This electronic thesis or dissertation has been downloaded from the King’s Research Portal at https://kclpure.kcl.ac.uk/portal/

Topics in Discrete Stochastic Processes on Graphs with Application to Computer Science

Kang, Nan

Awarding institution:
King’s College London

The copyright of this thesis rests with the author and no quotation from it or information derived from it may be published without proper acknowledgement.

END USER LICENCE AGREEMENT

Unless another licence is stated on the immediately following page this work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International licence. https://creativecommons.org/licenses/by-nc-nd/4.0/

You are free to copy, distribute and transmit the work

Under the following conditions:

- Attribution: You must attribute the work in the manner specified by the author (but not in any way that suggests that they endorse you or your use of the work).
- Non Commercial: You may not use this work for commercial purposes.
- No Derivative Works - You may not alter, transform, or build upon this work.

Any of these conditions can be waived if you receive permission from the author. Your fair dealings and other rights are in no way affected by the above.

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.

Download date: 04. Mar. 2022
Topics in Discrete Stochastic Processes on Graphs with Application to Computer Science

Nan Kang

Thesis is submitted to King’s College London, University of London, in partial fulfilment of the requirements for the degree of the Doctor of Philosophy

Department of Informatics
KING’S COLLEGE LONDON
June 2021
To my mother, Jing Qi
I would like to express my deep and sincere gratitude to my esteemed supervisors, Professor Colin Cooper and Professor Tomasz Radzik, for providing invaluable guidance throughout my Ph.D. study and related research. It was an incredible privilege and honour to work and study under their supervision. Their immense knowledge and motivation have encouraged me in all the time of my Ph.D. journey; and their patience and plentiful experience have given me more power and spirit to excel in the academic research. I appreciate all their contributions of time, effort and ideas to make my research experience productive and stimulating.

Colin’s expertise widened my research from various perspectives. His insightful feedback pushed me to sharpen my thinking and brought my work to a higher level. More than assisting with the theoretical, technical and practical aspects of my research, which have been instrumental, he encouraged me, challenged me, treated me with respect and as an equal, provided opportunities, and has fostered enormous growth in me as a researcher. I am profoundly grateful for having worked with him during these formative years.

Tomasz continuously provided encouragement and was always willing and enthusiastic to assist in any way he could. His intellectual suggestions were very inspirational. The joy and enthusiasm he has for his research was contagious and motivational for me, even during tough times in the Ph.D. pursuit.

I would also like to thank Dr. Nicolás Rivera for his insightful advice and the stimulating discussions. His motivation, dynamism, vision and sincerity have deeply
inspired me. Working collaboratively with him has shaped me to be a more skilled and knowledgeable researcher with a better understanding of the professional context in which my research sits.

My appreciation also goes out to my family and friends for their tremendous love, support, encouragement and understanding all through my studies.
Abstract

This thesis is devoted to the study of stochastic processes on graphs, which is a system of random variables that evolve over time while undergoing probability fluctuations according to probabilistic rules. We study four discrete stochastic processes throughout this thesis.

In the first process, we investigate the Best-of-3 voting dynamics for two colours (or opinions) in a class of non-complete graphs. As the main result, we show that if the graph has minimum degree $n^\Omega(1/\log \log n)$, and each vertex is initially coloured red with probability $1/2 + \delta$, blue otherwise, then the dynamics converges to a monochromatic red configuration within $O(\log \log n) + O(\log \delta^{-1})$ steps with high probability. This, for sufficiently large $\delta$, implies a doubly logarithmic convergence time in sufficiently dense graphs. The main purpose of this model is to close the gap between Best-of-3 and Best-of-5 in terms of consensus time, as it was known from precious work that both 2-majority and 3-majority take $O(\log n)$ time while 5-majority takes $O(\log \log n)$ time.

While there has been a rich line of research concerning majority consensus population protocols, we consider a diversity or task allocation population protocol, as many real-world systems evolve in an opposite way to reaching consensus: they converge to a state of diversity. In the second process - Diversification protocol, there are $n$ agents each with one of $k$ colours, and each colour $i$ has an associated weight $w_i \geq 1$. The main goal is to achieve diversity such that the system has roughly $nw_i/w$ agents with colour $i$ at any given time, where $w$ is the sum of the weights of all $k$
colours. This protocol converges to an approximately balanced state in $O(w^2 n \log n)$ steps, which is optimal up to the $w$ term, according to the lower bounds of rumour spreading.

In the third process, we study two switch Markov chains, called Triangle-switch processes, for sampling random regular graphs that favour triangles, in which local edge transformations are applied by choosing a pair of distinct edges $xy$ and $ab$ of the graph uniformly at random, and replacing with a uniformly chosen perfect matching of the vertices $\{x, y, a, b\}$ to create or remove triangles in the graph. Note that the degree of any vertex is preserved in each switch move. With restricted attention to 2-regular graphs, we analyse the properties of the transition matrix of the chain and investigate the spectral properties and conductance in order to have a deeper insight on the mixing time of the associated random walk.

As for the last process, we propose three models for generating random spanning trees, which can be considered as a type of random recursive forest and we call it the Permutation forest. A permutation forest is produced by first randomly ordering the vertices of the underlying graph $G$. For each vertex $u$ we check all its neighbours higher than itself in the permutation ordering and choose the first, or the last, or a random neighbour according to three protocols: Pick-First, Pick-Last, and Pick-UAR, respectively. If no neighbour appears after $u$, then $u$ becomes the root of a tree. We study the structure of the random forests generated under these tree protocols from an Erdős Rényi random graph $G(n, p)$, and provide analytical results for various properties, namely, the number of components, maximum component size and maximum tree height. It is worth mentioning that the expected number of components in a permutation forest depends only on the degree sequence, and not on the structure of the underlying graph.
## Contents

**List of Figures**

1. Introduction  
   1.1 Applications of stochastic processes  
   1.2 Random walks  
   1.3 Markov chains  
   1.4 Interacting particle systems  
   1.5 Random graphs and randomised algorithms  
   1.6 Thesis contribution and outline  

2. Best-of-Three Voting on Dense Graphs  
   2.1 Introduction  
   2.1.1 Main results  
   2.1.2 Illustrations  
   2.2 Model and proof strategy  
   2.2.1 Lower levels: Sprinkling process  
   2.2.2 Upper levels: Duplicating process  
   2.3 Open problems for further research  

3. Diversity, Fairness, and Sustainability in Population Protocols  
   3.1 Introduction  
   3.2 Protocol and results
3.2.1 Properties and definition of the protocol .......................... 53
3.2.2 Main results ................................................................. 57
3.3 Analysis of the Diversification protocol ................................. 59
  3.3.1 Phase 1: The rise of the minorities .................................. 61
  3.3.2 Phase 2: Reaching equilibrium - proving diversity .............. 64
  3.3.3 Phase 3: A finer equilibrium - achieving fairness ............... 68
3.4 Conclusions and future work ............................................ 74
3.5 Omitted proofs ............................................................... 75
  3.5.1 Proof of Section 3.3.1 .................................................. 75
  3.5.2 Proofs of Section 3.3.2 ................................................ 78
  3.5.3 Proof of Section 3.3.3 ................................................ 89
3.6 Auxiliary results ............................................................. 92

4 Triangle-Switch Processes on 2-Regular Graphs .......................... 94
  4.1 Introduction ..................................................................... 94
  4.2 Basic triangle-switch process on $G^*(n, 2)$ ......................... 98
    4.2.1 Basic triangle-switch, u.a.r. variant .......................... 99
    4.2.2 Variant $P(\text{make } C_3) = p$, $P(\text{break } C_3) = q$ .... 104
    4.2.3 Variant u.a.r. triangle-switch, ignore $C_4$ and $C_5$ ....... 105
    4.2.4 Triangles and the total number of cycles, u.a.r variant .... 108
  4.3 Linear triangle-switch process, an alternative triangle process .... 113
    4.3.1 Case $p > q$ ......................................................... 118
    4.3.2 Case $p = q$ ......................................................... 118
    4.3.3 An investigation of the spectral gap and mixing time of matrix $A$ in the case $p = q$ .................. 121
    4.3.4 A theorem of Fill, and its application to the mixing time of $A$ in the case $p = q$ .............. 123
    4.3.5 Experimental results .............................................. 128
5 Permutation Forests

5.1 Introduction ......................................................... 136
5.2 General results ..................................................... 141
  5.2.1 Expected number of components in $G(n,p)$ ................. 143
  5.2.2 Remarks on Theorem 45 ...................................... 144
5.3 Pick-First model ................................................... 146
  5.3.1 Monotone increasing paths .................................. 147
  5.3.2 Expected degree of a vertex ................................. 148
  5.3.3 Size of the largest component in Pick-First model ...... 150
5.4 Pick-Last model ................................................... 152
  5.4.1 Expected size of the largest component in Pick-Last model . 154
  5.4.2 Maximum height of trees in Pick-Last model .............. 158
5.5 Pick-UAR model .................................................... 164
  5.5.1 Expected size of the largest component in Pick-UAR model . 165
  5.5.2 Maximum height of trees in Pick-UAR model .............. 169
5.6 Experimental results for Permutation Forest models ............ 172
  5.6.1 Permutation forests around $p = 1/\sqrt{n}$ ................. 173
  5.6.2 Largest component as a function of $p$ ..................... 174
  5.6.3 Maximum tree height as a function of $p$ ................. 177
List of Figures

2.1 Example: A graph with initial colouring on its vertices. 29
2.2 Best-of-1 process on Figure 2.1. 30
2.3 Best-of-3 process on Figure 2.1. 31
2.4 Example of the construction of Sprinkling process. (Vertices are ordered from left to right, and collisions are marked in green.) 37
2.5 Part of an example DAG with clashes. 43
2.6 Duplicates of vertices in clash and their subgraphs. 43
3.1 Summary of the three phases of the analysis. 61
4.1 A switch, in which edges $xy$ and $wz$ are replaced by $xw$ and $yz$ 96
4.2 A triangle-switch. 97
4.3 A typical output of the basic u.a.r. triangle-switch on the subset $\mathcal{G}^*$ of 2-regular graphs. Here $n = 300$ vertices, and the simulation was run for 10,000 steps. 100
4.4 The number of triangles during 100,000 simulation steps for $k = 10,000$, sampled every 100 steps. The starting mix was $n/2$ vertices on triangles, and $n/2$ on hexagons. 113
4.5 Rule A: make a triangle with probability $p$. 114
4.6 Rule B: merge a triangle with probability $q$. 114
4.7 The underlying digraph $D$ of the transition matrix $A^t$. 122
5.1 Permutation forest obtained from the Pick-First process. Underlying
go graph $G$ shown on left-hand side, with vertices ordered 1, 2, 3, 4, 5 . . .
5.2 Permutation forest obtained from the Pick-Last process.
5.3 Permutation forest obtained from the Pick-UAR process.
5.4 Order of magnitude results for maximum component size in the three
models.
5.5 Order of magnitude results for maximum tree height $H$ in the three
models.
5.6 Pick-First process at $p = \sqrt{c/n}$ for $c = \frac{1}{5}$, 1, 5 and $n = 10,000$.
5.7 Pick-UAR and Pick-Last processes at $p = \frac{1}{\sqrt{n}}$ and $n = 10,000$.
5.8 Order of magnitude results for maximum component size.
5.9 Emergence of component of size about $n$ in Pick-First model for $n =
1000$. The vertical line is at $p = \frac{1}{\sqrt{n}}$.
5.10 Details of emergence of component of size about $n$ in Pick-First model
for $n = 1000$ and $n = 100,000$. The vertical line is at $p = \frac{1}{\sqrt{n}}$.
5.11 Sizes of largest (blue) and the second largest (red) component as a
function of $p$, with $n = 10^5$. Note: $p = \frac{1}{\sqrt{n}}$ corresponds to $p = 0.0032$.
5.12 Size of largest component as a function of $p$. For UAR, $n = 10,000$,
for Last, $n = 100,000$.
5.13 Order of magnitude results for maximum tree height $H$ in the three
models.
5.14 Size of longest path in Pick-First as a function of $p$, for $n = 100,000$.
5.15 Size of longest path in Pick-Last model as a function of $p$, for $n =
10,000$, $n = 100,000$ and $n = 1000,000$. The horizontal line is at
log $n$/log log $n$.
5.16 Size of longest path in Pick-UAR model as a function of $p$, for $n =
1,000$, $n = 10,000$ and $n = 100,000$. The horizontal line is at $2 \log n$. 
Chapter 1

Introduction

In order to analyse dynamic real-world phenomena and real-life situations, one mathematical approach used by researchers is to construct models that fit the available information and investigate the model to gain an understanding of the process. Many of these models involve randomness in their evolution rules and between interacting autonomous entities; in other words, the processes evolve, change, and develop over time in a random but prescribed manner, often resulting from randomised decisions and interactions among various autonomous agents. Such processes are called *stochastic processes* or *random processes*, which are mathematical models used in probability theory to describe the evolution of random systems over time. They can be categorised into discrete-time stochastic process and continuous-time stochastic process, depending on whether the system is observed or evolves at discrete time points or continuously as time passes.

1.1 Applications of stochastic processes

Stochastic processes, which started from research on physical phenomena modelled as a class of random variables evolving over time, have been intensively studied over decades. This has led to the emergence of new and fascinating mathematical models and probabilistic techniques. The theory and methodology of stochastic processes
have been widely applied to various fields, such as physics, mechanics, computer
science, image and signal processing, biology, finance, among others.

We next illustrate the general topic with a few examples. In computer science,
stochastic models are often used to describe dynamic networks, and investigate long-
term behaviour in order to compute quantities of interest such as probabilities, ex-
pectations, and other distributional characteristics.

Markov Chain Monte-Carlo (MCMC) is a sampling method for obtaining infor-
mation about probability distributions (see [58] for a more technical exposition). The
MCMC method generates random variables based on a random number generator for
systematic random sampling from complicated and often intractable probability distri-
butions, often where the next sample is dependent on the current existing sample.
Often direct computation is not feasible as it may be complex, time-consuming, or
too expensive. No matter what the distribution, a Markov chain can be set up so that
the steady-state distribution of the random variables produced by MCMC has the
desired unconditional distribution from which it is sampled. Therefore, one can char-
acterise a probability distribution using MCMC without knowing its mathematical
properties.

Queueing systems, which originated from modelling the number of phone calls at
the Copenhagen Telephone Exchange [99], consist of several servers to perform tasks,
and are used to describe dynamics that involve spontaneous arrivals and departures,
random waiting and service times, which occur in systems such as computers execut-
ing tasks, or customers waiting to be served.

Gaussian processes have received increased attention in the field of machine learn-
ing as they provide a practical approach in challenging learning problems. They are
commonly used for regression and classification tasks with reliable estimation of un-
certainty [93].

In biology, a model of great interest is the Moran process [87], which models the
spread of genetic mutations in a population of a community. It is a discrete-time
stochastic process on graphs, as adapted by [78]. In a Moran process, there are two
types of nodes, either mutant or non-mutant, and each type has a fitness (a positive
real number \( r > 0 \) for mutants and 1 for non-mutants). At each time-step, a random
node \( u \) (with or without mutation) is selected, with probability proportional to its
current fitness, to reproduce a copy of itself and eliminate the type of a random
neighbour. This neighbour is chosen uniformly at random among all the neighbours
of \( u \) and will be replaced by the copy of node \( u \).

Percolation processes are used to study phenomena that may spread (see [95] for
the first theoretical investigation), such as fire propagation, rumour spreading among
people, disease spreading throughout a population, or a virus spreading between
computers.

Other examples of stochastic processes include renewal processes, which study the
successive occurrences of events in terms of random variables, such as the lifespan of
lightbulbs, the emission of radioactive particles; Poisson processes which occurs from
the study of the arrivals of phone calls in a certain period of time; and Brownian
motion which models the constantly undergoing random motion or oscillations of
particles in a medium, such as the movement of dust particles or pollutants in the
air, or the motion of pollen in the water due to random collisions with water molecules.

Since the field of stochastic processes is large and evolving, we cannot give an
exhaustive study and accurately capture all the state-of-art topics in a short in-
troduction. Rather than attempt to cover everything, we focus on certain specific
examples of discrete stochastic processes on graphs relevant to this thesis. The main,
but not necessarily disjoint, topics include: random walks, Markov chains, interacting
particle systems, random graphs and randomised algorithms.
CHAPTER 1. INTRODUCTION

1.2 Random walks

One significant type of stochastic processes are discrete random walks, which are the realisation of random variables occurring in time space that depicts the random trajectory of particles or messages moving from one vertex to another in a graph or through a continuous medium. Measured as a sequence of increments at discrete time-steps, the motion of the particle is the fundamental topic in this problem. There has been extensive research on the evolution of the displacement of the random walk from its initial position and the long-term asymptotic behaviour of the walk. Due to their simple rules and rich behaviour, random walks are important as a model for a variety of other random processes, both in theory and in real life applications. Studying the associated random walk allows us to model various processes that evolve through small random changes and thus quantify different properties of interest.

A well-known and classical simple random walk on one-dimensional space is the ‘Gambler’s ruin’ problem. This problem which was studied by Pascal [49] in 1656 initiated the theory of random walks and diffusion processes (see [66] for the relationship between random walk theory and diffusion processes). Many approaches using random walks have been proposed for exploring graphs and testing graph connectivity. Other computer science applications based on random walks, include the work of Israeli and Jalfon [67] on a randomised protocol (called the self-stabilising mutual exclusion algorithm) that provides a solution to the problem of fault tolerance in distributed systems. A substantial work on weighted discrete walks, reversible and otherwise, is provided in [4].

Multiple random walks consist of a set of particles each of which makes an independent simple random walk on an undirected graph. There are two types of particles: oblivious or interactive particles, according to the behaviour arising when they meet at the same vertex. In oblivious random walks, particles have no interactions upon meeting; while in random walks with interactive particles, they can interact in various ways such as reproducing, annihilating, coalescing, or exchanging
In a coalescing random walk, a particle is initially located on each vertex of a connected undirected graph. The system is often modelled with an independent Poisson clock at each vertex that rings at rate 1. When a clock rings, the particle on that vertex moves to an adjacent vertex and coalesces with any other particle at the same vertex if it is already occupied. The coalesced particles proceed as one from the time when they meet. The behaviour of coalescing random walks on finite graphs is of great interest in the field of computer science. Cooper et al. [33] studied the coalescence time of random walks on a variety of different finite graphs. Cox [40] studied the expected time for the random walks to coalesce into a single particle (coalescence time) on $d$-dimensional tori.

1.3 Markov chains

A Markov process is a stochastic process that satisfies an elementary property, called the Markov property or memoryless property: namely that the future of the system is independent of the past given that the present state of the system is known. In a probabilistic sense, the present state of the system contains all the relevant information required to predict the future behaviour of the system. In other words, a Markov process has no memory of the past or how the present state was reached; its future only depends on the present state. If the state space is discrete (finite or countable), such processes are called Markov chains, i.e., a Markov chain is a discrete-time stochastic process possessing the Markov property. Typically the transition probabilities of Markov chains are time independent, and thus do not depend on the particular step of the process, but only on the current state.

Actually, Markov chains are a wide generalisation of random walks; and random walks are a special class of Markov chains with specific properties. Thus the theory of finite Markov chains does not differ much from the theory of random walks on graphs. Indeed Markov chains can be viewed as random walks on a directed graph.
with weighted edges. Comparably, time-reversible Markov chains can be considered as random walks on undirected graphs. A powerful technique called coupling, through which random variables can be compared and approximated, is useful in many key areas of probability theory such as Markov chains, random walks, percolation, interacting particle systems and diffusion processes. In Chapter 3, we couple the trajectory of the states of agents in a system with a Markov chain which represents the system in equilibrium, to measure hitting times.

A Markov chain can be described by its initial distribution and the transition probabilities. Since in most cases it is intractable to compute the transient distribution at the $t$-th step, another way to characterise a Markov chain is studying the limiting behaviour, which is determined by the stationary distribution. The mixing time of a Markov chain, roughly speaking, is the necessary number of time-steps $T$ such that the distribution of the state after $T$ steps will be ‘close’ (in some well-defined meaning of ‘close’) to its stationary distribution. However, it might happen that a Markov chain has a stationary distribution but does not converge to it, namely, if the chain is periodic (consider a random walk on a bipartite graph as a simple example).

