Extensive parallel processing on scale free networks

Peter Sollich,1 Daniele Tantari,2 Alessia Annibale,3,1 and Adriano Barra4

1Mathematics Department, King’s College London, The Strand WC²R 2LS, London, UK.
2Dipartimento di Matematica, Sapienza Università di Roma, P.le Aldo Moro 2, 00185, Roma, Italy.
3Institute for Mathematical and Molecular Biomedicine, King’s College London, Hoddinck Building, London SE1 1UL, U.K.
4Dipartimento di Fisica, Sapienza Università di Roma, P.le A. Moro 2, 00185, Roma, Italy.
(Dated: November 10, 2014)

We adapt belief-propagation techniques to study the equilibrium behavior of a bipartite spin-glass, with interactions between two sets of $N$ and $P = \alpha N$ spins each having arbitrary degree, i.e., number of interaction partners in the opposite set. An equivalent view is then of a system of $N$ neurons storing $P$ diluted patterns via Hebbian learning, in the high storage regime. Our methods allows analysis of parallel pattern processing on a broad class of graphs, including those with pattern asymmetry and heterogeneous dilution, while previous replica approaches assumed homogeneity. We show that in a large part of the parameter space of noise, dilution and storage load, delimited by a critical surface, the network behaves as an extensive parallel processor, retrieving all $P$ patterns in parallel without falling into spurious states due to pattern cross-talk and typical of the structural glassiness built into the network. Parallel extensive retrieval is more robust for homogeneous degree distributions, and is not disrupted by asymmetric pattern distributions. For scale free pattern degree distributions, Hebbian learning induces modularity in the neuron network so that our work gives the first theoretical description for extensive information processing on modular and scale free networks.

FIG. 1: (Left) Example of a Hebbian scale free network of $N = 500$ neurons and $P = \alpha N$ patterns with $\alpha = 3$. Pattern degrees are distributed as $P(\epsilon) \propto e^{-\gamma}$ for $\epsilon \geq 1$, with $\gamma = 2.2$. (Right) The network structure is modular as shown by the generalized topological overlap matrix [9, 10], with (top) modules arranged in a hierarchical fashion. See SM for further information on network generation.
level, because is the basis of extensive parallel retrieval of multiple patterns accomplished by the network. This is in agreement with experimental findings on intra-cellular protein networks, which have scale-free degrees but where interactions among hubs are strongly suppressed in favor of a modular structure, which minimizes cross talk among different modules [12, 13].

To confirm this scenario, we consider an equilibrated system of $N$ binary neurons $\sigma_i = \pm 1$ at temperature (fast noise) $T = 1/\beta$, with Hamiltonian

$$H(\sigma|\xi) = -\frac{1}{2} \sum_{i,j} \frac{P}{\mu} \xi_i^\mu \xi_j^\mu \sigma_i \sigma_j, $$

where pattern entries $\{\xi_i^\mu\}$ are sparse (i.e. the number of non-zero entries of a pattern is finite). We can then use a factor graph representation of the Boltzmann weight as $\prod_{\mu} F_{\mu}$, with factors

$$F_{\mu} = e^{\langle 2/\beta \rangle \sum_{i,j \in O(\mu)} \xi_i^\mu \xi_j^\mu \sigma_i \sigma_j} = \langle e^{\xi_i^\mu \xi_j^\mu \sigma_i \sigma_j} \rangle_z, $$

where $O(\mu) = \{ i : \xi_i^\mu \neq 0 \}$ and $z$ is a zero mean Gaussian variable with variance $\beta$ [24]. We denote by $e_\mu = |O(\mu)|$ the degree of a pattern $\mu$ and by $d_i = |N(i)|$ the degree of a neuron $i$, with $N(i) = \{ \mu : \xi_i^\mu \neq 0 \}$. We consider random graph ensembles with given degree distributions $P(d)$ and $P(e)$, and nonzero $\xi$’s independently and identically distributed (i.i.d.). Conservation of links demands $N(d) = P(e)$ where averages are taken over $P(d)$ and $P(e)$. The message from factor $\mu$ to node $j$ is the cavity distribution $P_{\mu}(\sigma_j)$ of $\sigma_j$ when this is coupled to factor $\mu$ only, which we can parameterize by an effective field $\psi_{\mu \to j}$. The message from node $j$ to factor $\mu$ is the cavity distribution $P_{\mu}(\sigma_j)$ of $\sigma_j$ when coupled to all factors except $\mu$, which we can parameterize by the field $\phi_{j \to \mu}$. The cavity equations are then [16]

