurdi*. Recuperado de <http://visualsocialmedialab.org/projects/the-iconic-image-on-social-media>
- Wan, J., Wang, D., Hoi, S. C. H., Wu, P., Zhu, J., Zhang, Y., & Li, J. (2014). Deep Learning for Content-Based Image Retrieval: A Comprehensive Study. *Proceedings of the ACM International Conference on Multimedia – MM '14*, 157-166. <https://doi.org/10.1145/2647868.2654948>

AUTORAS & AUTORES

Ana Marta M. Flores

Doutora e mestre em Jornalismo pela Universidade Federal de Santa Catarina (UFSC), no Brasil, com período de estudos na Universidade de Lisboa, em Portugal. Pesquisa jornalismo especializado em moda no contexto convergente e atualmente investiga os Estudos de Tendências aplicados à inovação no jornalismo. É pesquisadora do Núcleo de Estudos e Produção em Hipermídia Aplicada ao Jornalismo (Nephi-Jor), eixo do Grupo de Pesquisa em Hipermídia e Linguagem/CNPq, Brasil, e membro do grupo de investigação *Social Media Research Techniques* (SMART), do iNOVA Media Lab, em Portugal.

André Mintz

Doutorando em Comunicação Social pela Universidade Federal de Minas Gerais (UFMG), Brasil, com bolsa concedida pela CAPES. Membro do R-EST – estudos em redes sociotécnicas e no Núcleo de Pesquisa em Conexões Intermidiáticas (NucCon). Mestre em Comunicação Social pela UFMG e em Media Arts Cultures pela Aalborg University, Dinamarca, Lodz University, Polônia e Danube University Krems, Áustria. Temas de interesse incluem: métodos digitais, plataformas de mídia social, estudos sociais de ciência e tecnologia, estudos da imagem e arte contemporânea.

Anne Helmond

Anne Helmond é professora assistente de Novos Media e Cultura Digital da Universidade de Amsterdão, Holanda. Ingressou na *Digital Methods Initiative* como *designer-analista* desde a fundação, em 2007. Na sua investigação, concentra-se em plataformas e aplicações de *social media* a partir da perspectiva infraestrutural. Os seus interesses de investigação incluem métodos digitais, estudos de *software*, plataforma, aplicações e infraestrutura, além de histórico da *web*. Em sua

dissertação “The web as platform: Data flows in social media”, ela desenvolveu a noção de “plataformização”, para entender a dupla lógica da extensão das plataformas de mídia social na web e, simultaneamente, seu esforço para tornar dados da web externos “prontos para plataforma”. Atualmente, possui uma bolsa Veni da Organização Holandesa de Investigação Científica (NWO) para o projeto “Ecosistemas de Apps: uma história crítica de aplicações” (2017-2020).

Beatrice Gobbo

Desde novembro de 2018, é doutoranda em Design no Politécnico de Milão, em Itália. É membro do *DensityDesign Lab*, um grupo de investigação focado em visualização e *information design*. Atualmente investiga o papel do *design* de comunicação e de *information visualisation* na área de *computer science*.

Bernhard Rieder

Professor Associado em Novos Media na Universidade de Amsterdão, Holanda. Além de desenvolver investigação com base nos métodos digitais, seus interesses de pesquisa centram-se em história, teoria e política de software, em particular o papel dos algoritmos nos processos sociais e na produção do conhecimento. É autor do livro a ser lançado “Engines of Order. A Mechanology of Algorithmic Techniques”.

Bruno Latour

Bruno Latour é professor emérito associado ao médialab e ao programa de artes políticas (SPEAP) da Sciences Po em Paris, França. Desde janeiro de 2018 é bolseiro no *Zentrum fur Media Kunst* (ZKM) e professor no HfG, ambos em Karlsruhe, Alemanha. Além de curador de Zonas Críticas na ZKM (abertura em maio de 2020), também é curador da Bienal de Arte de Taipé, Taiwan, juntamente com Martin Guinard (abertura em outubro de 2020). Membro de várias academias, recebeu em 2013 o Prêmio Holberg. Escreveu e editou mais de 20 livros e publicou mais de 150 artigos.