A common tool for bounding the mixing time of a Markov chain is the spectral gap of the underlying transition matrix of the chain. This allows one to prove convergence and bound the closeness of the distribution at time $t$ to the limit given a starting distribution. The spectral gap is closely related to a term ‘conductance’ in the field of electrical networks. In fact, time-reversible Markov chains and thus random walks on undirected graphs have an intimate connections to electrical networks. The conductance of a graph, also called the Cheeger constant [25], is a geometric parameter measuring edge expansion. It controls the speed of convergence of random walks on an undirected graph to the stationary distribution. For a random walk on a graph $G$ whose current position is vertex $v \in V(G)$, it moves to a neighbour $u$ of vertex $v$ with probability proportional to the conductance of the edge $(u, v) \in E(G)$. The notion of conductance proposed by Jerrum and Sinclair [69] is a useful way to
prove rapid mixing of Markov chains, and it bounds the mixing time up to square factors.

We study a switch Markov chain with certain rules in Chapter 4, and investigate the mixing time of the associated random walk by the spectral gap conductance. Switch Markov chains are known to be rapidly mixing for regular graphs and graphs with a given degree sequence. It was shown in [73] that the mixing time is polynomial in $n$ for random regular bipartite graphs, using a Markov chain whose transitions are defined by switching two selected edges under certain ways. Cooper et al. [31] further extended this switch chain for generating random $d$-regular graphs and proved that the mixing time is bounded by $O(d^{24} n^{9} \log n)$. Then, Greenhill [62] proved the switch chain is rapidly mixing for non-regular cases, and gave a bound on the mixing time by $d_{\text{max}}^{14} M^{10} \log M$, where $d_{\text{max}}$ is the maximum degree such that $3 \leq d_{\text{max}} \leq \sqrt{M}/4$ and $M$ is the sum of degrees.

1.4 Interacting particle systems

Interacting particle systems (IPS) concern the rigorous analysis of certain types of models for complex phenomena involving a large number of interrelated components, whose motion and interaction is governed by certain local rules. Conceptually, it can be described as a stochastic process on a network, where particles meet pairwise and exchange information or modify their state (or opinion), in a certain way depending on their previous state. From a mathematical perspective, interacting particle systems can be considered as an extension of Markov processes, often consisting of countably many particles which, if without interactions, would locally evolve as Markov processes with countable states. However, due to the interactions, particles may change the transition rates, states or even existence of other particles, in which case an individual particle may no longer evolves as a Markov process, since the its transition depends on the entire configuration of the system. Thus, while some links with Markov processes are maintained, interacting particle systems may differ from
the standard Markov processes and are often used to model stochastic phenomena with spatial structures.

Interacting particle systems are by now a mature yet still growing area of probability theory, whose ergodic and limiting behaviour is a central topic and key area of research (see [48] for a comprehensive review). Connections have been established in recent years between interacting particle systems and a number of other subjects of mathematical research from both theoretical and applied sides. One tractable model of interacting particle systems that has been extensively studied is the voter model, introduced by Clifford and Sudbury [30]. The voter model is a discrete-time stochastic process on a graph, where each vertex is initially assigned a distinct opinion. At each time-step, every vertex selects a random neighbour and adopts its opinion. Voter models are analogous to a system of coalescing random walks in the case that each particle is initially located at each vertex, i.e., a walk started from each vertex. The correspondence between the voter model and coalescing random walks was first established in [64], where the authors stated the duality relationship that the consensus time of particles with different initial opinion corresponds to the coalescence time of random walks.

The problem of reaching consensus on a graph by means of local interactions is a flourishing area of research in the field of distributed computing. The consensus time of the voter model and the coalescence time of random walks were studied in [40] for $d$-dimensional tori. A variant of the voter model is considered in [46], where a two-party setting is applied such that each particle has binary opinions initially. In [63], Hassin and Peleg proved that the probability that a particular opinion wins in the voter model under two-party setting is proportional to the degree of that opinion in the initial configuration. The two-party voter model was also studied in by Cooper et al. for $d$-regular expanders, they proved in [34] that the consensus time is $O(\log n)$ provided the initial difference between the sizes of the two opinions is sufficiently large. Apart from voting, another well-known updating rule of consensus problems
is the *majority* rule, under which a vertex adopts the majority opinion of selected
neighbours rather than adopting one random opinion. We study a three-majority
model under two-party setting on general expander graphs, in which each vertex,
with initial binary opinion, adopts the majority opinion of three random neighbours
at each step; and investigate the consensus time such that the system agree to the
initial majority in Chapter 2.

1.5 Random graphs and randomised algorithms

Much attention has been paid to random graphs, mainly concerning binomial random
graphs $G(n, p)$ and the uniform random graphs $G(n, m)$. The concept of random
graphs was first introduced by Erdős and Rényi [50], and Gilbert [57], as an extension
of the probabilistic approach to discover certain graph properties and behaviour. In
the $G(n, m)$ model, a graph is chosen uniformly at random from the probability
space of all graphs with $n$ vertices and $m$ edges, where each graph is chosen with
equally likely probability, i.e. $1/\binom{n}{2}$. The $G(n, p)$ model is a distribution over
a sample of $n$-vertex graphs, generated by adding each edge independently with
probability $p$. Thus the expected number of edges is $\binom{n}{2}p$. Relating back to $G(n, m)$
graphs, equivalently, the probability of generating a random graph with $n$ vertices
and $m$ edges is $p^m(1 - p)^{(n) - m}$. Due to the flexibility in choosing the value of
$p$, the $G(n, p)$ random graphs tend to be prevalent in random graph theory (see [22]
and [68] for comprehensive studies). Researchers have gained insight on the numerous
properties of $G(n, p)$ random graphs, especially the phase transition behaviour of
certain properties. There is a threshold around $p = 1/n$, where $G(n, p)$ transitions
quickly from having only logarithmic size components, to possessing a single linear
sized giant component and all other components are $O(\log n)$.

Every connected graph has a spanning tree, which is a tree subgraph includ-
ing every vertex. Random spanning trees are extensively investigated probabilistic
structures in graph theory. It is a classic problem of great interest in computational
probability theory to generate a random spanning tree of a graph according to certain
distribution, such as uniform distribution or a distribution given by the weights on the
edges of the graph. A stunning connection, independently discovered by Broder [23]
and Aldous [6] gives rise to an algorithm for sampling a random spanning tree by
simply simulating a random walk on the graph until it covers all vertices, and only
the first incoming edges (corresponding to the first visit to vertices) are added to
the spanning tree. Kruskal revealed a strong connection between the Erdős-Rényi
random graphs and the minimum spanning trees in [76], that a forest \( F(n, p) \) can be
generated such that the sizes of the connected components exhibit the same phase
transition as \( G(n, p) \) when \( p = p(n) \sim 1/n \). In Kruskal’s algorithm, a minimum
spanning forest can be built in \( m \log n \) running time by a greedy approach. Given a
weighted undirected graph, we first sort the edges by their weights in an increasing
order and copy all vertices from the original graph. Then we add an edge one by one
into the growing subgraph according to the ordering, if and only if it does not create
a cycle. We repeatedly add edges until all nodes are reached.

Random recursive trees are one of the simplest and well-investigated random
trees models, with numerous applications in theoretical computer science. A random
recursive tree is a rooted tree uniformly distributed among all recursive trees with
a given cardinality. A random recursive tree of \( n \) vertices labelled from 1 to \( n \) can
be constructed by first choosing a single vertex labelled 1 as the root. Then we
recursively attach a vertex \( u \) for \( 2 \leq u \leq n \) to a vertex \( v \), where \( v \) is selected
uniformly at random from the label set \( \{1, \ldots, u-1\} \). It was proved in [89] that
the height \( H \) of a random union-find (disjoint-set) tree (the distance from root to the
furthest-away leaf) satisfies \( H/\log n \to e \) as \( n \to \infty \). The highest degree \( d_{\max} \) of a
vertex in a random recursive tree is shown in [44] to be \( d_{\max}/\log n \to 1 \). Further in-
depth studies have been done independently in [2] and [59]. A uniform random forest
model is studied in [8], in which each path from the root is labelled with an increasing
sequence of labels. In their model, given a vertex set \( \{1, \ldots, n\} \), a random forests
 CHAPTER 1. INTRODUCTION

is constructed by first designating 1 as a root, and then determining the connection of it or re-selecting a new root. To do this, for each vertex \( i \in \{2, \ldots, n\} \), a vertex \( v \) is chosen from the set \( \{0, 1, \ldots, i-1\} \) with probability \( p_i(v) \). If \( v \neq 0 \) then \( i \) is attached to \( v \); if \( v = 0 \) then \( i \) is selected to be a new root. With most focus on the special case when \( p_i(v) = 1/i \), the authors provided results on the expected number of trees as well as the number of vertices of given degree.

We will introduce three models for generating recursive spanning forests from random graphs under different constructing protocols in Chapter 5. Under one of the three protocols, the forest generated from a \( G(n, p) \) exhibits threshold behaviours for different range of \( p \) and shows some analogy to \( G(n, p) \) random graphs when it comes to the emergence of a giant component. In the other two processes, there is no obvious threshold, and the size of the largest component grows directly with \( p \). We will also analyse the properties of the generated trees, such as the number of components, maximum component size, maximum tree height, and provide experimental results to compare the models.

1.6 Thesis contribution and outline

The essence of this work consists of the analysis of four discrete stochastic processes on graphs in detail, using the knowledge and techniques from probability theory, combinatorics, linear algebra, etc. for the analysis.

First, we investigate the Best-of-three voting dynamics for two colours/opinions (with initial colour difference of \( \delta \)) in a class of non-complete graphs. In this process, at each time-step, every vertex chooses three neighbours at random and adopts the majority colour. We show that the dynamics converges to the initial dominant colour in \( O(\log \log n) + O(\log(1/\delta)) \) steps with high probability, provided that the underlying graph has minimal degree \( n^{\Omega(1/\log \log n)} \). This, for sufficiently large \( \delta \), implies a doubly logarithmic convergence time in sufficiently dense graphs. There is a rich line of research concerning the analysis of simple distributed dynamics for the consensus
problem. While most of the research on this topic assume that the underlying graph is complete, we consider the non-complete case, where one technical issue arises from the fact that the analysis has to keep track of the configurations of the opinions around each vertex at each time-step since the graph is not complete. To manage this, we show how to model the way a vertex decides its colour as a colouring process of a random DAG with a single source and certain levels. Based on that the coupled process where the choices of colour within a certain time-step are pairwise independent - albeit less of them will in expectation be of the initially dominant colour, we majorise the distribution of the cardinality of the minority colour at a certain time by a binomial distribution. Once the minority colour has diminished enough, the colouring process is then analysed by dividing the DAG into two subgraphs. We then argue and show that after some additional time-steps with high probability any vertex is of initially dominant colour.

Next, we analyse the Diversification protocol for generating a weighted partition of a population protocol with random scheduling and random transition. Over the years, population protocols with the goal of reaching consensus have been studied in great depth. However, many systems in the real-world do not result in all agents eventually reaching consensus, but rather in the opposite: they converge to a state of rich diversity. We study an interesting twist on the majority consensus population protocols that have received extensive attention in recent years. In this new twist, a population protocol system of $n$ agents (vertices) on a complete graph is initialised with an arbitrary distribution of $k$ colours (tasks), where each colour $i$ has a corresponding weight $w_i \geq 1$, that can think of for now as a natural number, and $w$ is the sum of these weights. That is, each agent is initialised with a pair $(i, w_i)$ for some colour $i$. Steps are scheduled with uniform randomness. The primary goal is to achieve diversity, meaning the system converges for a while to a state in which for each colour $i$, the fraction of agents currently adopting colour $i$ is close to $w_i/w$. Two secondary goals include fairness, in which the fraction of rounds each agent
spends with colour $i$ is close to $w_i/w$ over a long period of time; and sustainability, a deterministic property that disallows any colour from permanently leaving the system. As with some of the classic majority consensus results, a pleasingly simple strategy achieves these three properties. The Diversification protocol has each agent maintain a ‘shade’ bit that is either 1 (dark) or 0 (light). All agents start with dark shades and will not change their colour unless their shade is light. The shift from dark to light occurs when an agent with colour $i$ interacts with another agent with colour $i$, at which point it will switch colours with probability $1/w_i$, making heavily weighted colours much less likely to enter a vulnerable light shade state. This strategy immediately provides sustainability, as these interactions are asymmetric - only the initiator flips the coins to determine whether or not to change shades, so there must always be at least one other agent with this colour in the system with a dark shade in a step where an agent shifts from light to dark (and changes colour).

From a technique perspective, roughly speaking, the progress towards good diversity is captured by potential functions which are shown to be super-martingales allowing the application of useful concentration bounds. We show that this protocol converges in $O(w^2n \log n)$ rounds to the state of diversity and will stay in this state for a large number of rounds with high probability.

Then, we introduce Triangle-switch processes on 2-regular graphs, which is a graph transformation based on local edge-switches to ‘make’ or ‘break’ triangles, where two edges exchange one endpoint without changing the degree of any vertex. Markov chains that perform random triangle switches are a probabilistic approach for generating random graphs. A basic principle in social networks, which is referred to as triadic closure, is that if two people have a common friend, then there is an increased likelihood that they will become friends in the future [92]. However, many standard random graph generating models like $G(n, p)$, $G(n, m)$ and $G(n, d)$ fail to take into account a high clustering coefficient of networks, meaning an increased number of triangles. Note that a clustering coefficient measures the connection of
neighbourhood. We study two processes that tend to produce many triangles, as triangles commonly exist in citation networks and they indicate mutual friendship in social networks. One process is called the basic triangle-switch process, in which we investigate different variants of the process and analyse the time for triangles to converge. The other process, the linear triangle-switch process, is an alternative process where we make a triangle with probability $p$ and break one with probability $q$. In this process, we consider two cases $p > q$ and $p = q$, and study the properties of the transition matrix. For the case $p = q$, we discuss some spectral properties, and have an insight on the mixing time of the associated random walk with experimental simulations of the process on the convergence.

Finally, we propose a process referred as Permutation forest process, for generating a random spanning forest of a given graph $G$ in a simple and natural way, in which we introduce three protocols: Pick-First, Pick-Last, and Pick-UAR protocols. A permutation forest is produced by first choosing a random permutation on the vertices, and then for each vertex add an edge that connects it to the first or last or a random neighbour appearing after it in the ordering according to three protocols respectively. If a vertex appears after all its neighbours, no edge will be added, and such vertex is a root in the forest. Only the chosen edges for each vertex remain in the permutation forest and all other edges from the underlying graph are discarded. We analyse the expected number of connected components in the permutation forest for general case. Based on the observations that the number of components is the number of roots, and the probability that a given vertex is a root depends only on its degree, we prove that the expected number of components is a function of the degree sequence. Most of our attention is paid to the Erdős Rényi random graphs $G(n, p)$, in which case we establish concentration results on the expected number of components which is asymptotic to $1/p$, and investigate some structural differences, both theoretically and experimentally, that the permutation forests exhibit under the three protocols: maximum component size and maximum tree height.
CHAPTER 1. INTRODUCTION

I declare that, to the best of my knowledge, the results contained in this thesis are original and my own work, conducted under the supervision of Professor Colin Cooper and Professor Tomasz Radzik. Notably,

- Chapter 2 is based on joint work [72] with Nicolás Rivera, published at SPAA ’19;

- Chapter 3 is based on joint work [71] with Frederik Mallmann-Trenn and Nicolás Rivera, published at PODC ’21.
Chapter 2

Best-of-Three Voting on Dense Graphs

2.1 Introduction

Algorithms and protocols that solve consensus problems play an important role in distributed computing, analysis of social networks, etc. Usually, in these processes, vertices of a graph revise their opinions in a systematic and distributed way based on opinions of other neighbours, typically by sampling some of their neighbours. The aim of these protocols is to eventually reach a state where all vertices share the same opinion, and ideally this final state reflects the characteristics of the initial mix of opinions, e.g., the initial majority.

Among the protocols that solve consensus problems, one well-known protocol is the Best-of-$k$ model, in particular the cases with $k = 1, 2$ and $3$. In this protocol, we consider a graph $G = (V, E)$ with $n$ vertices, in which each vertex has an initial opinion (colour), and at each time-step, every vertex adopts the opinion of the majority of a sample of $k$ neighbours (uniformly with replacement). If there is no clear majority several rules can be applied, usually $i)$ the vertex keeps its opinion or $ii)$ the vertex picks a random one from the popular opinions among the $k$ neighbours.
The Best-of-1 model is the well-established voter model. This protocol solves the consensus problem in connected non-bipartite graphs. It is widely known that the probability of the system reaching consensus on a particular colour is proportional to the sum of the degrees of the vertices whose initial opinion is such colour. Specifically, a particular colour ‘wins’ with probability equals to the initial proportion of that colour in the configuration. Although the voter model can be used to solve consensus problems, it is not the desired protocol for applications where consensus to majority is required. Another shortcoming of the voter model is that the convergence time is linear, which is not fast enough in many cases.

The Best-of-$k$ model with $k = 2, 3$ partially overcomes the aforementioned problem, and it converges to majority under appropriate circumstances. Moreover, it converges considerably faster compared to the voter model. This model has been extensively studied when the underlying topology is a complete graph. In [15], the authors investigated the Best-of-3 dynamics breaking ties at random. They considered $q$ initial different opinions and proved that if the initial imbalance in the number of opinions between the first and second majorities is $\Omega(\min\{\sqrt{2q}, (n/\log n)^{1/6}\}) \sqrt{n/\log n}$, then consensus is reached with high probability in $O(\min\{q, (n/\log n)^{1/3}\} \log n)$ steps on the initial majority. Similar results can be proved for Best-of-2 dynamics [56].

When it comes to non-complete graphs, the Best-of-$k$ process seems very difficult to study. In this regard, [34] studied the Best-of-two process on a $d$-regular graph where each vertex has one out of two opinions (colours), say, red or blue. The authors showed that if the imbalance between the number of red and blue opinions is greater than $Kn \sqrt{1/d + d/n}$ initially, where $K$ is a large constant, then with high probability the process reaches consensus towards majority in $O(\log n)$ time-steps. In [35], the result was extended and refined to general graphs with large expansion. Denote by $R_0$ and $B_0$ the initial sets of vertices with red and blue opinions respectively. Then assume that $d(R_0) - d(B_0) \geq 4\lambda_2 d(V)$, where $d(X)$ denotes the sum of the degrees of the vertices in $X$, and $\lambda_2$ is the second largest absolute eigenvalue of the transition
matrix associated with the graph, then with high probability consensus is reached in \( O(\log n) \) rounds and opinion red wins. In regular graphs, the aforementioned condition implies an \( \Omega(n) \) gap between the sizes of the set \( R_0 \) and \( B_0 \). In [38], the result is extended to a larger number of initial opinions but with stronger assumptions.

Best-of-\( k \) with odd \( k \geq 5 \) was studied in [1] for the two-party model on random graphs with a given degree sequence. Under their setting, initially, each vertex is blue independently with probability \( 1/2 - \delta > 0 \), and red otherwise. Then, it is demonstrated that if \( \delta \) is sufficiently large, and \( k \geq \hat{d}_{\text{min}} \), then consensus is reached in \( O(\log \log n) \) time-steps and opinion red wins. Here \( \hat{d}_{\text{min}} \) is the effective minimum degree, which is the smallest integer that appears \( \Theta(n) \) times in the degree sequence.

### 2.1.1 Main results

In this chapter, we study the Best-of-\( k \) protocol in the particular case \( k = 3 \) under the two-party setting, where initially each vertex is blue independently with probability \( 1/2 - \delta > 0 \), otherwise red. By applying two models to analyse the process of a vertex updating its opinions, we prove convergence to majority in \( O(\log \log n) \) time-steps with high probability under certain conditions. Our main result is the following.

**Theorem 1.** Consider a graph \( G \) of \( n \) vertices with minimum degree \( d = n^\alpha \), where \( \alpha = \Omega((\log \log n)^{-1}) \), and suppose the initial opinion of each vertex is blue independently with probability \( 1/2 - \delta \), otherwise red, where \( \delta \geq (\log d)^{-C} \) for some \( C > 0 \). Then, w.h.p.\(^1\) the Best-of-Three protocol reaches consensus in \( O(\log \log n) + O(\log(\delta^{-1})) \) time-steps to the initial majority.