$$P_{\mu}(\sigma_j) = \text{Tr}_{\{\sigma_k\}} F_{\mu}(\sigma_j, \{\sigma_k\}) \prod_{k \in O(\mu) \setminus j} P_{\mu}(\sigma_k),$$

$$P_{\nu}(\sigma_j) = \prod_{\mu \in N(j) \setminus \nu} P_{\mu}(\sigma_j).$$

Given the site factorization, conditional on $z$, of the factors (1), translating these equations into ones for the effective fields is straightforward:

$$\psi_{\mu \to j} = \text{tanh}^{-1} \langle \sigma_j \rangle_{\xi_i^\mu} =$$

$$\frac{\text{sinh}(z \xi_j^\mu \prod_{k \in O(\mu) \setminus j} \cos(\phi_{k \to \mu} + z \xi_k^\mu))}{\cos(\phi_{k \to \mu} + z \xi_k^\mu)} z, $$

$$\phi_{j \to \nu} = \sum_{\mu \in N(j) \setminus \nu} \psi_{\mu \to j}. $$

These equations, once iterated to convergence, are exact on tree graphs. They will also become exact on graphs sampled from our ensemble in the thermodynamic limit, because the sparsity of the $\xi_i^\mu$ makes the graphs locally tree-like, with typical loop lengths that diverge (logarithmically, see SM) with $N$ [14, 15].

For large $N$, we can describe the solution of the cavity equations in terms of the distribution of messages or fields, $W_\psi(\psi)$ and $W_\phi(\phi)$. Denoting by $\Psi(\{\phi_{k \to \mu}\}, \{\xi_k^\mu\})$ the r.h.s. of (4), convergence of the cavity iterations then implies the self-consistency equation

$$W_\psi(\psi) = \sum_e \frac{e^P(e)}{\langle e \rangle} \langle \delta(\psi - \Psi(\{\phi_{1 \to 1}, \phi_{e-1}, \xi^1, ..., \xi^e\}) \rangle, $$

where the average is over i.i.d. values of the (nonzero) $\xi^1, ..., \xi^e$ and over i.i.d. $\phi_{1 \to 1}, \phi_{e-1}$ drawn from $W_\phi(\psi)$, and similarly

$$W_\phi(\phi) = \sum_d \frac{dP(d)}{\langle d \rangle} \langle \delta(\phi - d^{-1} \sum_{\mu = 1} \psi_{\mu}) \rangle, $$

where the average is over i.i.d. $\psi_1, ..., \psi_{d-1}$ drawn from $W_\psi(\psi)$. Field distributions can then be obtained numerically by population dynamics (PD) [16]. For symmetric $\xi$-distributions, a delta function at the origin for both $W_\psi$, $W_\phi$ is always a solution, and we find this to be stable at low $\beta$. At high $\beta$, on the other hand, the $\psi$ can become large (see Fig. 2), hence also the $\phi$, and the spins $\sigma_i$ will typically be strongly polarized. The fields $\beta \xi_i^\mu \sum_{j \in O(\mu)} \xi_j^\mu \sigma_j$ then fluctuate little, and the $\psi$ as suitable averages of these fields cluster near multiples of $\beta$ (for $\xi = \pm 1$).

Our main interest is in the retrieval properties, encoded in the fluctuating pattern overlaps $m_{\mu} = \sum_{i \in O(\mu)} \xi_i^\mu \sigma_i$. Since the joint distribution of the $\sigma_i$ in $O(\mu)$ is $F_{\mu}(\{\sigma_i\}) \prod_{i \in O(\mu)} P_{\mu}(\sigma_i)$, the distribution of the pattern overlap $m_{\mu}$ is

$$\text{Tr}_{\{\sigma_i\}} \left\{ \delta(m_{\mu} - m) \exp(\sum_{i \in O(\mu)} \xi_i^\mu \sigma_i) \right\}_z, $$

$$\text{Tr}_{\{\sigma_i\}} \left\{ \exp(\sum_{i \in O(\mu)} \xi_i^\mu \sigma_i) \right\}_z. $$