Carlos d'Andréa

Professor permanente do Programa de Pós-Graduação em Comunicação (PPG-COM) da Universidade Federal de Minas Gerais (UFMG), no Brasil. É coordenador do R-EST – estudos redes sociotécnicas e membro do Núcleo de Pesquisa em Conexões Intermediáticas (NucCon). Em 2017/2018, foi investigador visitante do Departamento de Media Studies da Universidade de Amsterdão, Holanda (bolsa CAPES de pós-doutoramento no exterior). Entre os seus temas de interesse

destacam-se: Plataformas de Mídias Sociais; Métodos digitais de pesquisa; Coleta, processamento e visualização de dados; Ciência, Esporte e controvérsias; STS / Teoria Ator-Rede; jornalismo contemporâneo.

Elaine Teixeira Rabello

Professora adjunta do Instituto de Medicina Social da Universidade do Estado do Rio de Janeiro, Brasil, integra também o Grupo de Estudos Sociais em Tecnociência e Saúde – BIOMEDSCI. Atualmente é investigadora pós-doutoral do Public Administration and Policy Group da Wageningen University & Research – Holanda, trabalhando no tema de estudos visuais de controvérsias online sobre saúde e alimentação. Também é investigadora convidada da Fundação Oswaldo Cruz e coordenadora executiva da Rede Zika e Ciências Sociais, Brasil. Psicóloga, mestre e doutora em Saúde Coletiva, desenvolve pesquisa com métodos digitais e STS, focando a dinâmica da circulação do conhecimento em saúde, seus usos, benefícios e consequências para diferentes atores sociais.

Inês Amaral

Doutorada em Ciências da Comunicação pela Universidade do Minho, desde 2012, no âmbito do Programa de Doutoramento da Fundação para a Ciência e Tecnologia. É professora associada da Faculdade de Letras da Universidade de Coimbra, Portugal, e investigadora do Centro de Estudos de Comunicação e Sociedade (Universidade do Minho). Leciona na área de comunicação digital e dedica-se à análise de redes sociais desde 2008. Tem investigado sociabilização em redes sociais online, consumo dos media na era digital e movimentos sociais.

Fábio Castro Gouveia

Tecnologista em saúde pública da Fundação Oswaldo Cruz – Brasil, líder do Grupo de Pesquisa Ciência, Dados, Redes e Metrias – (Scimetrics), participa como investigador na *Rede Zika Ciências Sociais*. Biólogo, mestre em Microbiologia e Imunologia e doutor em Química Biológica (Educação, Gestão e Difusão de Biociências), fez um pós-doutoramento curto como *Visiting Fellow* da *Katolieke Universiteit Leuven*, Bélgica, foi selecionado no edital 2009 do *Coimbra Group Scholarships Programme for Young Professors and Researchers from Latin American Universities*. Desenvolve pesquisas na área Ciência da Informação com ênfase em Estudos Métricos da Informação (Cientometria, Webometria, Altmetria e Indicadores de Ciência, Tecnologia e Inovação), Métodos Digitais, STS, Data Science e Tecnologia Blockchain, e na área de Divulgação Científica e Comunicação em Saúde, com ênfase em estudos sobre internet e mídias sociais.

Gabriele Colombo

Doutor em Design pelo Instituto Politécnico de Milão, Itália, Gabriele Colombo é integrado ao DensityDesign, o laboratório de investigação do instituto e também ao Departamento de Arquitetura e Artes da Universidade IUAV de Veneza, Itália. É professor do curso de mestrado em Design de Comunicação – *Final Synthesis Studies*, no Instituto Politécnico de Milão, onde leciona “Métodos Digitais e Design de Comunicação”. Sua investigação e atividades de ensino têm como foco o design visual de ferramentas que apoiam a investigação social digital, bem como o design de novas estratégias para comunicação, exploração, análise e valorização de coleções de imagens e vídeos.

Janna Joceli Omena

Investigadora doutoral em Mídias Digitais da Universidade Nova de Lisboa I NOVA FCSH (Faculdade de Ciências Sociais e Humanas), membro do iNOVA Media Lab onde lidera o #SMARTdatasprint I Digital Media Winter Institute, ambos em Lisboa, Portugal. No iNOVA Media Lab, é coordenadora do grupo de investigação Social Media Research Techniques (SMART). Além da teoria de métodos digitais, tem particular interesse nos estudos de multiplataforma, estudos de software e análise visual de redes. Atualmente, dedica-se à área da técnica das plataformas de media sociais, e às possibilidades ou limitações para a investigação digital. Tweeta em @jannajoceli, mais sobre a sua investigação pode ser encontrado em thesocialplatforms.wordpress.com.