Compared to previous work, [1] is the closest to ours, as they also look at forward conditions to ensure doubly logarithmic consensus time towards majority and they work on non-complete graphs. In order to reach doubly logarithmic speed the graph requires a tractable local structure around each vertex, and we need to be able to

\(^1\)We use w.h.p. to denote ‘with high probability’ meaning with probability at least \( 1 - 1/n \).
keep track of the configuration of opinions around each vertex at each time-step. In this regard, the techniques used in [34] and [35] are not necessarily useful to tackle the problem in our case, even though they work on a large class of graphs. This is because in their work, the authors track the number of red (and blue) opinions instead of the actual configuration of the opinions of the vertices. Although tracking the number of red opinions is easier, the obtained result is not precise enough, and indeed, the technique gives \(O(\log n)\) steps towards consensus, which is not as fast as desired. Moreover, by tracking only the number of red vertices, we lose the extra information of how the opinions of vertices are distributed given by the fact that vertices start with randomised opinions. Additionally, the proof technique used in [35] works under adversarial setting where the adversary can reorganise the opinions among the vertices and keep the total number of each opinion fixed, thus the initial location of the opinion does not matter.

With respect to [1], our result is weaker in some respects and much stronger in others. First of all, both proofs are based on a sort of ‘time-reversal duality’ while instead of tracking the opinion of a vertex \(v\) from time 0 to fixed value \(T\), we obtain the opinion at time \(T\) by looking at the opinions at time \(T - 1\), and to determine those we look at the opinions at time \(T - 2\) etc. The process of keeping track of the opinion of a vertex is more complex, since it depends on several random variables which are dependent and thus difficult to analyse. To avoid dealing with such problem directly, [1] decided to work in the setting of \(k \geq 5\), which allows them to assume that certain vertices have the ‘bad’ opinion (i.e., minority) even if they actually have the ‘good’ opinion (majority). This helps them reduce the dependency caused by the opinion updating process so as to transform the real process into a simpler and easier to analyse process. When \(k \geq 5\), assuming one opinion is ‘bad’ does not particularly damage the speed of convergence to consensus, as we can hope that the other \(k - 1\) opinions have the ‘good’ majority. However, since for some vertex one ‘bad’ opinion is assumed, they rely on the other \(k - 1\) opinions getting the good majority quite
often in the process. In other words, they need to ensure a large initial imbalance between the initial numbers of the two opinions, thus their result holds only when the initial probability of being blue is way less than $1/2$ (i.e., $1/2 - q$ for sufficiently large constant $q \in (0, 1/2)$). Due to this reason, their result cannot be extended to $k = 3$, as assuming a ‘bad’ opinion will affect the majority significantly. Indeed, if one of the other two opinions is the ‘bad’ opinion, then the vertex will adopt it. Our proof partially overcomes this problem. We work with $k = 3$ and allow the initial probability of being blue to be $1/2 - \delta$, where $\delta$ is arbitrarily close to 0 and we can even choose it tending to 0 as the graph grows. Finally, our analysis works on the family of dense graphs with minimum degree $n^{\Omega(1/\log \log n)}$, whereas in [1] the authors consider sparse graphs of a given degree sequence with average degree $o(\log n)$ among other constraints. Note that both classes of graphs are disjoint.

2.1.2 Illustrations

In order to see how the Best-of-$k$ process proceeds and how different interacting rules affect the process, we will visualise the models by giving an example of a graph with initial two-party colouring under the Best-of-$k$ protocols. Figure 2.1 illustrates a graph of 9 vertices, each of which has an initial colour of either red or blue.

![Graph Illustration]

Figure 2.1: Example: A graph with initial colouring on its vertices.

Consider the synchronous Best-of-1 model on this graph, we investigate the changes
of opinions in one step for each node. Recall that in each round, every vertex picks one neighbour and adopts its opinion. As shown in Figure 2.2, after the first step, vertex 1 may remain to be blue since it has probability of $2/3$ to choose a blue neighbour, while vertices 2 and 4 may become red as it has 2 red neighbours out of 3. Although both of vertices 5 and 6 have probability $1/3$ to sample a blue neighbour, we assume they become blue in this example after the first step. Then, in the second step, vertices 1, 5 and 6 become red and vertices 4 and 7 adopt blue. After this, some neighbours of vertices 4 and 7 might become blue again if they select these vertices. The process keeps going like this and it takes a much longer time to achieve a consensus (if it is reachable) than the Best-of-3 model which will be discussed next.

![Figure 2.2: Best-of-1 process on Figure 2.1.](image)

Applying the Best-of-3 model on this graph, we know that the process ends in three steps, since after the second step only one blue vertex left whose neighbours are all red.

As for Best-of-$k$ with $k \geq 4$, the process is similar to the Best-of-3 with a fast consensus time. However, those models require more access to the neighbours which is expensive in application and may not to be possible in some cases.

In the current work, given a graph $G = (V, E)$ of $n$ vertices with minimum degree $d$, we study a particular example of such a protocol under a binary-choices setting.
(only two available choices) in the case of \( k = 3 \). The Best-of-three protocol is a synchronous discrete time process, in which each vertex randomly samples three neighbours with replacement at each time-step, and changes its colour to the majority colour of their three neighbours. With an initialisation that each vertex is assigned a colour \( B \) with probability \( p_0 \in [0, 1] \) and \( R \) with probability \( 1 - p_0 \), we would like to know how many rounds it takes for this protocol to reach a consensus such that the final colour of vertices is consistent with the initial majority.

### 2.2 Model and proof strategy

Let us recall our model and introduce some notation. Let \( G = (V, E) \) be a graph with \( |V| = n \), in which each vertex is blue (\( B \)) independently with probability \( 1/2 - \delta \), otherwise red (\( R \)) with probability \( 1/2 + \delta \).

Define a configuration of the opinions of each vertex at time \( t \) to be \( \xi_t = (\xi_t(1), \ldots, \xi_t(n)) \). The evolution of the opinions \( (\xi_t)_{t \geq 0} \) is as follows: For \( v \in V \), define \( \xi_0(v) \) as the initial opinion of \( v \). For each \( t \geq 1 \), every vertex \( v \) independently samples three random neighbours \( w^1_v, w^2_v, w^3_v \) with replacement, and sets \( \xi_{t+1}(v) = \text{majority}\{\xi_t(w^1_v), \xi_t(w^2_v), \xi_t(w^3_v)\} \). Note that the value of \( (\xi_{t+1}(v))_{v \in V} \) is only
CHAPTER 2. BEST-OF-THREE VOTING ON DENSE GRAPHS

determined by $(\xi_t(v))_{v \in V}$ plus some independent randomness, i.e., $(\xi_t)_{t \geq 0}$ is a Markov chain.

Our proof strategy is based on verifying that the probability of a vertex being blue after $T$ time-steps is $o(1/n)$, where $T = O(\log \log n)$, i.e. $\Pr(\xi_T(v) = B) = o(1/n)$ holds after $T$ time-steps. This gives $\Pr(G$ is monochromatic red at step $T) = 1 - \Pr(\bigcup_{v \in V} \{\xi_T(v) = B\}) = 1 - o(1)$ by applying a union bound. Therefore, the emphasis of our work is essentially on computing $\Pr(\xi_T(v) = B)$. From the definition of the process we know that $\xi_T(v)$ is determined by the opinions of three random neighbours of vertex $v$, say $w_1^v, w_2^v, w_3^v$, at time $T - 1$, i.e., $\xi_{T-1}(w_i^v)$ for $i = \{1, 2, 3\}$. Similarly, each $\xi_{T-1}(w_i^v)$ is determined by the opinions at time $T - 2$ of three random neighbours of $w_i^v$. We can continue recursively until the point where we query for the opinions of vertices at time 0, whose joint distribution is known. The above recursive (random) structure can be represented as a directed acyclic graph (DAG).

A DAG $H$ is a directed graph with no directed cycles. The in-degree of a vertex $v$ is the number of edges incoming to $v$, while the out-degree is the number of edges outgoing from $v$. A root in $V(H)$ is a vertex with in-degree 0. Note that there is only one root in the DAG constructed under the Best-of-three protocol. A leaf in $V(H)$ is a vertex with out-degree 0. Given $v \in V(H)$, we define $H[v]$ as the subgraph induced by all the vertices $w$ that can be reached from $v$, i.e., there exists a directed path from $v$ to $w$. As in the analysis of this Best-of-three process we will consider some random DAGs, and we shall denote them by $\mathcal{H}$ while $H$ is used to denote a fixed, deterministic DAG.

Let us construct the random voting-DAG associated with $\xi_T(v_0)$, where $v_0$ is the root by construction. In our work, we call it a voting-DAG to specify the DAG that has out-degree at most three. Define the set $Q_T = \{v_0\}$, and for $t \in \{0, \ldots, T - 1\}$ define $Q_t \subseteq V$ as the (random) subset of all vertices queried to determine the opinions of the vertices in $Q_{t+1}$ at time $t + 1$, e.g., $Q_{T-1}$ is the set of three random neighbours of $v_0$ required to determine $\xi_T(v_0)$, etc. We define the random voting-DAG $\mathcal{H} =$
\(\mathcal{H}_{v_0}\) by setting \(V(\mathcal{H}) = \{(v, t) \in V(G) \times \{0, \ldots, T\} : v \in Q_t\}\), and we say that 
\(((v, t + 1), (w, t)) \in E(\mathcal{H})\) if and only if \((v, t + 1) \in Q_{t+1}\) and one of the three vertices sampled by \(v\) to compute \(\xi_{t+1}(v)\) is \(w\). We divide the vertices of \(\mathcal{H}\) into levels, where level \(t \in \{0, \ldots, T\}\) contains all the vertices \((v, t) \in V(\mathcal{H})\). Note that each vertex at level \(t + 1\) has out-degree exactly 3, connecting to vertices at level \(t\) and that directed paths go from higher to lower levels.

Given a realisation \(H\) of \(\mathcal{H}\) with root \((v_0, T)\) we can simulate \(\xi_T(v)\) as follows. First, set the opinion of vertices \((v, 0) \in Q_0\) to be independently \(\mathbf{B}\) with probability \(1/2 - \delta\), otherwise \(\mathbf{R}\). Then recursively compute the opinions of vertices at level \(t + 1\) as the majority of the three neighbours at level \(t\), for \(t \in \{0, \ldots, T - 1\}\). Denote by \(X_H(v, t)\) the colour of vertex \((v, t)\) in \(H\). By summing up over all possible realisations of \(\mathcal{H}\), it is clear that the colour of \((v_0, T)\) has the same distribution as \(\xi_T(v_0)\), i.e.,

\[\mathbb{P}(\xi_T(v_0) = \mathbf{B}) = \mathbb{P}(X_H(v_0, T) = \mathbf{B})\,.
\]

Note that \(X_H(v_0, T)\) involves two independent sources of randomness. One source generates the voting-DAG \(\mathcal{H}\), and the other settles the colours of the leaves of \(\mathcal{H}\) (vertices at level 0) independently. Note that \((v_0, T) \in V(H)\) for any realisation \(H\) of \(\mathcal{H}\), so the random variable \(X_H(v_0, T)\) is well-defined. Finally, the process of defining \(X_H\) as above to colour the realisation \(H\) of \(\mathcal{H}\) is referred as the colouring process.

Given \(\mathcal{H} = H\) we have that the opinions of the vertices at level 0 are i.i.d\(^2\), as they do not depend on the structure imposed by \(H\), which however is unfortunately not true for levels \(t > 0\). Recall that \(H[(v, t)]\) is the subgraph induced by all the reachable vertices from vertex \(v\) at level \(t\). Indeed, it is clear that the colour \(X_H(v, t)\) depends only on the colouring of the leaves of \(H[(v, t)]\). Therefore the variables \(X_H(v, t)\) and \(X_H(w, t)\) with \((v, t), (w, t) \in V(H)\) are independent if and only if \(V(H[(v, t)]) \cap V(H[(w, t)]) = \emptyset\). Since the structure distribution of \(\mathcal{H}\) (and so its sample \(H\)) strongly depends on the underlying structure of \(G\), it is very unlikely to have the above independence condition for all vertices in \(H\). However, let us assume for a

\(^2\)We use i.i.d as the abbreviation for ‘independent, identically distributed’.
moment that the independence condition above is satisfied for all pairs of vertices sharing the same level. In such a case, \( H \) is a directed ternary tree with root \((v_0, T)\).

Let \( B_{t-1} \) be the number of blue vertices at level \( t - 1 \) among three random samples \( Q_{t-1}(v) \) of a vertex \( v \) at level \( t - 1 \), and \( b_t \) be the probability that any vertex at level \( 0 < t \leq T \) is blue. Observe that \( B_t \) follows a Binomial distribution \( \text{Bin}(3, b_t) \), thus the probability \( b_t \) follows the recursions:

\[
b_0 = \frac{1}{2} - \delta, \quad b_t = P(B_{t-1} \geq 2) = b_{t-1}^3 + 3b_{t-1}^2(1 - b_{t-1}) = 3b_{t-1}^2 - 2b_{t-1}^3.
\]  

(2.1)

Therefore, a simple computation shows that by choosing \( T = O(\log \log n + \log(\delta^{-1})) \) we get \( b_T = o(n^{-1}) \).

As the probability that \( H \) is realised as a ternary tree is considerably low, the above recursion does not necessarily reflect the true process. In order to deal with the inner dependency between levels, we divide the graph \( H \) into two subgraphs, one from level \( T \) to \( T' \) and another from level \( T' \) to 0, where \( T' = O(\log \log d) + O(\log(\delta^{-1})) + 1 \) is fixed in the forthcoming Lemma 4, and \( T = O(\log \log d) + T' \) is fixed in the forthcoming Lemma 7. For the subgraph from level \( T' \) to 0 we couple the colouring \( X_H \) with another colouring \( X'_H \) such that if colour \( B \) represents 1 and colour \( R \) represents 0, then \( X_H(v, t) \leq X'_H(v, t) \) for all \((v, t) \in V(H)\). The process where \( X'_H \) arises is called the Sprinkling process. By introducing an error term to deal with the dependency in this process, we have an easier way to study \( X'_H \) as opinions among vertices at the same level are independent given \( H \). Moreover, if we do not reveal \( H \) in advance (i.e., randomise over \( H \)), the distribution of the colours of vertices at level \( t \) for \( t \in \{0, \ldots, T'\} \) is i.i.d. and follows a recursion (see Equation (2.3)) similar to Equation (2.1). Unfortunately, this recursion (2.3) cannot be applied any further for \( t > T' \), since after \( T' \) steps it reaches a fixed point where the error term becomes significant in the recursion. Nevertheless, the above strategy is good enough to prove that, with probability \( 1 - o(n^{-1}) \), the number of blue vertices at level \( T' \) is sufficiently small. As for the subgraph of \( H \) from \( T' \) to 0, the only way that the root of \( H \) gets colour \( B \) is that the structure of \( H \) from level \( T \) to \( T' \) is particularly bad for \( R \). We
can prove that the event of $H$ having such a structure occurs with probability $o(n^{-1})$ by a process called the Duplicating process.

**Remark 2.** The random voting-DAG $H(v_0)$ can be viewed as the trajectory of a Coalescing and Branching random walk or, for short, COBRA walk (see [17],[37], [86] for recent research). A COBRA walk is a discrete process on a graph $G$ where vertices are occupied by particles. At each time-step, each particle makes $k-1$ copies of itself and they locate at the same vertex, then all the particles in the graph independently move to a random neighbour. After that, if a vertex is occupied by more than one particle they coalesce into one. The process keeps repeating forever. In our setting, $H$ represents the trajectory of $T$ steps of a COBRA walk with $k=3$ starting with one particle, located at $v_0$. Level $T-t$ of $H$ represents the set of occupied vertices at time $t$ of the COBRA walk, and the edges between level $T-t$ and $T-t-1$ represent the movements of the particles between times $t$ and $t+1$. The COBRA walk with parameter $k=1$ is the classic Coalescing random walk process which is the dual process of the voter model (or best-of-1 according to our notation).

The proof is presented in two parts. In Section 2.2.1 we work with the lower levels of the voting-DAG (i.e., close to the leaves) by the Sprinkling process; while in Section 2.2.2 we study the colouring structure close to the root by the Duplicating process.

### 2.2.1 Lower levels: Sprinkling process

Let $G = (V,E)$ be a graph of $n$ vertices with minimum degree $d = n^{\beta/\log \log n}$ where $\beta > 0$. For the simplicity and convenience of analysis, we associate the opinion $B$ to value 1 and $R$ to value 0.

Let $T' \leq T$, and consider the following protocol, which is called the Sprinkling process. Note that $T' = O(\log \log d) + O(\log(\delta^{-1})) + 1$ will be fixed in the forthcoming Lemma 4, and $T = O(\log \log d) + T'$ will be fixed in the forthcoming Lemma 7.
Suppose we only know the structure of the voting-DAG from level 0 up to level \( T' \), then we choose an arbitrary order of the vertices at level \( T' \), say \((v_1, T'), \ldots, (v_m, T')\) where \( m = |Q_{T'}| \). For each vertex at level \( T' \) from \( v_1 \) to \( v_m \), we start revealing the three sampled neighbours of them at level \( T' - 1 \) one by one. We say that a collision happens at \((v, T')\) if \((w, T' - 1)\) was revealed by \((v, T')\) and it was already revealed by another vertex before \( v \) in the order at level \( T' \) or by \( v \) itself. (See the graph on the left-hand side of Figure 2.4.) In such a case, we first erase \(( (v, T'), (w, T' - 1) ) \) from the edge set of \( H \). Then we add a new vertex to \( V(H) \) at level \( T' - 1 \), say \((q, T' - 1)\), and a new edge from \((v, T')\) to \((q, T' - 1)\). Next we set the outdegree of \((q, T' - 1)\) to be 0, and set the opinion of \((q, T' - 1)\) to be deterministically 1 (or \( B \) in colour language) irrespective of the actual colour of \((w, T' - 1)\). (See the graph on the right-hand side of Figure 2.4.)

Applying the Sprinkling model when revealing the neighbours of vertices at level \( T' \) gives us a collision-free level, where any two vertices at this level do not have common neighbours. After this we repeat the Sprinkling process on levels \( T' - 1, T' - 2, \ldots \) up to level 1. At the end we will have a new voting-DAG \( H' \) with \( V(H) \subseteq V(H') \), and all vertices in \( V(H') \setminus V(H) \) have opinion 1 (colour \( B \)) deterministically. Apart from the Sprinkling process, the rest of the colouring process of \( H' \) is the same as what we do in \( H \). We colour all normal vertices (which are not artificially added) at level 0 with colour \( B \) with probability \( 1/2 - \delta \), otherwise red. As those vertices without a collision also exist in \( V(H) \), a coupling that uses the same (random) initial colours in both \( H \) and \( H' \) gives us \( X_H(v, t) \leq X_{H'}(v, t) \) for all \((v, t) \in V(H)\), as a result of the extra blue vertices we added in the Sprinkling process. Denote by \( H' \) the result of the above process applied to the random voting-DAG \( \mathcal{H} \).

Since the Sprinkling process gives a collision-free subgraph from level 0 to \( T' \) in \( \mathcal{H}' \), it holds that \( \{ X_{\mathcal{H}'}(v, t), (v, t) \in V(\mathcal{H}) \} \) are independent random variables for \( t \in \{0, \ldots, T'\} \). In spite of the fact that their distribution is not identical (and is difficult to compute because of its dependency on several factors, such as the colours
Figure 2.4: Example of the construction of Sprinkling process. (Vertices are ordered from left to right, and collisions are marked in green.)

at level 0 and the (random) structure of $\mathcal{H}$ and thus $\mathcal{H}'$, we will prove that for any vertex $v$ at level $t \in \{0, \ldots, T'\}$,

$$P(X_{\mathcal{H}'}(v, t) = B | (v, t) \in V(\mathcal{H})) \leq p_t,$$

(2.2)

where $p_t$ satisfies the recursions: $p_0 = 1/2 - \delta$, and for $t \in \{1, \ldots, T'\}$

$$p_t = (3p_{t-1}^2 - 2p_{t-1}^3) (1 - \varepsilon_{t-1})^3 + (2p_{t-1} - p_{t-1}^2) \binom{3}{1} \varepsilon_{t-1}(1 - \varepsilon_{t-1})^2 + 1 \cdot \binom{3}{2} \varepsilon_{t-1}^2(1 - \varepsilon_{t-1}) + 1 \cdot \binom{3}{3} \varepsilon_{t-1}^3$$

$$\leq (3p_{t-1}^2 - 2p_{t-1}^3) + 6p_{t-1} \varepsilon_{t-1} + 3\varepsilon_{t-1}^2 + \varepsilon_{t-1}^3,$$

(2.3)

where $\varepsilon_{t-1} = 3^{T-t+1}/d$.