Defining this as $P(m, \{\phi_{i \to \mu}\}, \{\xi_i^\mu\})$, in the graph ensem-
with coefficients $\Xi(\xi^\mu, \xi^\nu, \{\xi^\mu\})$ given by

$$\langle \sinh(z\xi^\mu)\sinh(z\xi^\nu)\prod_{\ell\in\{\mu,\nu\}\setminus\{k\}} \cosh(z\xi^\ell) \rangle_z = \frac{\prod_{\ell\in\{\mu,\nu\}\setminus\{k\}} \cosh(z\xi^\ell)_{\pm}}{\prod_{\ell\in\{\mu,\nu\}\setminus\{k\}} \cosh(z\xi^\ell)_{\mp}}.$$

The self-consistency relations for the field distributions $W_\phi$ and $W_\phi$ then show that as long as the mean fields are small, they are related to leading order by

$$\langle \phi \rangle = \langle \phi \rangle \sum_e P(e) \langle e \rangle \langle \Xi(\xi_1, \ldots, \xi_e) \rangle_{\langle e \rangle}, \quad \langle \phi \rangle = B_d\langle \psi \rangle,$$

where $B_d = \sum_e P(d)d(d-1)/\langle d \rangle$ is one of the two branching ratios of our locally tree-like graphs, the other being $B_c = \sum_e P(e)e(e-1)/\langle e \rangle$. If the means are zero then the onset of nonzero fields is detected by the variances, which are related to leading order by

$$\langle \phi^2 \rangle = \langle \phi^2 \rangle \sum_e P(e) \langle e \rangle \langle \Xi^2(\xi_1, \ldots, \xi_e) \rangle_{\langle e \rangle}, \quad \langle \phi^2 \rangle = B_d\langle \psi^2 \rangle.$$

Symmetric pattern distributions – When the $\xi$ are symmetrically distributed, then also the field distributions are always symmetric and there can be no instability from growing means; cf. (8). The bifurcation has to result from the growth of the variances, which from (11) occurs at $A = 1$ with

$$A = B_d \sum_e P(e) \langle e \rangle \langle \Xi^2(\xi_1, \ldots, \xi_e) \rangle_{\langle e \rangle}.$$  

This factorizes as $A = B_d A_c(\beta)$ with the dependence on the noise and on the topology (i.e. the distribution of the $e$’s) contained in the second factor $A_c(\beta)$. For $\beta \to 0$ the variance of $z$ goes to zero and $A_c(0) = 0$. For $\beta \to \infty$, the $z$-averages are dominated by large values of $z$ where $\sinh^2(z) \approx \cosh^2(z)$, so $A_c(\infty) = B_c$. Hence there is no bifurcation when $B_d B_c < 1$, in agreement with the general percolation condition for bipartite trees [18].

The key advantage of our method is that we can easily investigate the parallel processing capabilities of a bipartite graph with arbitrary degrees $\{e_p\}$. Here we have a pattern-dependent dilution of the links $P(\xi) \propto \prod_{\ell,\mu} P(\xi^\ell_{\ell,\mu}) \prod_{\ell,\mu} \delta_{e_{\ell,\mu}} \sum_{z\mid\xi^\ell_{\ell,\mu}}$ with

$$P(\xi^\ell_{\ell,\mu}) = \frac{e_p}{2N} (\delta_{z_{\ell,\mu}, 1} + \delta_{z_{\ell,\mu}, -1}) + (1 - \frac{e_p}{N}) \delta_{z_{\ell,\mu}, 0}$$

leading to $P(d) = \text{Poisson}(\alpha(c))$ while $P(e) = P^{-1} \sum_{\ell,\mu} \delta_{e_{\ell,\mu}}$. If we keep the mean degree fixed at $\langle e \rangle = c$, the critical point for $\beta \to \infty$ is found at