Jonathan Gray

Jonathan Gray é professor de Estudos Críticos de Infraestrutura no Departamento de Humanidades Digitais do King’s College London, Reino Unido. Atualmente escreve um livro sobre “mundos de dados”, concentrando-se a sua investigação na política de dados abertos e informações públicas. Além disso, interessa-se pela vida social dos métodos digitais no contexto de advocacia, jornalismo e políticas públicas. Também é investigador associado da Digital Methods Initiative, da Universidade de Amsterdão, Holanda; investigador associado do médialab Sciences Po, em Paris, França, e bolsheiro do Tow Center for Digital Journalism, da Universidade de Columbia, EUA. Mais informações podem ser encontradas em jonathangray.org e no Twitter em @jwyg.

Liliana Bounegru

Liliana Bounegru é investigadora de novos media, métodos digitais e jornalismo digital, estando atualmente a fazer o pós-doutoramento na Universidade

de Oxford, Reino Unido. Bounegru é cofundadora de Public Data Lab e da Digital Methods Initiative, em Amsterdão, Holanda, e do médialab, Sciences Po, em Paris, França. Através do seu trabalho como responsável do programa de jornalismo de dados do Centro Europeu de Jornalismo, coeditou o Data Journalism Handbook e o DataDrivenJournalism.net. Além disso, executou a primeira edição do Data Journalism Awards, do Doing Journalism with Data MOOC, da School of Data Journalism inúmeras ações de formação e conferências. Tweeta @bb_liliana, e mais sobre ela pode ser encontrado em lilianabounegru.org.

Mathieu Jacomy

Mathieu Jacomy é doutorando de tecnoantropologia no TANTLab, da Universidade de Aalborg, Dinamarca. Foi engenheiro de investigação do médialab Sciences Po, em Paris, França, ao longo de dez anos, e é cofundador do Gephi. Desenvolve instrumentos digitais envolvendo visualização de dados e análise de redes para as ciências sociais e humanas. A sua investigação atual concentra-se na análise de redes visuais, métodos digitais e mapeamento de questões problemáticas (*issue mapping*). Contribui para o desenvolvimento do web crawler Hyphe, para o projeto MiniVan e, certamente, Gephi. Tweeta em @jacomy e escreve no blog em reticular.hypotheses.org.

Michele Mauri

Desde de outubro de 2014, é investigador do Instituto Politécnico de Milão, Itália. O seu trabalho de investigação é realizado em colaboração com o *DensityDesign Lab*, um grupo de investigação focado na visualização da informatização e no *design* de informações. No laboratório, coordena a investigação, o *design* e o desenvolvimento de projetos relacionados com a comunicação visual de dados e informações, em particular projetos de dados de origens digitais e métodos digitais. Em 2015, obteve o PhD em *Design* com a tese “*Design del non finito*”. Interessa-se por visualização de dados, estatística, fontes de dados de origens digitais e, é claro, *design* gráfico.

Theo Röhle

Professor de Estudos dos Media e Comunicação da Universidade de Karlstad, Suécia. Os seus campos de investigação incluem escudos algorítmicos, uma pesquisa sobre as respostas das organizações dos media aos ataques de navegação de negação de serviço (distributed denial-of-service – DDoS); culturas sobre a história do software.

Tommaso Venturini

Tommaso Venturini é investigador do CNRS Centre for Internet and Society. Também é investigador associado do INRIA e do médialab Sciences Po em Paris, ambos em França, além de membro fundador do Public Data Lab. Em 2017 e 2018, foi investigador da École Normale Supérieure de Lyon, França, e recebeu a bolsa de pesquisa avançada do Instituto Francês de Pesquisa em Ciência da Computação e Automação. Em 2016, foi professor de Métodos Digitais do Departamento de Humanidades Digitais do King's College London, Reino Unido. De 2009 a 2015, coordenou as atividades de investigação do médialab Sciences Po Paris.

Richard Rogers

Richard Rogers é professor catedrático de Novos Media e Cultura Digital da Universidade de Amsterdão, Holanda. É diretor do Govcom.org, grupo responsável pelo Issue Crawler e outras ferramentas info-políticas, e pela Digital Methods Initiative. Entre outros trabalhos, Rogers é autor de *Politics on the Web* (MIT Press, 2004), reconhecido como melhor livro do ano de 2005 pela American Society of Information Science & Technology (ASIS&T) e *Digital Methods* (MIT Press, 2013), premiado como livro de destaque de 2014 pela International Communication Association. O livro *Issue Mapping for an Ageing Europe*, escrito com Natalia Sánchez e Aleksandra Kil, foi publicado pela Amsterdam University Press (2015), e *Doing Digital Methods* (2019), pela Sage.