The proof goes by induction. Clearly the bound (2.2) applies for any vertex at level 0. Assume it works up to level $t-1$, and consider a vertex $(v, t) \in V(\mathcal{H})$ at level $t$. The event that $(v, t)$ is coloured by $B$ in $\mathcal{H}'$ is the same as: vertex $(v, t) \in V(\mathcal{H})$ has at least two neighbours that have opinion $B$ at level $t-1$; or there is exactly one collision in $(v, t)$ and at least one of its two normal neighbours are $B$; or there are 2 or 3 collisions in $(v, t)$. Note that at level $t-1$ there are at most $3^{T-t+1}$ vertices, therefore when revealing one neighbour of $(v, t)$, the probability of a collision is at most $3^{T-t+1}/d(v) \leq 3^{T-t+1}/d = \varepsilon_{t-1}$. Then, the expression in Equation (2.3) is obtained by revealing the neighbours of the vertices at level $t$ independently of the order. The first term is the probability that no collision occurs and there are at
least two blue vertices out of three normal vertices; the second term represents one collision with at least one blue vertex out of two normal vertices; and the last two terms mean two and three collisions respectively. We summarise the above argument in the following proposition.

**Proposition 3.** Let \( G = (V, E) \) be a graph of \( n \) vertices. Let \( v \in V \) be any vertex and consider \( \mathcal{H} \) the random voting-DAG associated to \( v \) of \( T \) levels. Let \( T' \leq T \), then the opinions at level \( T' \) can be majorised by a set of independent opinions where the probability of being \( B \) is given by \( p_{T'} \) as in Equation (2.3), where \( \varepsilon_{t-1} = 3^{T-t+1}/d \).

**Lemma 4.** Let \( G = (V, E) \) be a graph with minimum degree \( d \geq n^{\beta/\log \log n} \) for some \( \beta > 0 \), and assume the initial opinions of vertices are independently \( B \) with probability \( 1/2 - \delta \) with \( \delta \geq (\log d)^{-C} \) for some \( C > 0 \). Let \( v \in V \) be an arbitrary vertex. Then for any \( a > 0 \) there exists \( T \geq [a \log \log d] \), such that if considering the random voting-DAG \( \mathcal{H} \) of \( T \) levels, then the opinions at level \( T - [a \log \log d] \) are majorised by a vector of independent opinions where opinion \( B \) has probability \( o(d^{-1}) \).

**Proof.** We consider a voting-DAG of height \( T = [a \log \log d] + 1 + T_2 + T_3 \), where \( T_2 \) and \( T_3 \) will be chosen later. Recall that at level 0 all vertices have independent opinions with probability \( 1/2 - \delta \) of being \( B \), and \( R \) otherwise. Our proof consists of three steps:

1. **(i)** Opinions at level \( T_3 \) can be majorised by i.i.d. opinions with probability of being \( B \) equal to \( 1/2 - 1/(2\sqrt{3}) \);
2. **(ii)** Opinions at level \( T_2 + T_3 \) can be majorised by i.i.d. opinions with probability of being \( B \) equal to \( \text{polylog}(d)/d \);
3. **(iii)** Opinions at level \( 1 + T_2 + T_3 \) can be majorised by i.i.d. opinions with probability of being \( B \) equal to \( o(d^{-1}) \).

Note that ‘polylog’ denotes ‘polynomial-logarithmic’, so ‘polylog(n)’ means ‘some polynomial in \( \log(n) \)’, i.e. \( \log^k(n) \) for some constant \( k \).
CHAPTER 2. BEST-OF-THREE VOTING ON DENSE GRAPHS

We first check (iii) assuming (i) and (ii). For that, we ignore all previous levels and consider a voting-DAG of height \( h_1 = \lfloor a \log \log d \rfloor + 1 \) and the colour of the leaves are independently \( \text{B} \) with probability \( p_0 = \text{polylog}(d)/d \). By Equation (2.3),

\[
p_1 \leq 3p_0^2 + 6p_0 \varepsilon_0 + 3\varepsilon_0^2 + \varepsilon_0^3,
\]

and \( \varepsilon_0 = 3^{h_1/d} = \text{polylog}(d)/d \), hence

\[
p_1 = (\text{polylog}(d)/d)^2 = o(1/d).
\]

Next, we check (ii) assuming (i). Consider a voting-DAG of height \( h_2 = h_1 + T_2 \), and assume that leaves are independently \( \text{B} \) with probability \( p_0 = 1/2 - 1/(2\sqrt{3}) \). We choose \( T_2 \) to be

\[
T_2 = \min\{\min\{t \geq 0 : p_t \leq 12\varepsilon_t\}, 2 \log_2 \log d\}.
\]

For a voting-DAG of height \( h_2 \), we have that \( \varepsilon_t \leq 3^{h_2/d} = 3^{O(\log \log d)/d} = \text{polylog}(d)/d \), for any \( t \leq h_2 \). Then, for \( t \in \{1, \ldots, T_2\} \) we have that

\[
p_t \leq 3p_{t-1}^2 + 6p_{t-1}\varepsilon_{t-1} + 4\varepsilon_{t-1}^2 \leq 4p_{t-1}^2.
\] (2.4)

The first inequality of (2.4) is due to Equation (2.3), while the second follows from the fact that for \( t \leq T_2 \) we have \( p_{t-1} > 12\varepsilon_{t-1} \). Iterating the recursion, we have that

\[
p_t \leq (4p_0)^{2^t} \leq \left(4 \left(\frac{1}{2} - \frac{1}{2\sqrt{3}}\right)\right)^{2^t} = \left(1 - \frac{2}{\sqrt{3}}\right)^{2^t}.
\]

Let \( L = \log_2 \left(- \log(d)/\log(2 - 2/\sqrt{3})\right) \), and note that \( L \leq 2 \log_2 \log d \) for sufficiently large \( d \). If \( L < T_2 \) then we would have that \( p_L \leq 1/d \) but \( \varepsilon_L \geq 3^{h_2 - T_2}/d = 3^{h_1/d} = \text{polylog}(d)/d \), which contradicts the fact that \( p_L > 12\varepsilon_L \). Therefore, we conclude that \( T_2 = \min\{t \geq 0 : p_t \leq 12\varepsilon_t\} = O(\log \log d) \), and that opinions at level \( T_2 \) can be majorised by independent opinions with the probability of being blue equal to \( 12\varepsilon_{T_2} = 12 \times 3^{h_1/d} = \text{polylog}(d)/d \).

Finally, we check (i). We consider a voting-DAG of height \( h_3 = h_2 + T_3 \), and assume the initial opinions are i.i.d with probability of being blue \( p_0 = 1/2 - \delta \). We choose \( T_3 = \min\{\min\{t \geq 0 : \delta_t \geq 1/(2\sqrt{3})\}, C \log(\delta^{-1})\} \), where \( C \) is a suitable
constant greater than 10/\log(5/4). Let \( \delta_t = 1/2 - p_t \). Then, replacing \( p_t = 1/2 - \delta_t \) in Equation (2.3) and noting that \( 3\varepsilon_{t-1}^2 + \varepsilon_{t-1}^3 \leq \varepsilon_{t-1} \) give us

\[
\delta_t \geq \delta_{t-1} + \left( \frac{1}{2} \delta_{t-1} - 2\delta_{t-1}^3 - 4\varepsilon_{t-1} \right) .
\] (2.5)

The function \( f(x) = x/2 - 2x^3 \) is such that \( f(0) = 0 \), and it is increasing from 0 to 1/(2\sqrt{3}) where it reaches a local maximum. Note that if \( \delta_{t-1} \geq 12\varepsilon_{t-1} \) and \( \delta_{t-1} < 1/(2\sqrt{3}) \), then Equation (2.5) yields

\[
\frac{1}{2} \delta_{t-1} - 2\delta_{t-1}^3 - 4\varepsilon_{t-1} \geq \delta_{t-1} \left( \frac{1}{2} - 2\delta_{t-1}^2 - \frac{\varepsilon_{t-1}}{\delta_{t-1}} \right) \\
\geq \delta_{t-1} \left( \frac{1}{2} - \frac{1}{6} - \frac{1}{12} \right) = \frac{\delta_{t-1}}{4} ,
\]

implying that \( \delta_t \geq 5\delta_{t-1}/4 \). Note that as long as we can apply the previous recursion we have an increasing sequence of \( \delta_t \) while \( \varepsilon_t \) is decreasing, therefore if \( \delta_0 \gg \varepsilon_0 \) then \( \delta_t \gg \varepsilon_t \) for all \( t \leq T_3 \). To check that \( \delta_0 \gg \varepsilon_0 \), recall that \( \delta_0 = \delta \geq (\log d)^{-C} \) for some constant \( C > 0 \), and \( \varepsilon_0 \leq 3^{h_0}/d = 3^{O(\log \log d + \log \delta^{-1})}/d = (\text{polylog}(d))/d \), then \( \varepsilon_0 \ll \delta_0 \). We conclude that \( \delta_t > 5\delta_{t-1}/4 \) for all \( t \leq T_3 \). Let \( L = \frac{\log(\delta^{-1}/(2\sqrt{3}))}{\log(5/4)} \).

Note that \( \delta_L \geq (5/4)^L \delta_0 \geq 1/(2\sqrt{3}) \), implying that \( L \geq \min\{t \geq 0 : \delta_t \geq 1/(2\sqrt{3})\} \), therefore by our choice of \( C \) in \( T_3 \) we conclude that \( T_3 = \min\{t \geq 0 : \delta_t \geq 1/(2\sqrt{3})\} = O(\log(\delta^{-1})) \). \( \Box \)

### 2.2.2 Upper levels: Duplicating process

From the results in the previous section, we know that the opinions of vertices at level \( T' \), where \( T' = O(\log \log d + O(\log(\delta^{-1})) + 1 \) (see Lemma 4), in \( H \) are majorised by i.i.d. Bernoulli random variables with probability of being 1 (or colour \( B \) or blue) equal to \( o(d^{-1}) \) (see Proposition 3). In this section, we will deal with the levels from \( T' \) to level \( T \) (root).

Since there is no need to care about lower levels (from level 0 to level \( T' \)), we assume that \( \mathcal{H} \) is a voting-DAG of \( h + 1 = T - T' \) levels with root \((v_0, h)\) and that the vertices at levels 0 are independently \( B \) with probability \( o(d^{-1}) \), otherwise \( R \). Our
strategy to deal with this case is to show that for most realisations of the random voting-DAG $\mathcal{H}$, the number of vertices at the bottom (leaves) with colour $B$ is too small for the root of $\mathcal{H}$ to have colour $B$. We start by supposing that $\mathcal{H}$ is (deterministically) a ternary tree.

**Lemma 5.** Suppose that $H$ is a ternary tree of $h + 1$ levels. If the number of leaves with opinion $B$ is less than $2^h$, then the root has opinion $R$.

**Proof.** The statement is equivalent to that if the root is $B$ then there are at least $2^h$ vertices with opinion $B$ at level 0. The result holds easily by noting at least two neighbours of the root have opinion $B$, each of which is also the root of a (sub)-ternary tree of $H$.

For the case that the voting-DAG is not a ternary tree, the next lemma establishes that we can find a colouring on a ternary tree that assigns the same colour to the root, and the number of $B$ leaves in the ternary tree depends on the number of levels that involve collisions in the DAG. This process is referred as the Duplicating process.

**Lemma 6.** Let $H$ be a fixed voting-DAG of $h + 1$ levels with root $v_0 \in V(G)$. Given a colouring $\xi$ of the vertices at level 0, there exists a colouring $\xi'$ of the leaves of a ternary tree $H'$ of $h + 1$ levels such that the colouring process in $H$ and in $H'$ give the same colour to the root. Moreover, the number of $B$ leaves in $\xi'$ is at most $2^C B_0$, where $B_0$ is the number of $B$ leaves in $\xi$ and $C$ is the number of levels of $H$ that involve at least one collision.

**Proof.** The proof follows by induction on the number of levels. If the number of levels is 1, then $H$ is a single vertex and the result holds trivially. Suppose the result holds for $h$ levels, we will prove it for $h + 1$ levels. Let $X_H$ be the colouring of $H$ given the colouring $\xi$ of the leaves. Consider the root at level $h$ and let $e_1$, $e_2$ and $e_3$ be its three outgoing edges. We consider two cases: $i$) at least two of these edges share the
same endpoint at level $h - 1$ (i.e., level $h$ contains collisions), or $ii$) the edges do not share endpoints at level $h - 1$ (i.e., level $h$ is collision-free).

In the first case $i$), the opinion of $(v_0, h)$ is determined by the colour of the shared endpoint, say $(v, h - 1)$. In this case, we consider a voting-DAG $H'$ of $h + 1$ levels. At level $h$ we have the root $(v_0, h)$. At level $h - 1$ we put two disjoint copies of $H[(v, h - 1)]$ (without sharing vertices), and one ternary tree of $h - 1$ levels. Then we connect $(v_0, h)$ with the root of those three sub-graphs. We colour the leaves of $H'$ as follows. In the copies of $H[(v, h - 1)]$ the colours of the vertices are given by the original opinions settling in $\xi$, while the leaves of the ternary tree are attached to colour $R$. Note that the colour of the root of $H'$ is the same as the root of $H$ since the colour of the root of $H'$ is determined by the colour of the root of $H[(v, h - 1)]$ (the colour of the root of the ternary tree is irrelevant). By the induction hypothesis, $H[(v, h - 1)]$ can be transformed into a tree with at most $B'_0 2^{C'}$ leaves with opinion $B$, where $B'_0$ and $C'$ are the number of blue leaves and the number of levels involving at least one collision in $H[(v, h - 1)]$, respectively. By the construction of $H'$, all the collisions are represented in the copies of $H[(v, h - 1)]$ which are $C'$ many. Clearly $C' + 1 \leq C$ and $B'_0 \leq B_0$. Applying the induction hypothesis to the two copies of $H[(v, h - 1)]$, where $H[(v, h - 1)]$ is transformed in a ternary tree such that $B'_0 2^{C'}$ leaves have colour $B$. As $H'$ contains two copies of such graphs, after the induction step we get a ternary tree with $2 \cdot 2^{C'} B'_0 = 2^{C' + 1} B'_0 \leq 2^C B_0$ leaves with colour $B$.

Case $ii$) can be done similarly by applying the induction hypothesis to the three vertices at level $h - 1$.

Specifically, we give an example of how the Duplicating process works to obtain a ternary graph. As shown in Figure 2.5, suppose the edges from two vertices $e$ and $f$ respectively at level $i + 2$ share a same endpoint $b$ at level $i + 1$; and three edges from vertices $b$, $c$, and $d$ at level $i + 1$ respectively collide at a vertex $a$ at level $i$. We apply the Duplicating process by removing the edge between $f$ and $b$ and adding a copy of $H[(b, i + 1)]$ to vertex $f$ (see Figure 2.6, the root of the copy of $H[(b, i + 1)]$ is
named $b_1$; and then attach both vertex $c$ and $d$ a duplicate sub-graph $h[(a, i)]$ (see Figure 2.6, the roots of the copies of $H[(a, i)]$ are named $a_1$ and $a_2$ respectively).

Finally, we combine the two previous lemmas to show that with w.h.p. the root is $R$. The idea is to show that the number of levels involving collisions in the DAG is not large and therefore a straightforward application of Lemma 6 and Lemma 5, together with the fact that a leaf is $B$ with probability $o(d^{-1})$, tell us that the root of the DAG is $R$ w.h.p..

**Lemma 7.** Consider a random voting-DAG $\mathcal{H}$ with $h + 1$ levels, whose leaves have opinion $B$ with probability $o(d^{-1})$, otherwise $R$. Then, the probability that the root of $\mathcal{H}$ has opinion $B$ is $o(n^{-1})$.

*Proof.* Let $C_i$ be the indicator random variable taking value 1 if at least one clash occurs at level $i$. Recall that level $i$ involves a collision if two vertices at level $i$ share
a neighbour at level \(i - 1\). Consider the event \(E_i = \{\text{there are } m_i \text{ vertices at level } i\}\), where \(m_i \leq 3^{h-i}\) for \(i \in \{1, \ldots, h\}\), and start revealing the neighbours of the vertices at level \(i\) one by one. Then,

\[
P(C_i = 1 | E_i) = 1 - P(C_i = 0 | E_i) \\ \leq 1 - \left[ (1 - 1/d)(1 - 2/d) \cdots (1 - (m_i - 1)/d) \right] \\ \leq 1 - \left(1 - m_i/d\right)^{m_i} \leq 1 - \left(1 - m_i \cdot m_i/d\right) \\ = m_i^2/d \leq 9^h/d .
\]

Denote by \(C = \sum_{i=1}^{h} C_i\) the total number of levels that involve at least one collision. Since \(C_i\) only depends on the out-edges of vertices at level \(i\) of \(H\), then \(C\) can be majorised by a Binomial random variable that follows \(Bin(h, 9^h/d)\).

We first construct a ternary tree \(H'\) by applying Lemma 6 to \(H\). Let \(B\) and \(B'\) be the number of leaves with opinion \(B\) in \(H\) and \(H'\), respectively. Then \(B' \leq 2^C B\), and by Lemma 5 it holds that

\[
P(\text{root of } H \text{ is blue}) \leq P(B' \geq 2^h) \leq P(2^C B \geq 2^h) = P(B \geq 2^{h-C}) ,
\]

and

\[
P( B \geq 2^{h-C} ) \leq P \left( B \geq 2^{h-C} \left| C \geq \frac{h}{2} \right. \right) P \left( C > \frac{h}{2} \right) + \left[ P \left( B \geq 2^{h-C} \left| C \leq \frac{h}{2} \right. \right) P \left( C \leq \frac{h}{2} \right) \right] \\ \leq P \left( C > \frac{h}{2} \right) + P \left( B \geq 2^{h/2} \right) . \tag{2.6}
\]

To compute the first probability of Inequality (2.6), recall that the total number of levels \(C\) involving collisions can be majorised by \(Bin(h, 9^h/d)\). This gives

\[
P(C > \frac{h}{2}) \leq \sum_{k=[h/2]}^{h} \binom{h}{k} \left( \frac{9^h}{d} \right)^k \leq \sum_{k=[h/2]}^{h} \binom{he}{k} \left( \frac{9^h}{d} \right)^k \\ \leq \sum_{k=[h/2]}^{h} \left( \frac{2e \cdot 9^h}{d} \right)^k \leq \left( \frac{2e \cdot 9^h}{d} \right)^{h/2} \sum_{k=0}^{\infty} \left( \frac{2e \cdot 9^h}{d} \right)^k \\ \leq 2 \left( \frac{2e \cdot 9^h}{d} \right)^{h/2} .
\]
The last step holds as we claim that \(2e^{9h/d} \leq 1/2\), in which case the geometric series is approximated by 2. To see this, let \(h = a \log \log d\) for some constant \(a > 0\), then for any \(0 < b < 1\),

\[
2e^{9h/d} = 2e^{\exp\{a \log \log (\log_2 d)\}/d} = 2e^{(\log_2 d)^a \log^9/d} \leq d^{-b},
\]

which proves the claim. From the previous equation and by \(d = n^\alpha\) we obtain

\[
P(C > \frac{h}{2}) \leq 2d^{-bh/2} = n^{-bh\alpha/2}.
\]

We finish by checking that \(bh\alpha/2 > 1\), in which case \(h = a \log (\alpha \log_2 n)\). Then we get

\[
\frac{bh\alpha}{2} = \frac{ab}{2} \log (\alpha \log_2 n) = \frac{ab}{2} (\alpha \log \alpha + \alpha \log \log_2 n).
\]

For \(\alpha > c/\log \log_2 n\), we can choose \(a\) sufficiently large such that the above quantity is greater than 1, so that \(P(C > h/2) = o(n^{-1})\).