$$B_d B_c = \alpha c((c^2/c - 1) + \text{Var}(c)) = 1$$

while for large $\alpha$ one obtains for the critical line $\beta c^{-1}(\alpha) \approx \sqrt{\pi} \sqrt{c(c - 1) + \text{Var}(c)}$. Similar results are obtained with soft constraints $e_p$ on the degrees, i.e. by dropping the delta function constraint in $P(\xi)$ before (13): one now finds $B_d B_c = \alpha (c^2 + \text{Var}(c))$ and $\beta c^{-1}(\alpha) \approx \sqrt{\pi} \sqrt{c^2 + \text{Var}(c)}$. In both cases, the region where parallel retrieval is obtained is larger for degree distributions with smaller variance; the optimal situation occurs when all patterns have exactly the same number $c$ of non zero entries (Fig. 4, right). Notably, then, scale free networks, which perform best for information spreading [18], are not optimal for information processing (see SM). For the special case of homogeneous dilution $P(\xi^\ell_{\ell,\mu}) = \pm 1 = c/(2N)$ we easily recover previous results [8]: the distributions of pattern degrees $e$ and neuron degrees $d$ are Poisson$(c)$ and Poisson$(\alpha c)$ respectively, so $B_d = \alpha c$, $B_c = c$ so that there is no bifurcation for $\alpha c^2 < 1$. The network acts as a parallel processor here for any $\beta$ because the bipartite network consists of finite clusters of interacting spins in which there is no interference between different patterns [8]. At higher connectivit, the critical line defined above by $A = 1$ indicates the temperature above which this lack of interference persists even though the network now has a giant connected component. Fig. 4 (left) compares theory to PD results, where we locate the transition as the onset of nonzero.
second moments of the field distributions. The impact of the transition on the overlap probability distribution of a pattern with fixed $c$ can be seen from the PD results in Fig. 3 (middle and right panels). Crossing the transition line, parallel retrieval is accomplished at low temperatures, but it degrades when $\alpha$ is increased (see shrinking peaks in the middle panel), or $c$ is increased, eventually fading away for sufficiently large $\alpha$ and $c$ (right panel).

![Fig. 4: Transition lines (theory, with symbols from PD numerics) for different pattern degree distributions. Left: $c \sim \text{Poisson}(c=1)$. Right: changing $P(\psi)$ at constant $\langle e \rangle = 3$; $P(\psi) = \delta_{0.3}$ (blue); $P(\psi) = \delta_{2} + \delta_{3} + \delta_{4}$ (green); $P(\psi) = \delta_{2} + \delta_{4}$ (pink); $P(\psi)$ power law as in preferential attachment graphs, with $\langle e^2 \rangle = 21.66$ (orange).](image)

We can also analyze the case of asymmetric patterns, where we take for the nonzero pattern entries $P(\xi_i^\mu = \pm 1) = (1 \pm a)/2$ with a degree of asymmetry $a \in [-1, +1]$. One can show (see SM) that at zero temperature the bifurcation occurs when $B_\alpha B_\beta = a^{-2}$; when $a$ tends to zero the transition point goes to infinity and we retrieve the symmetric case. Beyond the bifurcation, non-centered field probability distributions (see Fig. 5) produce a nonzero global magnetization typical of ferromagnetic systems. However, a bifurcation towards growing field variances at zero means can also still occur. The physical bifurcation is the one taking place first on increasing $\beta$; Fig. 6 shows that at large $\alpha$ this is the one to growing means, at small $\alpha$ to growing variances.

To our knowledge, ours is the first study to quantify analytically the impact of heterogeneous degree distributions on the resilience of parallel processes on graphs. Degree heterogeneities in monopartite graphs [19] are well known to affect their resilience [20, 21] and the dynamics of processes that they support (transport, epidemics etc.) [22, 23], due to the fact that hubs enhance the spread of information across the network. Our method paves the way for exploring similar qualitatively important phenomena in bipartite systems.

AB thanks GNFM-INdAM (Progetto Giovani Agliari2014) and DT acknowledge Sapienza University (Avvio alla Ricerca 2014) for financial support. PS acknowledges funding from the EU under REA grant agreement nr. 290038 (NETADIS). Elena Agliari and Yamir Moreno are acknowledged insightful discussions.

![Fig. 5: Histogram of the fields $\psi$ in the ferromagnetic region, for $c = 1$, $\beta = 1$ and different levels of bias: $a = 0.9$ with $\alpha = 9$ (left) and $a = 1$ with $\alpha = 8$ (right). Field distributions are obtained by PD starting from positive fields, to break the gauge symmetry. For $a = 1$ (right) there are only positive fields as expected: when all patterns have positive entries there are no conflicting signals, even above the percolation threshold.](image)

![Fig. 6: Transition lines to growing field means (theory, green) and variances (theory, red), showing a good match to numerical PD data (dots); here $c = 1$ and pattern bias $a = 1, 0.95, 0.9$ from left to right. The first line to be crossed from high $T = \beta^{-1}$ gives the physical transition.](image)

[24] Eq. (1) corresponds to Gaussian $\tau_i$ in the bipartite SG; for discrete $\tau_i$ one would average over $z = \pm \beta$ with probability $1/2$ each.