As for the other term of the Inequality (2.6), we know that the probability of a vertex at the bottom level in \(\mathcal{H}\) is \((3^{h+1}/d)^2 = o(1/d)\) from Lemma 4. Then, the number of blue leaves \(B\) in \(\mathcal{H}\) can be majorised by \(\text{Bin}(3^h, 9^{h+1}/d^2)\).

\[
P(B \geq 2^{h/2}) \leq \sum_{k=\lfloor h/2 \rfloor}^{3^h} \binom{3^h}{k} \left(\frac{9^{h+1}/d^2}{k} \right) \leq \sum_{k=\lfloor h/2 \rfloor}^{3^h} \binom{3^h}{k} \left(\frac{3^h}{d} \right)^k
\]

\[
\leq \sum_{k=\lfloor h/2 \rfloor}^{3^h} \left(\frac{3^h e}{\lfloor h/2 \rfloor} \right)^k \left(\frac{3^h}{d} \right)^k \leq \left(\frac{2e^{9h}}{dh} \right)^{\lfloor h/2 \rfloor} \sum_{k=0}^{\infty} \left(\frac{2e^{9h}}{dh} \right)^k
\]

\[
\leq 2 \left(\frac{2e^{9h}}{dh} \right)^{h/2}.
\]

The last step holds as long as \(2e^{9h}/(dh) = o(1)\), which was shown before. Note that we already demonstrated that \((2e^{9h}/d)^{h/2} = o(n^{-1})\), then we conclude that \(P(B > 2^{h/2}) = o(n^{-1})\).

\[\square\]

### 2.3 Open problems for further research

We have demonstrated that the consensus time of any graph of \(n\) vertices with a minimum degree \(d = n^{O(1/\log_2 \log_2 n)}\) under the Best-of-three protocol is \(O(\log_2 \log_2 n) +\)
\[ O(\log_2(\delta^{-1})) \] with high probability for constant \( \delta \in (0, \frac{1}{2}) \).

As for the further research, we are interested in extending our analysis of the Best-of-three protocol to cover more general classes of graphs. Since our arguments do not require any information of the structure of the graph, we may especially consider some special cases, such as grids, hypercubes, random graphs, and expanders, in which the structures might help us to drive down the probability of the occurrence of collisions or give us another way to deal with them. Generally speaking, further research may concern the following classes of graphs:

- Expanders with large degree
- Expanders with constant degree
- Hypercubes

We conjecture that our results will still be valid for the consensus time of expanders with large degree. It is harder to analyse the protocol for expanders of constant degree. As a starting point, we could consider the real process of random samplings of vertices, and examine how different types of collisions influence the probability of a vertex being blue at each level.

Apart from this, it is also important to figure out what is the worst case for the root to be in red at the end of the process. We also want to move beyond expanders to hypercubes, in which case we need to focus on the structure of the graphs. For instance, consider two vertices of a hypercube which are distant from each other: their neighbours may not overlap or the size of the intersection of their neighbours could be small, which implies that they these two vertices have low chance to collide. So, it might help us deal with the dependency by considering the such structure of hypercube.
Chapter 3

Diversity, Fairness, and Sustainability in Population Protocols

3.1 Introduction

Diversification, i.e., the action of diversifying something, is omnipresent: companies balancing the ratio of male to female employees, farmers sowing a large variety of crops, investment funds diversifying their portfolios to be more resilient, bees reallocating tasks to guarantee the survival of the hive, etc.

In order to understand how diversification works on a global scale through local choices, we study the problems from a population protocols perspective: through simple local choices. Consider a population of \( n \) agents, where each agent has one of \( k \) colours and each colour has an associated weight greater than or equal to 1. The main question we ask ourselves is: assuming limited interactions between the agents, does there exist a population protocol that converges to a state where each colour has a support which is roughly proportional to its weight?

Before answering such a question, it is relevant to explain why this problem is
not trivial. At first we may wonder why do we not just trivially choose a random colour with probability proportional to its weight. There are a couple of reasons for not doing this. First, in order to carry out such simple protocol, each agent needs to know the normalisation constant of the distribution, and thus we need to know all the different colours in the system and their respective weights, which requires too much memory and computation. Second, if one colour is removed from the system (e.g. an external agent recolours all red agents blue, which could arise when a task is fulfilled and no longer necessary), other agents will not be notified of this change. The same happens if nature changes the colour of an agent by a completely new one (e.g. an ant notices that the nest temperature is too hot and starts fanning). Therefore, the trivial protocol is not robust to changes in the structure of the population, and therefore it is not good to solve the diversification task. In order to make this protocol more robust, agents would need to constantly broadcast their list of colours, implying a lot of time, memory, local computation, and large messages, likely exceeding the capabilities of simple organisms such as ants.

In this chapter, we propose a more biologically plausible protocol, the Diversification protocol, where each agent only requires one extra bit of memory (besides remembering its own colour and the corresponding weight), and agents only need to communicate their own colour and weight. We will show that the Diversification protocol achieves diversity, and we will also show that our protocol is fair and sustainable. By fairness, we mean that over a long enough period, each agent $u$ has each colour $i \in [k]$ a portion of time which is roughly proportional to the weight $w_i \geq 1, w_i \in \mathbb{R}$ associated to the colour. For example, in the context of task-allocation, this means that all agents perform every task for roughly the same amount of time. Sustainability refers to the process guaranteeing (with probability 1) that no colour ever vanishes. Our protocol is also robust to structural changes: even when an adversary adds agents and colours, the protocol quickly returns into a state of diversity and fairness. Note that none of these properties implies one another.
CHAPTER 3. DIVERSITY, FAIRNESS, AND SUSTAINABILITY IN POPULATION PROTOCOLS

Our Diversification process is not only simple, but also efficient in converging, as it takes \( O(w^2 n \log n) \) time-steps to converge, where \( w = \sum_{i=1}^{k} w_i \) is the sum of the \( k \) weights of the different colours in the system. Assuming that the total weight \( w \) does not depend on the size of the population, the protocol is even asymptotically optimal, as for example, if a colour is supported by exactly one agent out of \( n \), then such colour has to propagate to at least a set of agents of size \( \Theta(n) \), which takes at least time \( \Omega(n \log n) \) in a population protocol by simple broadcasting arguments.

To see this, we consider a broadcasting process on a complete graph of size \( n \). Initially, one node knows a piece of information and we would like to know how long it takes to inform all the other \( n - 1 \) nodes. Suppose there is a Poisson clock that rings at rate 1 on each node. When the Poisson clock of a node \( u \) rings, \( u \) chooses a random neighbour to inform. Therefore, we obtain the following probability

\[
P(u \text{ informs } v \text{ during time interval } [t, t + dt]) = \frac{1}{n - 1} dt.
\]

This is because: the Poisson clock on node \( u \) rings during time interval \([t, t + dt)\) with probability \( 1 \cdot dt \); and node \( u \) picks node \( v \) to inform with probability \( 1/(n - 1) \). Moreover, if we consider the Poisson clocks on edges rather than on nodes, the above event is equivalent to \( n(n - 1) \) many Poisson processes on the edges with rate \( 1/(n - 1) \); and the Poisson process associated with edge \((u, v)\) happens during time interval \([t, t + dt)\) with probability \( dt/(n - 1) \). Let \( T_i \) be the first time that exactly \( i \) many nodes are informed, and \( T_1 = 0 \). Then, \( T_{i+1} - T_i \) is the time required for the \((i + 1)\)-th node to be informed. Let \( S_{T_i} \) be the set of nodes that are informed at \( T_i \), then \( |S_{T_i}| = i \). Note that there are \( i(n - i) \) many edges between the set \( S_{T_i} \) and its complement \( S_{T_i}^c \), each of which has a Poisson clock with rate \( 1/(n - 1) \). Thus, \( T_{i+1} - T_i \) follows a Poisson distribution with rate \( i(n - i)/(n - 1) \). Therefore,

\[
\mathbb{E}[T_{i+1} - T_i] = \frac{n - 1}{i(n - i)}
\]

\[
\Rightarrow \mathbb{E}[T_n] = \mathbb{E} \left[ \sum_{i=1}^{n-1} (T_{i+1} - T_i) \right] = \sum_{i=1}^{n-1} \frac{n - 1}{i(n - i)} = \frac{2(n - 1)}{n} \sum_{i=1}^{n-1} \frac{1}{i} \sim 2 \log n.
\]
CHAPTER 3. DIVERSITY, FAIRNESS, AND SUSTAINABILITY IN POPULATION PROTOCOLS

Note that, with the Poisson clocks setting, there are $n$ agents in expectation that are active in each unit time (by ‘active’ we mean ‘chosen to inform others’); while in the discrete setting, where one agent is chosen to inform others at each step if it has the information, there are at most 1 active agent in expectation in each unit time. Thus, in the discrete case, the process is $n$ times slower in expectation. Hence, for a colour supported by exactly one agent at the beginning to propagate to at least $\Theta(n)$ agents asynchronously, it takes at least $O(n \log n)$ time.

Interestingly, the questions of diversity, fairness and sustainability have not been asked in the context of population protocols, yet$^1$. Instead, research has focused almost uniquely on understanding the opposite question: how can the population agree on one colour?

The closest works to ours are [106, 104, 105] by Yasumi et al., where they study population protocols designed to obtain equi-sized partitions. In these works the authors focused on deterministic scheduling (worst-case), and investigated several space-complexity problems under several assumptions on, e.g., initial states, power of the (deterministic) adversarial scheduling, etc, where the main objective is to find protocols that use the least number of states. Our approach is rather different as we assume a random schedule, where no adversary is trying to boycott our protocol by slowing down interactions between agents, and thus we focus on finding a fast and simple protocol.

Diversity can be seen as an opposite problem of consensus, which is the most studied problem in the field of population protocols. Here, $n$ agents start with one of $k$ distinct colours and the goal is to converge quickly to a single colour. The required time is called the consensus time. This finds applications in understanding spreading phenomena, for example the spread of infectious diseases, rumours and opinions in societies, or viruses/worms in computer network. Some well studied

$^1$There has been some work [106, 104, 105] by Yasumi et al. on finding equi-sized partitions in deterministic population protocol settings.
consensus protocols include 2-Choices and 3-Majority. In the former, every agent \( u \) samples two random agents in each time-step and updates its opinion if both sampled agents have the same opinion. In the latter, each vertex \( u \) samples three random agents and revises its opinion according to the majority opinion of the three samples. See [34, 14, 35, 16, 38, 85, 41, 72] for references of the state-of-the-art. While 2-Choices and 3-Majority are not exactly a population protocol in the sense that interactions are not pairwise, those protocols have been extended to actual population protocols, with the help of extra memory per agent [14]. The state-of-the-art for population protocols reaching consensus is [18] (with elaborate communication and timing strategies). Other related spreading/consensus processes include the Moran process [78, 45], contact processes, and other classic epidemic processes [20, 79, 88].

Another protocol closely related to ours is the anti-voter model, where each agent starts with one of two opposite colours. In this protocol agents interact in the same way as the Voter model, but agents adopt the opposite colour of the sampled neighbour. It is proved that in the long term agents reach an equilibrium and that agents are constantly changing colours [4, 94]. However, it is worth noting that this protocol works only for two colours and that agents know the colours, thus it is not straightforward to extend this protocol to more colours, without the use of extra memory by the agents.

Another related type of protocols are averaging processes [5], in which upon a pairwise interaction, both agents adopt the average of their opinions. In [90], the authors considered various averaging processes, in the context of load-balancing. Their most related process is the so-called diffusion load-balancing process in the matching model, in which each node starts with some load and in every round the nodes are randomly matched and average their load. The authors obtained bounds on the convergence time and the load discrepancy over time. Recently, authors of [84] considered the noisy-averaging population protocol. Here, each agent has an initial value and the goal is to agree on the average. The catch is that the communication is noisy.
and the value communicated by the agents can be altered. The authors studied the
behaviour of the simple averaging protocol in which two agents are chosen uniformly
at random and both set their values to the average of the received value and their
own value.

Population protocols have also been used in the context of community detection. Most of the research has focused on the stochastic block model where each agent belongs to one of two communities. Any two agents from the same community share an edge with probability $p$ and two agents from different communities share an edge with probability $q$, where $q < p$. The goal is to recover the hidden partition. See [85, 12, 42] for more details.

There has also been a large body of research on task allocation. Note that our problem is a special case of task allocation where each task has a demand proportional to $w_i$. See [39, 47, 91] for ant-inspired research on task allocation. The main difference from our work is that this line of research assumes that ants receive feedback from the tasks indicating whether there are too many or too few ants working on the tasks. In our setting, there is no such feedback; agents receive implicit feedback through sampling the current colour of other agents. Becchetti et al. [13] studied a synchronous/parallel process of the balls-into-bins problem, in which a fixed number of $n$ balls are assigned to $n$ bins. In every round, one ball is chosen from each non-empty bin and reallocated to one of the $n$ bins uniformly at random. They proved that starting from an arbitrary configuration, this process converges to a configuration with maximum load $O(\log n)$ in linear time, and it stays in $O(\log n)$ for $\text{poly}(n)$ rounds, with high probability.

## 3.2 Protocol and results

We study the following population protocol. Initially, there are $n$ agents, each of which has one out of $k$ colours; and each colour $i \in [k]$ has an associated weight
The system evolves as follows: at every time-step an agent $u$ is chosen out of $n$ agents (scheduled) u.a.r.\footnote{We use ‘u.a.r.’ to denote ‘uniformly at random’}. The scheduled agent $u$ then samples another agent $v$ from its $n - 1$ neighbours u.a.r., and observes the colour of $v$ as well as its weight, and then processes such information, leading to a possible change of colour and the corresponding weight. We denote by $c_u(t)$ the colour of agent $u$ after the $t$-th time-step, and $c_u(0)$ its initial colour. Additionally, we denote by $C_i(t)$ the number of agents with colour $i$ at time-step $t$, i.e., $C_i(t) = |v \in [n] : c_v(t) = i|$.

### 3.2.1 Properties and definition of the protocol

We now introduce the properties we will study in this Diversification protocol: diversity, fairness, and sustainability. The primary goal is to achieve diversity, and the two secondary goals are fairness and sustainability. We assume that both the total weight $w$ and the number of colours $k$ are independent of the size of the population, and both are constant.

(i) **Diversity.** A protocol is diverse if there exists $t_0 \geq 0$ and $T = \Omega(n^\alpha)$, where $\alpha > 0$ is a constant, such that for any $t \in \{t_0, \ldots, T\}$ and $i \in [k]$ w.h.p.\footnote{We use ‘w.h.p.’ to denote ‘with high probability’ meaning with probability at least $1 - 1/n$.}

\[
\left| \frac{C_i(t)}{n} - \frac{w_i}{w} \right| = \tilde{O} \left( \frac{1}{\sqrt{n}} \right).
\] (3.1)

Note that, $\tilde{O}(f(n))$ is the soft $O$ notation, meaning $O(f(n) \cdot \text{polylog}(f(n)))$, i.e. $O(f(n) \cdot (\log f(n))^s)$ for some constant $s$.

In words, the population stabilises in a subset of configurations such that each colour $i$ appears in the population proportionally to its weight $w_i$ for $\Omega(n^\alpha)$ time-steps. Note that if $w_i = 1$ for all colour $i$, then achieving diversity is equivalent to finding a uniform partition of the vertices.

(ii) **Fairness.** A protocol is fair if there exists $T = \Omega(n^\beta)$, where $\beta > 0$ is a constant, such that for any $T' > T$, it holds for any $i \in [k]$ and any agent $u$
w.h.p. that
$$\frac{\left|\{t' \in [0, T']: c_u(t') = i\}\right|}{T'} = (1 \pm o(1)) \frac{w_i}{w}.$$ 

In words, each agent has, roughly, colour $i$ for a proportion $w_i/w$ of the times.

(iii) **Sustainability.** A protocol is sustainable if for all $t > 0$ and any colour $i$, there exists at least one agent with colour $i$ at time $t$. In other words, no colour ever vanishes.

**Diversification Protocol.** We proceed to define the Diversification protocol which satisfies all the above properties. In this protocol each agent has an extra bit of memory, which can be observed by other agents when they interact. This bit represents the degree of confidence agents have in their current colour. Agents whose bit is 1 will not change their current colour, whereas agents whose bit is 0 are open to change their current colour. Pictorially, we call colours with bit 0 *light*, and those with bit 1 *dark*, both of which are also referred to as *shade*. Therefore, a dark colour needs to become light before changing into a completely different colour. Let $b_u(t)$ represents the extra bit of agent $u$ after the $t$-th iteration. We assume that $b_u(0) = 1$ for all agents $u$. The Diversification protocol is defined as follows: suppose that, in the $(t + 1)$-th time-step, agent $u$ is scheduled (chosen u.a.r out of all $n$ agents) and it samples another agent $v$ u.a.r.. If $u$ has a light colour, and $v$ has a dark colour, then $u$ adopts the colour of $v$ (and the associated weight of that colour) and its shade. If both $u$ and $v$ have the same dark colour $i$, then $u$ changes the shade of its colour to light (i.e. it keeps the current colour and changes the value of the extra bit to 0) with probability $1/w_i$, where $w_i \geq 1$ is the associated weight of colour $i$. Formally,
the changes occur according to the following rule:

\[(c_u(t+1), b_u(t+1)) = \begin{cases} 
(c_v(t), 1) & \text{if } b_u(t) = 0 \text{ and } b_v(t) = 1, \\
(c_u(t), 0) \text{ w.p. } \frac{1}{w_{c_u(t)}} & \text{if } b_u(t) = b_v(t) = 1, \text{ and } c_u(t) = c_v(t), \\
(c_u(t), b_u(t)) & \text{otherwise.}
\end{cases}\]  

Note that when all weights are equal to 1, a dark colour agent changes its shade to light with probability 1 if it interacts with another agent with the same dark colour. In this case, our protocol gives a deterministic protocol for the uniform partition problem.

The intuition of the protocol, from the agent point of view, is that if agent \(u\) observes another agent \(v\) with the same colour, it suggests that this colour is over-represented. Of course, it is very crude using only one observation to decide if one colour is over-represented. Nonetheless, in expectation this approach works: the colour distribution approaches the target distribution (e.g., uniform distribution if all weights are 1). Since there are \(n\) agents performing the same protocol over long periods, emergent behaviour appears: the colour distribution will be concentrated around its expected value, the target distribution.

To ensure that the protocol indeed converges towards the target distribution, two rules are important. The first one is that only light-shaded colours can change their colour (first line of Equation (3.2)). The second rule is that dark-shaded colours change their shade to light if they observe another same colour with dark shade, with probability equals to the inverse of the weight of the colour, otherwise they keep the shade (second line of Equation (3.2)). This causes heavy-weighted colours to change less often than light-weighted ones, and thus we expect to see more of the former than the latter.

A rough idea of why this protocol works is that for each colour \(i\), the ‘rate’ at which colour \(i\) decreases by 1 is \(C_i(t)^2/(n^2w_i)\), whereas the ‘rate’ \(i\) increases by 1 is
CHAPTER 3. DIVERSITY, FAIRNESS, AND SUSTAINABILITY IN POPULATION PROTOCOLS

$C_i(t) \frac{\sum_{j=1}^{k} C_j(t)}{(n^2 w)}$, hence equilibrium is achieved when all values $C_i(t)/w_i$ for $i \in [k]$ are roughly the same, i.e. $C_i \approx nw_i/w$, leading to diversity. Fairness is obtained by noting that an agent changes its colour to colour $j$ with probability proportional to $C_j(t)$. Furthermore, since the protocol eventually reaches a configuration with $C_j(t) \approx nw_j/w$, then for large $t$, an agent has colour $j$ approximately a $w_j/w$ fraction of the time. Sustainability simply follows from the fact that an agent with a dark shade of a colour can change its colour only after it becomes the light-shaded version, which happens only when the agent encounters another agent with the same dark-shaded colour. Since only one agent changes its colour at a given time-step, no dark-shaded colour can ever vanish. Moreover, sustainability is preserved if new agents are added to the system, or new colours are added, as long as the new colours are initially dark, and they do not replace the last dark version of another colour.

In the next section we will formalise these ideas, and show that the Diversification protocol satisfies the diversity and fairness properties.

**Derandomisation.** Before finishing this section, we describe a derandomised version of our protocol that avoids the sampling procedure in the second line of Equation (3.2). Here we shall assume that all weights are non-negative integers. The derandomised Diversification protocol is described as follows: instead of having only light and dark shades, we have $1 + w_i$ different shades of grey which are enumerated from 0 (light) to $w_i$ (dark). At each time-step, an agent is selected u.a.r. (scheduled) from the $n$ agents, and then the scheduled agent samples another agent from its $n-1$ neighbours. For an agent $u$ which has colour $i$ with shade greater than 0, whenever it is scheduled, if it chooses another agent with the same colour and shade greater than 0, agent $u$ reduces the shade of its colour by 1. If an agent $u$ with colour $i$ and shade 0 chooses an agent $v$ with shade greater than 0, then $u$ adopts the colour of $v$, say $j$, and sets its shade to $w_j$. In all other interactions, the agents do nothing. Note this algorithm requires $\lceil \log_2(1 + w_i) \rceil$ extra bits of memory when it adopts colour $i$. 

56
3.2.2 Main results

For simplicity, in this work, we assume that the number of colours $k$ and the total sum of weights $w$ are constants. However, we state most of the intermediate results in terms of $k$ and $w$, but we do not attempt to optimise the terms involving $k$ or $w$.

**Theorem 8.** For constant $k$ and $w$, the Diversification protocol achieves diversity, fairness and sustainability.

To prove the above theorem we just need to prove that the protocol is diverse and fair as we already argued that the protocol achieves sustainability: a dark colour cannot change its colour in one step. In order to change colour, it has to change its shade from dark to light first, which requires interacting with another agent with the same dark colour. Therefore, if initially all colours are dark, it holds for each colour $i$ in the system that at each time-step there exists at least one agent with the dark version of colour $i$, thus leading to sustainability.

For diversity, we rely heavily on the following theorem, which directly gives the desired result.

**Theorem 9.** There exists $T = O(w^2n \log n)$, such that, w.h.p., for every colour $i \in [k]$,

$$\sum_{i=1}^{k} \sum_{j=1}^{k} \left| \frac{C_i(t)}{w_i} - \frac{C_j(t)}{w_j} \right|^2 = O(wn \log n) ,$$

for all $t$ in the interval $[T, n^8]$.

Dividing by $2k^2$ on both sides of the above equation, expanding the square term on the left-hand side, and rearranging some terms yield

$$\frac{1}{k} \sum_{i=1}^{k} \left( \frac{C_i(t)}{w_i} \right)^2 - \frac{1}{k^2} \left( \sum_{i=1}^{k} \frac{C_i(t)}{w_i} \right)^2 = O(wn \log n/k^2) .$$

Let $x = (1/k) \sum_{i=1}^{k} C_i(t)/w_i$, then

$$\frac{1}{k} \sum_{i=1}^{k} \left( \frac{C_i(t)}{w_i} - x \right)^2 = O(wn \log n/k^2) . \quad (3.3)$$
From here we will deduce that
\[
\frac{C_i(t)}{w_i} = \frac{n}{w} + O\left(\sqrt{(w/k)n \log n}\right) \tag{3.4}
\]
holds after \(O(n \log n)\) time-steps for at least \(\Omega(n^8)\) time-steps with high probability.

To derive Equation (3.4) we notice that Equation (3.3) gives
\[
C_i(t) = w_i x + w_i \cdot O\left(\sqrt{(w/k)n \log n}\right).
\]
Hence, by taking sums of \(C_i\) and \(w_i\) and noting \(\sum_{i=1}^{k} C_i(t) = n\), we get
\[
n = w x + w \cdot O\left(\sqrt{(w/k)n \log n}\right),
\]
from which we know that \(x = n/w + O(\sqrt{(w/k)n \log n})\), and thus Equation (3.4) holds.

The previous result given in Equation (3.4) implies that the Diversification protocol is diverse. While we proved that \(\alpha\) in the definition of diversity in Equation (3.1) can be chosen as \(\alpha = 8\), a simple inspection of our proof shows that \(\alpha\) can be chosen as an arbitrarily large constant. The proof of Theorem 9 can be found in Section 3.3.

The proof of fairness is provided in Section 3.3.3, which is built on the results proved in Section 3.3. The main idea is that after \(O(n \log n)\) time-steps the whole system stabilises for at least \(\Omega(n^8)\) time-steps in configurations where colours (including both light and dark versions) are almost perfectly distributed among the agents (according to the appropriate weight values). Therefore, instead of following the colours of agents in the Diversification protocol, we follow agents in a new system where the (dark and light) colours are perfectly distributed among the agents.

The main contribution of this paper is to propose and demonstrate that the Diversification protocol achieves diversity, fairness, and sustainability. The analysis of the protocol is mainly probabilistic, and it is divided into three phases. In the first phase the main technique is to couple certain statistics of our protocol with biased random walks on integers in order to achieve rapid convergence (despite the fact that
this is the slowest part of the process as the initial configurations of colours are arbitrary). In the second phase we introduce two potential functions that decrease over time, and we provide a general concentration inequality to analyse those potentials. Finally, in the third and final phase we couple the trajectory of the states of agents, that is \((c_u(t), b_u(t))_{t \geq 0}\), with a Markov chain \(P\) that represents the system in ‘perfect equilibrium’, and show that both processes roughly hit every colour the same amount of times. To prove our coupling is correct we make use of Chernoff’s bounds for Markov chains.

### 3.3 Analysis of the Diversification protocol

We first introduce some notation that will be used throughout the analysis of our protocol. For simplicity, we assume that the colours are enumerated from 1 to \(k\) with corresponding weights \(w_1, \ldots, w_k\), where \(w_i \geq 1\) for \(i \in [k]\), and we set \(w = \sum_{i=1}^{k} w_i\). Recall that \(c_u(t)\) denotes the colour of agent \(u\) after the \(t\)-th iterations, and \(b_u(t)\) its corresponding shade indicator after the \(t\)-th iterations: 1 for dark shade and 0 for light shade. We denote by \(A_i(t)\) and \(a_i(t)\) the number of agents having colour \(i\) at time \(t\) with dark shade (bit value 1) and light shade (bit value 0), respectively, i.e., for \(t > 0\),

\[
A_i(t) = |\{u : c_u(t) = i, b_u(t) = 1\}|, \quad \text{and} \\
a_i(t) = |\{u : c_u(t) = i, b_u(t) = 0\}|.
\]

Define \(A(t) = \sum_{i=1}^{k} A_i(t)\) and \(a(t) = \sum_{i=1}^{k} a_i(t)\), and let \(\xi(t)\) denote the process containing all the information at time \(t\), that is,

\[
\xi(t) = (A_1(t), \ldots, A_k(t), a_1(t), \ldots, a_k(t)) .
\]

Let \(\Omega = \{(A_1, \ldots, A_k, a_1, \ldots, a_k) : A_i \geq 1, a_i \geq 0, \sum_{i=1}^{k} (A_i + a_i) = n\}\) be the space state where the process \(\xi(t)\) takes values. Clearly the initial state is such that \(A_i(0) \geq 1\) for all colours \(i\).
As usually done in population protocol, our analysis is divided into several phases. In each of these phases the configuration of the system (i.e., the colours of the agents) will be attaining properties that are maintained for long periods of time (much longer than the duration of those phases), and each successive phase will make use of the previously achieved properties. The main idea revolves around showing that perfect equilibrium is achieved at the values:

$$\frac{a_i(t)}{n} = \frac{w_i}{1 + w} \quad \text{and} \quad \frac{A_i(t)}{n} = \frac{w_i}{1 + w},$$

for all \(i \in \{1, \ldots, k\}\). The analysis of our protocol requires three phases which we proceed to describe.

**Phase 1** starts at time 0 and ends at time \(\tau_1 = O(w^2 n \log n)\). In this phase we will show that w.h.p. the equality in Equation (3.5) can be obtained up to multiplicative constants (see Theorem 14). This property is shown to hold for at least \(\Omega(n^{10})\) time-steps, enough to be carried to the next phases of the analysis. The analysis of this phase is in Section 3.3.1.

**Phase 2** starts exactly at the end of Phase 1, and lasts for \(\tau_2 = O(wn \log n)\) time-steps. This phase is divided into two consecutive subphases. The first subphase lasts for \(\tau_{2,1} = O(wn \log n)\) time-steps, while the second subphase lasts for \(\tau_{2,2} = O(wn \log n)\) time-steps. Clearly \(\tau_2 = \tau_{2,1} + \tau_{2,2}\).

In this phase we will analyse two potential functions \(\phi\) and \(\psi\), given by

$$\phi(t) = \sum_{i=1}^{k} \sum_{j=1}^{k} \left( \frac{A_i(t)}{w_i} - \frac{A_j(t)}{w_j} \right)^2, \quad \text{and} \quad \psi(t) = \sum_{i=1}^{k} \sum_{j=1}^{k} \left( \frac{a_i(t)}{w_i} - \frac{a_j(t)}{w_j} \right)^2.$$

Then, in the first subphase we show that the potential \(\phi(t)\) decreases to \(O(wn \log n)\) w.h.p. (by Lemma 15), and in the second subphase, \(\psi(t)\) is shown to decrease to \(O(wn \log n)\) w.h.p. (by Lemma 16). These properties hold for at least \(n^8\) time-steps (see Theorem 17). The analysis of Phase 2 is in Section 3.3.2.

The analysis of the first two phases is enough to prove diversity, however, to prove fairness we need an extra phase.
Phase 3 starts exactly after the end of Phase 2, and lasts for $\tau_3 = O(wn \log n)$ time-steps. In this phase we show that the system is very close to perfect equilibrium, indeed, Equation (3.5) holds up to a small additive error, and such property holds up to time $n^8$ w.h.p. (see Theorem 22). To get such approximation, we will show that the potential function $\sigma^2$ given by

$$\sigma^2(t) = \left( \frac{A(t)}{w} - a(t) \right)^2$$

decreases to $O(wn \log n)$, which combining with the fact that the potential functions $\phi(t)$ and $\psi(t)$ are already of the same size, gives us enough information to obtain the desired approximation. Phase 3 is analysed in Section 3.3.3.

For quick reference, Figure 3.1 below shows a brief summary of the three phases.

![Figure 3.1: Summary of the three phases of the analysis.](image)

### 3.3.1 Phase 1: The rise of the minorities

In the first phase of the analysis, we handle colours that are largely over-represented or largely under-represented, and we will show that the system stabilises in a configuration such that for any $i \in [k]$,

$$\frac{a_i(t)}{n} \approx \frac{w_i}{1 + w} \quad \text{and} \quad \frac{A_i(t)}{n} \approx \frac{w_i}{1 + w},$$

(3.6)

where the approximation holds up to a multiplicative constant. For this, we will demonstrate that the process, little by little, starts improving its configuration, in
the sense that we are closer to achieve Equation (3.6) over time. At the beginning, the process is rather slow at making progress as some colours may have only a small number of agents supporting them, but as long as the configuration improves, the speed at which the process improves the configuration increases as well.

To discuss this in detail, we first show that $a(t) = \sum_{i=1}^{k} a_i(t)$ quickly increases to roughly $(1 - \varepsilon)n/(1 + w)$ from $a(0) = 0$. We achieve this by coupling the process $a(t)$ with a biased random walk. To see why $a(t)$ has bias towards increasing its value, note that, at the beginning, all agents have dark colours, and therefore interactions between agents of the same colour are likely to happen between agents of the same shade, thus enlarging the chance of increasing the value of $a(t)$. After obtaining a healthy proportion of light-coloured agents (which are available to change their current colour to any other colour), we can show that under-represented colours start growing by another biased random walk argument. At first, we show that under-represented dark-shaded colours, say $A_i$, have bias toward increasing their representation. Intuitively, at the beginning, it is unlikely for an agent to sample an under-represented colour $i$ and therefore it is unlikely that an agent switches to colour $i$, however, it is even less likely that an agent of colour $i$ switches to a different colour, so at the beginning, $A_i(t)$ grows very slowly. Then, as the popularity of colour $i$ grows, it becomes more likely to sample $i$, thus $A_i(t)$ will start to grow increasingly faster, until $A_i(t)/n$ is close to $w_i/(1 + w)$, at which point the bias of the process starts being less relevant to make significant progress. The same can be applied to each under-represented light-shaded colour, say $a_i(t)$, and also to over-represented colours with dark and light shades respectively. By the end of Phase 1 the bias presented in the system is very small, and thus coupling with biased random walks is not enough to prove improvements in the approximation of Equation (3.6).

For a formal analysis it is convenient to define the following regions in terms of a
constant parameter $\varepsilon \in (0, 1/4)$:

\[
R_1 = \left\{ \xi \in \Omega : \frac{a}{n} \geq (1 - \varepsilon) \frac{1}{1 + w} \right\},
\]

\[
S_1 = \left\{ \xi \in \Omega : \frac{a}{n} \geq (1 - 2\varepsilon) \frac{1}{1 + w} \right\},
\]

\[
R_2 = \left\{ \xi \in \Omega : \forall i \in [k], \frac{A_i}{n} \geq (1 - 3\varepsilon) \frac{w_i}{1 + w} \right\} \cap S_1,
\]

\[
S_2 = \left\{ \xi \in \Omega : \forall i \in [k], \frac{A_i}{n} \geq (1 - 4\varepsilon) \frac{w_i}{1 + w} \right\} \cap S_1,
\]

\[
S_3 = \left\{ \xi \in \Omega : \forall i \in [k], \frac{A_i}{n} \leq (1 + 4\varepsilon w) \frac{w_i}{1 + w} \right\} \cap S_2,
\]

\[
S_4 = \left\{ \xi \in \Omega : \frac{a}{n} \leq (1 + 4\varepsilon w) \frac{1}{1 + w} \right\} \cap S_3.
\]

Note that $R_j \subseteq S_j$. Define $T_i = \min\{t \geq T_{i-1} : \xi(t) \in R_i\}$ and $T_0 = 0$, and let $T_i' = \min\{t \geq T_i : \xi(t) \in \Omega \setminus S_i\}$.

**Lemma 10.** Let $c > 0$ be a sufficiently small constant. For any $\xi(0) \in \Omega$, we have $T_1 = O(nw/\varepsilon)$ with probability at least $1 - \exp(-c\varepsilon n)$. Moreover, if $\xi(0) \in R_1$ then $T_1' > \exp(cn\varepsilon^2/w)$ with probability at least $1 - \exp(-cn\varepsilon^2/w)$.

**Lemma 11.** Let $c > 0$ be a sufficiently small constant. For any $\xi(0) \in \Omega$, we have $T_2 = O(nw \log n/\varepsilon)$ with probability at least $1 - \exp(-cn\varepsilon^2/w)$. Additionally, if $\xi(0) \in R_2$ then $T_2' > \exp(cn\varepsilon^2/w)$ with probability at least $1 - \exp(-cn\varepsilon^2/w)$.

The proofs of Lemma 10 and Lemma 11 are provided in Section 3.5.1.

**Lemma 12.** Assume $\xi(t) \in S_2$, then $\xi(t) \in S_3$.

**Proof.** Using the lower bounds given in the definition of $S_2$ and $S_1$, we have

\[
A_i \leq n - a - \sum_{j \neq i} A_j \leq n - \frac{(1 - 2\varepsilon)n}{1 + w} - \frac{(1 - 4\varepsilon)n(w - w_i)}{1 + w} \leq \frac{nw_i}{1 + w} + \frac{4n\varepsilon w w_i}{1 + w}.
\]

**Lemma 13.** Assume $\xi(t) \in S_3$, then $\xi(t) \in S_4$. 

63
CHAPTER 3. DIVERSITY, FAIRNESS, AND SUSTAINABILITY IN POPULATION PROTOCOLS

Proof. Using the lower bounds given in the definition of $S_3$ and $S_2$, we have

$$a \leq n - \sum_{i=1}^{k} A_i \leq n - \frac{(1 - 4\varepsilon)nw}{1 + w} \leq \frac{n}{1 + w} + \frac{4\varepsilon nw}{1 + w}.$$ 

\[ \square \]

For any $\delta > 0$ define the set of configurations $\mathcal{E} = \mathcal{E}(\delta) \subseteq \Omega$ by

$$\mathcal{E}(\delta) = \left\{ \xi \in \Omega : \frac{A_i}{w_i} \in \left[ \frac{(1 - \delta)n}{1 + w}, \frac{(1 + \delta)n}{1 + w} \right] \forall i \in [k], \text{ and } a \in \left[ \frac{(1 - \delta)n}{1 + w}, \frac{(1 + \delta)n}{1 + w} \right] \right\}.$$ 

(3.7)

Applying the previous lemmas in order with $\varepsilon = \delta/(4w)$ we obtain the following result.

**Theorem 14.** Let $\delta > 0$ be fixed, then there exists $\tau_1 = O(w^2 n \log n)$ such that

$$\mathbb{P} \left( \bigcap_{t=\tau_1}^{\tau_1 + n^{10}} \{ \xi(t) \in \mathcal{E} \} \right) \geq 1 - \exp \left( -\Omega \left( \frac{n}{w^3} \right) \right).$$

Note that $\tau_1$ in the above theorem implicitly depends on the choice of $\delta$. For our purposes, we just fix $\delta$ as a small enough constant (for example $\delta = 0.0001$ is more than enough for our analysis). The second phase of the process starts exactly at time $\tau_1$ of Theorem 14, where we know that w.h.p. $A_i(t)/w_i$ and $a(t)$ are equal up to multiplicative constants, and this is valid for at least $n^{10}$ time-steps.

The proof of the above theorem can be found in Section 3.5.1.

### 3.3.2 Phase 2: Reaching equilibrium - proving diversity

The second phase starts when the proportions $a_i/w_i$ are all equal up to a multiplicative constant, as well as the proportions $A_i/w_i$. It is convenient to restart time back to 0 for the analysis of phase 2 of the process. Therefore we set the starting configuration of phase 2 as $\xi(0)$. We know that $\xi(0) \in \mathcal{E}$ and it has the property $\xi(t) \in \mathcal{E}$ for at least $n^{10}$ time-steps.
As it was mentioned at the beginning of Section 3.3, in the second phase we will analyse two potential functions \( \phi \) and \( \psi \), given by

\[
\phi(t) = \sum_{i=1}^{k} \sum_{j=1}^{k} \left( \frac{A_i(t)}{w_i} - \frac{A_j(t)}{w_j} \right)^2,
\]

(3.8)

and

\[
\psi(t) = \sum_{i=1}^{k} \sum_{j=1}^{k} \left( \frac{a_i(t)}{w_i} - \frac{a_j(t)}{w_j} \right)^2.
\]

(3.9)

Also, recall that Phase 2 is divided into two consecutive subphases - subphase 2.1 and subphase 2.2, lasting for \( \tau_{2,1} \) and \( \tau_{2,2} \) time-steps, respectively (see Figure 3.1). In the first subphase we prove that \( \phi(t) \) reduces its value to \( O(wn \log n) \), and in the second one we show the same for \( \psi(t) \).

It is worth mentioning that both potential functions implicitly depend on each other. Potential function \( \phi \) only depends on terms containing \( A_i(t) \)'s, however, there is an implicit dependency on light-shaded colours, as only light-shaded agents can transform into dark-shaded ones. The same holds for the potential function \( \psi \). Fortunately, the properties proved in Phase 1 of our analysis (that hold for long periods of time) allow us to control the implicit dependency between the potentials. In particular, we will verify that both \( \phi \) and \( \psi \) are approximately super-martingales, and both present a drift towards reducing their values. Indeed, we will show that their values halve every \( O(wn) \) time-steps. In order to obtain strong w.h.p. bounds for \( \phi \) and \( \psi \), we introduce a Chung-Lu-type [28] concentration to bound a general class of processes containing \( \phi \) and \( \psi \) (see Lemma 20), which may be of independent interest. Using our new bounds, we are able to show that the potentials quickly reach size \( O(wn \log n) \) (Lemma 15 and Lemma 16).

Recall that we restart time at the beginning of Phase 2. For the analysis of subphase 2.1 we define the event \( B_t \) by

\[
B_t = \{ \xi(s) \in E, \text{ for all } s \in \{0, 1, \ldots, t-1\} \},
\]

then we have the following result.
CHAPTER 3. DIVERSITY, FAIRNESS, AND SUSTAINABILITY IN POPULATION PROTOCOLS

Lemma 15. Given $r > 10$, there exists $C > 0$ and $\tau_{2,1} = O(wn \log n)$ such that with probability at least $n^{10-r}$,

$$\mathbb{1}_{B_t} \sum_{i=1}^{k} \sum_{j=1}^{k} \left( \frac{A_i(t)}{w_i} - \frac{A_j(t)}{w_j} \right)^2 \leq C wn \log n,$$

for all $t \in \{\tau_{2,1}, \tau_{2,1} + 1, \ldots, \tau_{2,1} + n^9\}$.

Note that in the above lemma we prove the result for $\mathbb{1}_{B_t} \phi(t)$ instead of $\phi(t)$. This is just for the sake of the formal analysis, as we know from Phase 1 that the events $B_t$ hold true w.h.p.

Similarly, we can prove that the second potential $\psi$ defined in Equation (3.9) also decreases very fast to a value of $O(wn \log n)$. However, we can only prove this after Subphase 2.1 is over as our analysis uses the fact that $\phi(t)$ is already small. Again, it is convenient to restart time at $\tau_{2,1}$ (i.e. we start Subphase 2.1 at time 0).

Define the set of configurations $\mathcal{E}' \subseteq \Omega$ as

$$\mathcal{E}' = \left\{ \xi \in \Omega : \sum_{i=1}^{k} \sum_{j=1}^{k} \left( \frac{A_i(t)}{w_i} - \frac{A_j(t)}{w_j} \right)^2 \leq C wn \log n \right\} \cap \mathcal{E},$$

(3.10)

where $\mathcal{E}$ is defined in Equation (3.7). Define the event $B'_t = \{\xi(s) \in \mathcal{E}', \forall s \in \{0, \ldots, t-1\}\}$, and we know it holds true after Phase 2.1 for at least $n^9$ time-steps by Lemma 15.

Lemma 16. Given $r > 9$, there exists $C' > 0$ and $\tau_{2,2} = O(wn \log n)$ such that with probability at least $n^{9-r}$,

$$\mathbb{1}_{B'_t} \sum_{i=1}^{k} \sum_{j=1}^{k} \left( \frac{a_i(t)}{w_i} - \frac{a_j(t)}{w_j} \right)^2 \leq C' wn \log n,$$

for all $t \in \{\tau_{2,2}, \ldots, \tau_{2,2} + n^8\}$.

Again, note that we prove the result for the process $\mathbb{1}_{B'_t} \psi(t)$ instead of $\psi(t)$, but, as before, this is just for having simpler analysis as in practice the events $B'_t$ hold w.h.p. in Phase 2.2.
Combining Theorem 14, Lemma 15, and Lemma 16 gives the main result of this section.

**Theorem 17.** Consider the Diversification protocol starting from an arbitrary configuration in $\Omega$. Then for any constant $r > 0$, there exists time-steps $\tau = \tau_1 + \tau_{2,1} + \tau_{2,2} = O(w^2 n \log n)$ and a constant $C > 0$ such that with probability at least $1 - O(n^{-r})$,

$$\sum_{i=1}^{k} \sum_{j=1}^{k} \left( \frac{A_i(t)}{w_i} - \frac{A_j(t)}{w_j} \right)^2 \leq C wn \log n$$

and

$$\sum_{i=1}^{k} \sum_{j=1}^{k} \left( \frac{a_i(t)}{w_i} - \frac{a_j(t)}{w_j} \right)^2 \leq C wn \log n,$$

for all $t \in \{\tau, \tau + 1, \ldots, \tau + n^8\}$.

Note that Theorem 9 is a straightforward corollary of Theorem 17. The strategy to prove Lemma 15 and Lemma 16 is to show that after $O(wn)$ time-steps the value of each potential function halves in the respective phases. In order to prove the potentials halve, we use the following lemmas that establish some properties about $\phi$ and $\psi$ respectively.

Denote by $F_t$ the filtration generated by the configurations of colours up to time $t$, and let $G_t$ be an increasing sequence of events such that $G_t \in F_{t-1}$ for $t \geq 1$ and $G_0 \in F_0$. The potential functions satisfy the following properties:

**Lemma 18.** Let $T \geq 0$ be an integer. Suppose the event $\{\xi_t \in E\}$ is contained in $G_{t+1}$ for all $t \geq T$, then there exist constants $C_1, C_2, C_3, C_4, C_5$ and $C_6$ (independent of $T$ and $k$, etc.) such that for any $t \geq T$,

1. $\mathbb{E}[\mathbf{1}_{G_{t+1}} \phi(t + 1)|F_t] \leq \mathbf{1}_{G_{t+1}} \phi(t) \left( 1 - \frac{C_1}{w} \right) + C_2$,

2. $\text{Var}(\mathbf{1}_{G_{t+1}} \phi(t + 1)|F_t) \leq C_3 \mathbf{1}_{G_t} \phi(t)/w + C_4 k$,

3. $|\mathbf{1}_{G_{t+1}} \phi(t + 1) - \mathbb{E}[\mathbf{1}_{G_{t+1}} \phi(t + 1)|F_t]| \leq C_5 \sqrt{\mathbf{1}_{G_t} \phi(t) k} + C_6 k$.

**Lemma 19.** Let $T \geq 0$ be an integer. Suppose that the event $\{\xi_t \in E\} \cap \{\psi(t - 1) \geq \max\{16\phi(t - 1), k^2\}\}$ is contained in $G_t$, then there exist constants $C_1, C_2, C_3, C_4, C_5, C_6$ (independent of $T$ and $k$, etc.) such that for any $t \geq T$,
1. \[\mathbb{E}[1_{G_{t+1}}\psi(t+1)|F_t] \leq 1_{G_t}\psi(t) \left(1 - \frac{\psi(t)}{n}\right) + C_2,\]

2. \[\text{Var}(1_{G_{t+1}}\psi(t+1)|F_t) \leq C_31_{G_t}\psi(t) + C_4k,\]

3. \[|1_{G_{t+1}}\psi(t+1) - \mathbb{E}[1_{G_{t+1}}\psi(t+1)|F_t]| \leq C_5\sqrt{1_{G_t}\psi(t)k} + C_6k.\]

Remark that the only difference between the previous two lemmas is that the second lemma requires an extra condition on the event \(G_t\). This difference is key, as any meaningful application of Lemma 19 requires knowing that \(\psi(t-1) \geq 16\phi(t-1)\), and thus \(\phi(t-1)\) needs to be already small for any simple application of it.

The previous two lemmas are not particularly useful by themselves, however we prove the following concentration inequality, which combining with the previous results leads to the proof of the main results (Theorem 17) of this phase.

**Lemma 20.** Let \(M(t)\) be a stochastic process adapted to a filtration \(F_t\). Suppose that \(M(t) \geq 0\) satisfies

(i) \(\mathbb{E}[M(t)|F_{t-1}] \leq (1 - \alpha)M(t-1) + \beta\), with \(0 < \alpha < 1\), and \(\beta > 0\); 

(ii) \(|\mathbb{E}[M(t)|F_{t-1}] - M(t)| \leq \gamma\); 

(iii) \(\text{Var}(M(t)|F_{t-1}) \leq \delta^2\).

Then for any \(\lambda > 0\),

\[P(M(t) \geq \mathbb{E}M(t) + \lambda) \leq \exp\left(-\frac{\lambda^2/2}{\delta^2/2\alpha - \alpha^2 + \gamma^3/3}\right).\]

Lemma 18 and Lemma 19 are proved in Section 3.5.2. The proofs of Lemma 15 and Lemma 16 are deferred to Section 3.5.2, and finally the proof of Lemma 20 can be found in Section 3.5.2.

### 3.3.3 Phase 3: A finer equilibrium - achieving fairness

Our next goal is to prove fairness of the Diversification protocol.

**Theorem 21.** For constant \(k\) and \(w\), the Diversification protocol achieves fairness.
The proof of Theorem 21 is provided at the end of this section via a Markov chain approximation. In order to demonstrate fairness, we first need to prove a finer characterisation of the colour distribution for large times. Such characterisation can only be proved by using the properties that are proved to hold after the first two phases of the process. Therefore, this new characterisation is proved in Phase 3, which starts immediately after Phase 2 and last for $\tau_3$ time-steps.

**Theorem 22.** Given $r > 9$, there exists $C > 0$ and $\tau = \tau_1 + \tau_{2,1} + \tau_{2,2} + \tau_3 = O(w^2 n \log n)$ such that with high probability $1 - O(n^{-r})$, it holds that

$$\left| A_i(t) - \frac{w_i n}{1 + w} \right| \leq C n^{3/4} (\log n)^{1/4}, \quad \text{and} \quad \left| a_i(t) - \frac{w_i n}{(1 + w)w} \right| \leq C n^{3/4} (\log n)^{1/4},$$

for all $i \in \{1, \ldots, k\}$ and for all $t \in \{\tau, \ldots, n^8\}$.

As usual, for the proof of the above theorem it will be convenient to restart time after phase 2. Recall that $B_t' = \{\xi(s) \in \mathcal{E}', \forall s \in \{0, \ldots, t - 1\}\}$, where $\mathcal{E}'$ is defined in Equation (3.10). For the proof of Theorem 22 we define the set $\widehat{\mathcal{E}} \subseteq \Omega$ as

$$\widehat{\mathcal{E}} = \left\{ \xi \in \Omega : \sum_{i=1}^{k} \sum_{j=1}^{k} \left( \frac{a_i}{w_i} - \frac{a_j}{w_j} \right)^2 \leq C' wn \log n \right\} \cap \left\{ \xi \in \Omega : \sum_{i=1}^{k} \sum_{j=1}^{k} \left( \frac{A_i}{w_i} - \frac{A_j}{w_j} \right)^2 \leq C' wn \log n \right\},$$

where $C' > 0$ is a large constant. By Theorem 17, we have $\xi(t) \in \widehat{\mathcal{E}}$ for all $t$ after Phase 2 for at least $n^8$ time-steps with probability at least $1 - O(n^{-r})$ (the constant $C'$ depends on $r$).

Define $\widehat{B}_t = \{\xi(s) \in \widehat{\mathcal{E}}, \forall s \in \{0, \ldots, t - 1\}\}$. The proof of Theorem 22 follows by combining Theorem 17 and the following lemma.

**Lemma 23.** Given $r > 9$ there exists $\widehat{C} > 0$ and $\tau_3 = O(wn \log n)$ such that with probability at least $n^{9-r}$, we have for all $t \in \{0, \ldots, n^8 - \tau_3\}$,

$$1_{\widehat{B}_t} \left( \frac{A(t)}{w} - a(t) \right)^2 \leq \widehat{C} n^{3/2} \sqrt{\log n}.$$
To see why Theorem 22 follows, we recall that Theorem 17 tells us that proportions
are respective between the number of vertices representing colours of the same shade,
however, we do not know how many vertices there are of each shade. Fortunately,
Lemma 23 tells us that if we wait for \( \tau_3 \) extra steps after the time when Theorem 17
starts working (given by \( \tau_1 + \tau_{2,1} + \tau_{2,2} \)), then Lemma 23 provides one extra equation,
which is enough to completely determine the number of agents supporting each dark
colour and light colour, up to a small additive error.

The proof of Lemma 23 is provided in Section 3.5.3, which follows the same steps
of the proof of Lemma 15 and Lemma 16.

Now we are ready to prove fairness of the Diversification protocol, by coupling
the trajectory of the states of agents with a Markov chain.

**Proof of Theorem 21.** Recall that each agent has one of \( k \) colours, and each colour
is either light or dark shaded. We denote by \( D_i \) the dark-shaded version of colour
\( i \), and \( L_i \) its light-shaded version. Therefore, in our protocol, we can imagine that
each agent is moving in the state space \( \{D_1, \ldots, D_k, L_1, \ldots, L_k\} \) according to some
transition rules that depend on the current state of the agent as well as the states of
all other agents (i.e. the trajectory of one agent is not a Markov chain by itself).

Let \( M^o(t) \) be the trajectory of a particular agent on the state space
\( \{D_1, D_2, \ldots, D_k, L_1, L_2, \ldots, L_k\} \). Even though \( M^o \) is not a Markov chain, we can
approximate it by one. Indeed, let \( M \) be the Markov chain on the state space
\( \{D_1, D_2, \ldots, D_k, L_1, \ldots, L_k\} \) with transition matrix \( P \) given by

\[
P(L_j, D_i) = \frac{w_i}{(1 + w)n}, \quad \text{for all } i, j ;
\]

\[
P(L_i, L_i) = 1 - \frac{w_i}{(1 + w)n}, \quad \text{for all } i ;
\]

\[
P(D_i, L_i) = \frac{1}{(1 + w)n}, \quad \text{for all } i ;
\]

\[
P(D_i, D_i) = 1 - \frac{1}{(1 + w)n}, \quad \text{for all } i .
\]
CHAPTER 3. DIVERSITY, FAIRNESS, AND SUSTAINABILITY IN POPULATION PROTOCOLS

while all other entries of $P$ are 0. Note that the matrix $P$ does not depend on $n$, and thus the mixing time of $P$ is constant.

We introduce $M$ as it describes the trajectory of a particle across all the states of the system when the system is perfectly balanced: with probability $1/n$ the agent is chosen to observe a random neighbour, and in perfect equilibrium each dark-shaded colour appears in a proportion $w_i/(1+w)$, and each light shaded colour in proportion $w_i/(w(1+w))$ according to Theorem 22. Despite the fact that such perfect equilibrium is very unlikely to achieve, Theorem 22 ensures that the system stays close to such equilibrium for long periods of time.

We claim that the stationary distribution of $M$ is given by

$$
\pi(L_i) = \frac{w_i}{n}, \quad \text{and} \quad \pi(D_i) = \frac{w_i}{1+w},
$$

which can be directly verified by the definition of stationary distribution, that is, 

$$
\pi(D_i) = \pi(D_i)P(D_i, D_i) + \sum_{j=1}^{k} \pi(L_j)P(L_j, D_i), \quad \text{and}
$$

$$
\pi(L_i) = \pi(L_i)P(L_i, L_i) + \pi(D_i)P(D_i, L_i). \quad (3.11)
$$

The trajectory of $M^o$ can be approximated by the transition matrix $P$. Indeed, let $err = c(\log n)^{1/4}/n^{1/4}$, for some sufficiently large constant $c$. Then by Theorem 22 we have that for any $t \in \{\tau, \ldots, n^8\}$, the following holds with probability $1 - \delta$, where $\delta = O(n^{-r})$,

$$
P(M^o(t+1) = L_i | M^o(t) = D_i, \mathcal{F}_t) = \frac{A_i(t)/w_i}{n(n-1)} = P(D_i, L_i) \pm err, \quad (3.12)
$$

since with probability $1/n$ the agent is chosen, and it changes its colour from dark shade to light shade with probability $A_i(t)/(w_i(n-1))$ by the definition of the protocol. In a similar fashion, we have

$$
P(M^o(t+1) = D_i | M^o(t) = D_i, \mathcal{F}_t) = P(D_i, D_i) \pm err, \quad
$$

$$
P(M^o(t+1) = D_j | M^o(t) = L_i, \mathcal{F}_t) = P(L_i, D_j) \pm err, \quad \text{for all } i, j,
$$

$$
P(M^o(t+1) = L_i | M^o(t) = L_i, \mathcal{F}_t) = P(L_i, L_i) \pm err. \quad (3.13)
$$
Note that other transitions have probability 0 as they are impossible by the design of the protocol (they also have entries 0 in the transition matrix $P$).

We fix a state, say $D_\ell$, and define the following transition matrix $P^+_{D_\ell}$, given by

$$
P^+_{D_\ell}(D_\ell, L_\ell) = P(D_\ell, L_\ell) - \text{err} ,
$$

$$
P^+_{D_\ell}(D_\ell, D_\ell) = P(D_\ell, D_\ell) + \text{err} ,
$$

$$
P^+_{D_\ell}(D_i, L_\ell) = P(D_i, L_\ell) + \text{err} , \quad \text{for } i \neq \ell ,
$$

$$
P^+_{D_\ell}(D_i, D_\ell) = P(D_i, D_\ell) - \text{err} , \quad \text{for } i \neq \ell ,
$$

$$
P^+_{D_\ell}(L_i, D_\ell) = P(L_i, D_\ell) + k \cdot \text{err} , \quad \text{for all } i ,
$$

$$
P^+_{D_\ell}(L_i, D_j) = P(L_i, D_j) - \text{err} , \quad \text{for } i, \text{ and } j \neq \ell ,
$$

$$
P^+_{D_\ell}(L_i, L_i) = P(L_i, L_i) - \text{err} , \quad \text{for all } i .
$$

The idea is that $P^+_{D_\ell}$ is a transition matrix, such that it increases the probabilities of all the transitions of $M^0$ that move the agent closer to the state $D_\ell$ and decreases the probabilities of the transitions of $M^0$ that prevent the agent from getting closer to $D_\ell$. Additionally, we can define $P^-_{D_\ell}$ in the same way as $P^+_{D_\ell}$ but change the sign for the err term, and thus representing a transition matrix that decreases transition probabilities that move the agent closer to $D_\ell$, and increases the others. Similarly, we define $P^+_{L_\ell}$ and $P^-_{L_\ell}$. It can be easily verify that the stationary distributions of $M^+$ and $M^-$ are the same as $M$ up to errors of order err.

It is convenient to restart time at time $\tau$, from which the approximation of Theorem 22 starts holding true. Recall that the event of Theorem 22 holds with probability $1 - \delta$, where $\delta = O(n^{-r})$ for fixed $r > 9$. Let $N^0_{D_\ell}(t)$ be the number of times $M^0$ has been in state $D_\ell$ up to time $t$, and similarly let $N^+_{D_\ell}(t)$ be the number of times a particle moving according to $P^+_{D_\ell}$ hits $D_\ell$ up to time $t$, and define $N^-_{D_\ell}$ similarly. By the construction of the transition matrices $P^+_{D_\ell}$ and $P^-_{D_\ell}$, with probability at least $1 - \delta$ the following majorisation between the random variables holds:

$$
N^-_{D_\ell}(t) \preceq N^0_{D_\ell}(t) \preceq N^+_{D_\ell}(t) ,
$$

72
for all $t \leq n^8 - \tau$.

To see this, note that starting at any state $x \neq D_\ell$ one can use a coupling argument to show that the time it takes for the process $M^o$ to reach $D_\ell$ from $x$ is at most the time it takes for the chain $P^+_{D_\ell}$ to reach $D_\ell$ from $x$. A similar argument can be made for the time that the chain stays put at $D_\ell$. With a similar argument we can see that $N^{-}_{D_\ell}(t) \precsim N^o_{D_\ell}(t)$.

In the low-probability case (with probability $\delta$) that the event of Theorem 22 does not hold, we consider the simple bound $0 \leq N^o_{D_\ell}(n^8) \leq n^8$.

Note that $P^+_{D_\ell}$ is a recurrent and ergodic Markov chain with (unique) stationary distribution denoted by $\pi^+$. Since the state space has size $2k$ which is finite, then the mixing time is finite as well. Therefore, by Chernoff’s bounds for Markov chains (see Theorem 26), there exists $c > 0$ such that

$$|N^+_{D_\ell}(t) - \pi^+(D_\ell)t| \leq c\sqrt{\pi^+(D_\ell)t \log n}$$

with probability $1 - \delta$.

Finally, we notice that $P^+_{D_\ell}$ is just a small perturbation of $P$, and it can be easily verified that $\pi^+(D_\ell) = \pi(D_\ell) + O(\text{err})$ (see Equation (3.11)), therefore

$$N^+_{D_\ell}(t) \leq \frac{w_\ell t}{1 + w} + O\left(\sqrt{\frac{w_\ell t \log n}{1 + w}} + (\text{err} \cdot t)\right),$$

with probability at least $1 - \delta$. The same argument applies to $N^-_{D_\ell}$.

We conclude the proof by noting that from the beginning of the process, the approximations in Equation (3.12) and Equation (3.13) only hold from time $\tau = O(wn \log n)$ as stated in Theorem 22, and such approximation is valid up to time $n^8$ with polynomially large probability. Therefore, we still need to count the number of hits to $D_\ell$ in the interval $[0, \tau]$. However, since $\tau = O(n \log n)$, even the worst-case bounds in such interval are second-order terms compared to the hits in the interval $[\tau, n^8]$. 

73
We deduce that with probability at least \(1 - O(\delta)\) we have for a particular agent \(u\) that
\[
\frac{|\{t \in [0, n^8] : c_u(t) = \ell, b_u(t) = 1\}|}{n^8} = \frac{w_\ell}{1 + w}(1 \pm o(1))
\]
Similarly, by repeating the same argument for \(L_\ell\), we obtain, with probability at least \(1 - O(\delta)\) that
\[
\frac{|\{t \in [0, n^8] : c_u(t) = \ell, b_u(t) = 0\}|}{n^8} = \frac{w_\ell}{1 + w + w}(1 \pm o(1))
\]
Thus applying a union bound on the above two equations gives that the amount of time that an agent \(u\) spent in colour \(l\) satisfies that, with probability at least \(1 - O(\delta)\),
\[
\frac{|\{t \in [0, n^8] : c_u(t) = \ell\}|}{n^8} = \frac{w_\ell}{w}(1 \pm o(1))
\]
By a union bound, the same holds for all colours \(i \in \{1, \ldots, k\}\) and all agents at the same time, with probability \(1 - O(nk\delta)\).

Note that we just prove the approximation up to time \(n^8\), but for larger times we divide time in segments of length \(n^8\), and in each segment we allow \(\tau\) time-steps in order to reach the approximation of Theorem 22, and then we apply the previous domination argument. Note that the probability that we have a successful approximation is \(1 - O(kn\delta) = 1 - O(n^{-r+1})\) in each interval where \(r\) is arbitrarily large. In case that the event of Theorem 22 does not hold, we just bound the number of visits by \(n^8\) and 0, respectively. Choosing \(r > 10\) ensures that w.h.p. fairness is achieved in the interval. Using the Chernoff’s bounds, one can also show fairness over longer intervals.

3.4 Conclusions and future work

We have introduced the notion of diversity, fairness and sustainability to the realm of population protocols. We showed that the simple Diversification protocol achieves those properties.
CHAPTER 3. DIVERSITY, FAIRNESS, AND SUSTAINABILITY IN POPULATION PROTOCOLS

Natural open problems are to analyse the derandomised version of our protocol (proposed in Section 3.2.1), and to extend our analysis to \( k \) and \( w \) depending on the number of agents \( n \). Another research direction is to find protocols that attain a stronger notion of diversity, where the error term in Equation (3.1) is much smaller than \( \tilde{O}(1/\sqrt{n}) \). We think that finding such protocol will lead to very interesting trade-off between time and space complexity. Another line of research is to investigate the diversification protocol in different graph topologies, other than the complete graph.

An interesting research direction is to describe what lies in between consensus and diversification? Finally, one can ask whether our protocol emerges naturally in the context of learning, or in other distributed systems, perhaps even in biology.

3.5 Omitted proofs

Recall that we assume \( w_i \) for \( i \in [k] \) and \( k \) are constants.

3.5.1 Proof of Section 3.3.1

Proof of Lemma 10. To begin with, note that

\[
\sum_{i=1}^{k} \frac{A_i^2}{w_i} = \frac{w}{w_i} \sum_{i=1}^{k} \left( \frac{A_i}{w_i} \right)^2 w_i \geq w \left( \sum_{i=1}^{k} \frac{A_i w_i}{w_i} \right)^2 = \frac{A^2}{w},
\]

where in the first inequality we used the Jensen’s inequality since \( w_i/w > 0 \) and \( \sum_{i=1}^{k} w_i/w = 1 \).

Let \( p \) be the probability of increasing \( a \) and \( q \) the probability of decreasing it. Using the above inequality, we get

\[
p = \sum_{i=1}^{k} \frac{A_i(A_i-1)}{n(n-1)w_i} \geq \frac{A^2}{wn(n-1)} - \frac{A}{n(n-1)}, \quad \text{and}
\]

\[
q = \sum_{i=1}^{k} \frac{a_i A}{n(n-1)} = \frac{aA}{n(n-1)}. \tag{3.14}
\]
Without loss of generality, assume that the current configuration is not in $R_1$, we must have that $a < n(1 - \varepsilon)\frac{1}{w+1}$ and therefore, $A \geq n - n(1 - \varepsilon)\frac{1}{w+1}$. Thus,

$$\frac{A}{a} \geq \frac{1 - \frac{1-\varepsilon}{w+1}}{\frac{1-\varepsilon}{1+w}} = \frac{1 + w - 1 + \varepsilon}{1 - \varepsilon} = \frac{w + \varepsilon}{1 - \varepsilon}.$$ 

Then, Equation (3.14) gives that the probability to decrease in an active time-step (i.e. when $a$ changes its value) is

$$\frac{q}{q + p} \leq \frac{aA}{aA + A^2/w - A} \leq \frac{w}{w + A/a - O(w/n)} \leq \frac{w}{2w/3 + (w + \varepsilon)/(1 - \varepsilon)} \leq \frac{1 - \frac{1+\varepsilon}{2}}{2 - \frac{1+\varepsilon}{3}}.$$ 

This yields a biased random walk that maintains the probability of increasing to be at least $1/2 + \varepsilon/3$ until $\xi(t) \in R_1$.

Considering only the active time-steps, where $a(t)$ either increases or decreases its value, and an interval of active steps of length $\ell$, we define the stopping time $T^* = \min_{\text{active step } t} \{\xi(t) \in R_1\}$. Let $Z_1, Z_2, \ldots$ be independent random variables where each variable is either 1 or $-1$ with probability $1/2 + \varepsilon/3$ and $1/2 - \varepsilon/3$ respectively. Let $X_t = \sum_{i=1}^{t}(Z_i - EZ_i)$, then $X_t$ is a martingale. Denote by $I$ and $D$ the number of increasing and decreasing steps respectively in the interval of active steps of length $\ell$. Then $I + D = \ell$ and $I - D = \sum_{i=1}^{\ell}Z_i = X_\ell + 2\ell\varepsilon/3$.

According to the Azuma-Hoeffding inequality, for the number of increases $I$ minus the number of decreases $D$ we have

$$P(I - D \leq \ell\varepsilon) = P\left(X_\ell < \frac{\ell\varepsilon}{3}\right) \leq \exp\left(-\frac{(\ell\varepsilon/3)^2}{2\ell}\right).$$

Choosing $\ell = 2n/\varepsilon$, it guarantees that $T^* \leq \ell$ with probability at least $1 - \exp(-\varepsilon n/9)$. It remains to show the relationship between active time-steps and regular time-steps. Recall,

$$p \geq \frac{A^2/w - A}{n^2 - n} \geq \frac{A^2 - Aw}{wn^2} = \Omega(1/w).$$
Hence, there is only a factor \( w \) difference throughout the entire regime until \( T^* \). This yields the first part of the claim, i.e. \( T_1 = \{ t \geq 0 : \xi(t) \in R_1 \} = O(nw/\varepsilon) \), with probability at least \( 1 - \exp(-c\varepsilon n) \).

For the second part, we consider the setting where the biased random walk starts at \((1 - 3\varepsilon/2)\frac{n}{w+1}\) and increases with probability at least \( 1/2 + \varepsilon/3 \). We apply Theorem 25 to determine the probability of hitting \((1 - \varepsilon)\frac{n}{w+1}\) before hitting \((1 - 2\varepsilon)\frac{n}{w+1}\).

Define \( Z_t = a_t - (1 - 3\varepsilon/2)\frac{n}{w+1} + (\varepsilon/2)\frac{n}{w+1} \), \( b = \varepsilon\frac{n}{w+1} \), and \( s = \frac{\varepsilon}{2}\frac{n}{w+1} \). Let \( T = \min\{ t \geq 0 \mid Z_t \in \{ 0, b \} \} \). Then, applying Theorem 25 and using that \( \frac{w-x}{1-x} \leq y \), we get

\[
P(Z_T = 0) = \left( \frac{1/2 - \varepsilon/3}{1/2 + \varepsilon/3} \right)^{a} - \left( \frac{1/2 - \varepsilon/3}{1/2 + \varepsilon/3} \right)^{b} \leq \left( \frac{1/2 - \varepsilon/3}{1/2} \right)^{s} \leq \left( 1 - \frac{2\varepsilon}{3} \right)^{s} \leq \exp\left(-c\varepsilon^2/w\right),
\]

for \( c > 0 \) small enough.

By taking union bound over \( \exp(cn\varepsilon^2/w) \) time-steps yields that \( T'_1 \geq \exp(cn\varepsilon^2/w) \) with probability at least \( 1 - \exp(-c\varepsilon^2/w) \), which completes the proof.

\[\square\]

**Proof of Lemma 11.** Let \( p \) be the probability of increasing \( A_i \), and \( q \) the probability of decreasing it. We have

\[
p = \sum_{j=1}^{k} \frac{a_j A_i}{n (n-1)} = \frac{a A_i}{n^2 - n}, \quad \text{and}
\]

\[
q = \frac{A_i (A_i - 1)}{n (n-1)} \frac{1}{w_i}.
\]

Assume that \( A_i \) does not satisfy the property of region \( R_2 \) and note that \( a \) is in region \( S_1 \) such that \( a \geq (1 - 2\varepsilon)n/(w+1) \), then

\[
\frac{A_i}{aw_i} \leq \frac{(1-3\varepsilon)w_i}{1+w} \leq \frac{1 - 3\varepsilon}{1 + 2\varepsilon}.
\]
As done in the proof of Lemma 10, we consider only active steps, then for \( \varepsilon \leq 2/5 \), it holds that

\[
\frac{p}{q + p} = \frac{a}{\frac{A_i - 1}{w_i} + a} \geq \frac{1 - 2\varepsilon}{1 - 2\varepsilon + 1 - 3\varepsilon} \geq \frac{1}{2} + \frac{\varepsilon}{4}.
\]

Using a similar biased random work analysis on the active steps as what we did in the proof of Lemma 10, we obtain that in \( O(n/\varepsilon) \) active steps \( A_i \) has the desired size with overwhelming probability.

Note that the ratio of active steps is proportional to \( p \), which is up to constants due to our lower bound on \( a \) and \( A_i/(wn) \). Thus increasing \( A_i \) by 1 takes \( O(\frac{nw}{A_i\varepsilon}) \) time-steps. Hence, the total time to raise \( A_i \) all the way up to linear in \( nw_i/w \) is dominated by the geometric series \( \sum_{i=0}^{\log n/2} 2^i \cdot O(\frac{nw}{2\varepsilon}) = O(\frac{wn \log n}{\varepsilon}) \).

Finally, similar to Lemma 10, \( A_i \) remains within the stated bounds with overwhelming probability. \( \square \)

### 3.5.2 Proofs of Section 3.3.2

#### Proof of Lemma 18 and Lemma 19

We proceed to prove Lemma 18 and Lemma 19. In the proofs we abuse notation and use letters such as \( c, c_1, c_2, C, C_1 \), etc. to denote constants, and for convenience we reuse the same notation/name for different constants.

**Proof of Lemma 18.** We first introduce some notation. For simplicity, let \( A_i = A_i(t) \) and \( A'_i = A_i(t + 1) \). We define the following variables \( d_i = \frac{A'_i}{w_i} - \frac{A_i}{w_i} \), and \( q_i = \frac{A_i}{w_i} \). Let \( D_r = \sum_{i=1}^{k} d_i^r \), and \( Q_r = \sum_{i=1}^{k} q_i^r \), for \( r = \{1, 2, \ldots\} \). Finally, denote \( q_{ij} = q_i - q_j \) and \( d_{ij} = d_i - d_j \).

With the above notation, the potential function \( \phi \) defined in Equation (3.8) be-
comes \( \phi(t) = \sum_{i=1}^{k} \sum_{j=1}^{k} q_{ij}^2 \). Therefore,

\[
\mathbb{I}_{G_{t+1}} \phi(t+1) = \mathbb{I}_{G_{t+1}} \sum_{i=1}^{k} \sum_{j=1}^{k} (q_{ij} + d_{ij})^2 = \mathbb{I}_{G_{t+1}} \left( \phi(t) + 2 \sum_{i=1}^{k} \sum_{j=1}^{k} q_{ij} d_{ij} + \sum_{i=1}^{k} \sum_{j=1}^{k} d_{ij}^2 \right), \tag{3.15}
\]

Note that \( d_i \) is the proportional change of the value \( A_i \) in one step, and that \( A_i \) increases its value by 1 with probability \( \frac{a A_i}{n(n-1)} \) and decreases by 1 with probability \( \frac{A_i(A_i - 1)}{w_i n(n-1)} \). Then,

\[
\mathbb{E}[d_i | F_t] = \frac{1}{w_i} \left( \frac{a A_i}{n(n-1)} - \frac{A_i(A_i - 1)}{n(n-1) w_i} \right) = \frac{1}{n(n-1)} (a q_i - q_i^2 + \frac{q_i}{w_i}), \text{ and} \]

\[
\mathbb{E}[d_{ij}^2 | F_t] = \frac{1}{w_i} \left( \frac{a A_i}{n(n-1)} + \frac{A_i(A_i - 1)}{n(n-1) w_i} \right) = \frac{1}{w_i n(n-1)} (a q_i + q_i^2 - \frac{q_i}{w_i}). \tag{3.16}
\]

Notice that \( |\mathbb{E}[d_i | F_t] - (aq_i - q_i^2)/n^2| \leq 3/n \). Then we have the following bound on the term \( \sum_{i=1}^{k} \sum_{j=1}^{k} q_{ij} d_{ij} \) in Equation (3.15): for some constant \( c > 0 \),

\[
\mathbb{I}_{G_{t+1}} \mathbb{E} \left[ \sum_{i=1}^{k} \sum_{j=1}^{k} q_{ij} d_{ij} | F_t \right] \leq \mathbb{I}_{G_{t+1}} \sum_{i=1}^{k} \sum_{j=1}^{k} \frac{q_{ij}}{n^2} [a q_i - q_i^2 - (aq_j - q_j^2)] + \sum_{i=1}^{k} \sum_{j=1}^{k} q_{ij} \frac{3}{n} \\
= \mathbb{I}_{G_{t+1}} \sum_{i=1}^{k} \sum_{j=1}^{k} q_{ij}^2 \frac{a - (q_i + q_j)}{n^2} + c \\
\leq -\mathbb{I}_{G_{t+1}} \frac{1 - 3\delta}{(1 + w)n} \phi(t) + c,
\]

where in the last step we used the facts that \( \{\xi(t) \in \mathcal{E}\} \subseteq G_{t+1} \). By the definition of \( \mathcal{E} \) in Equation (3.7) we have \( a \leq (1 + \delta)n/w \) and \( q_i \geq (1 - \delta)n/w \), where \( \delta \) is sufficiently small.

As for the term \( \sum_{i=1}^{k} \sum_{j=1}^{k} d_{ij}^2 \) in Equation (3.15), we notice that \( d_i d_j = 0 \) for \( i \neq j \), which yields

\[
\mathbb{I}_{G_{t+1}} \mathbb{E} \left[ \sum_{i=1}^{k} \sum_{j=1}^{k} d_{ij}^2 | F_t \right] = \mathbb{I}_{G_{t+1}} 2(k - 1) \mathbb{E} \left[ \sum_{i=1}^{k} d_i^2 | F_t \right] \leq \mathbb{I}_{G_{t+1}} 2(k - 1) \sum_{i=1}^{k} \frac{a q_i + q_i^2}{n^2} \leq \mathbb{I}_{G_{t+1}} \frac{4k^2 (1 + \delta)^2}{(1 + w)^2},
\]

79
where in the last step we apply \( a \leq (1 + \delta)n/w \) and \( q_i \leq (1 + \delta)n/w \) for sufficiently small \( \delta \), since \( \{\xi(t) \in \mathcal{E}\} \subseteq G_{t+1} \).

Notice that \( k \leq w \). Thus, replacing Equation (3.15) with the above calculations and bounds yields, for some constants \( c_1, c_2 > 0 \),

\[
\mathbb{E} \left[ 1_{G_{t+1}} \phi(t + 1) \right] = 1_{G_{t+1}} \left( \phi(t) - \frac{2(1 - 3\delta)}{(1 + w)n} \phi(t) + c_1 + \frac{4k^2(1 + \delta)^2}{(1 + w)^2} \right)
\]

\[
\leq 1_{G_{t+1}} \phi(t) \left( 1 - \frac{c_1}{wn} \right) + c_2 .
\]

We continue to prove Item 2. Note that with the notation introduced in the proof of Lemma 18, the potential function \( \phi \) can be expressed as

\[
\phi(t) = \sum_{i=1}^{k} \sum_{j=1}^{k} (q_i - q_j)^2 = 2kQ_2 - 2Q_1^2 , \quad \text{and}
\]

\[
\phi(t + 1) = 2k \sum_{i=1}^{k} (q_i + d_i)^2 - 2(Q_1 + D_1)^2
\]

\[
= 2k \left( Q_2 + D_2 + 2 \sum_{i=1}^{k} q_id_i \right) - 2Q_1^2 - 2D_1^2 - 4Q_1D_1
\]

\[
= \phi(t) + 4k \sum_{i=1}^{k} q_id_i - 4Q_1D_1 + (2k - 2)D_2 .
\]

We compute \( \text{Var}(1_{G_{t+1}} \phi(t + 1) - 1_{G_t} \phi(t) | \mathcal{F}_t) \) by using the previous two equations. First note that if \( 1_{G_{t+1}} = 0 \) then \( \text{Var}(1_{G_{t+1}} \phi(t + 1) - 1_{G_t} \phi(t) | \mathcal{F}_t) = \text{Var}(1_{G_t} \phi(t) | \mathcal{F}_t) = 0 \), since \( 1_{G_t} \phi(t) \in \mathcal{F}_t \) and \( 1_{G_{t+1}} \in \mathcal{F}_t \). If \( 1_{G_{t+1}} = 1 \), then \( 1_{G_t} = 1 \) and thus

\[
\text{Var}(1_{G_{t+1}} \phi(t + 1) - 1_{G_t} \phi(t) | \mathcal{F}_t) = 1_{G_{t+1}} \text{Var}(\phi(t + 1) - \phi(t) | \mathcal{F}_t)
\]

\[
= 1_{G_{t+1}} \text{Var} \left( \sum_{i=1}^{k} (4kq_i - 4Q_1)d_i + (2k - 2)D_2 \bigg\vert \mathcal{F}_t \right)
\]

\[
\leq 1_{G_{t+1}} \mathbb{E} \left[ \left( \sum_{i=1}^{k} (4kq_i - 4Q_1)d_i + (2k - 2)D_2 \right)^2 \bigg\vert \mathcal{F}_t \right]
\]

\[
\leq 2 \cdot 1_{G_{t+1}} \mathbb{E} \left[ \left( \sum_{i=1}^{k} (4kq_i - 4Q_1)d_i \right)^2 \bigg\vert \mathcal{F}_t \right] + 2 \cdot 1_{G_{t+1}} \mathbb{E} \left[ ((2k - 2)D_2)^2 \bigg\vert \mathcal{F}_t \right] .
\]

(3.17)
We upper bound each of the terms in the above equation as follows: By Equation (3.16) and the fact that \( a \leq (1 + \delta)n/w \) and \( q_i \leq (1 + \delta)n/w \) for sufficiently small \( \delta \) since \( \{\xi(t) \in E\} \subseteq G_{t+1} \), we know that \( E[d_i] \leq \frac{4d_i^2}{(1+w)^2} \) and \( E[d_i^2] \leq \frac{2(1+\delta)^2}{(1+w)^2} \). Also notice that \( E[d_i^{2r+1}] \leq E[d_i] \) and \( E[d_i^{2r}] \leq E[d_i^2] \). Then for term on the right-hand side of the last inequality in Equation (3.17) we have,

\[
2 \cdot 1_{G_{t+1}} E \left[ \left( (2k-2)D_2 \right)^2 | F_t \right] = 8 \cdot 1_{G_{t+1}} (k-1)^2 \sum_{i=1}^{k} E \left[ d_i^4 | F_t \right] \\
\leq 8 \cdot 1_{G_{t+1}} k^2 2(1+\delta)^2 k \leq 8(1+\delta)^2 k = c_4 k .
\]

For the term on the left-hand side of the last inequality in Equation (3.17), we notice that \( d_id_j = 0 \) for \( i \neq j \), and \( \phi(t) = 2kQ_2 - 2Q_1^2 \), then

\[
2 \cdot 1_{G_{t+1}} E \left[ \left( \sum_{i=1}^{k} (4kq_i - 4Q_1)d_i \right)^2 | F_t \right] \\
= 2 \cdot 1_{G_{t+1}} \sum_{i=1}^{k} (4kq_i - 4Q_1)^2 E[d_i^2 | F_t] = 16 \cdot 1_{G_{t+1}} 2(k^2Q_2 - kQ_1^2) E[d_i^2 | F_t] \\
\leq 16 \cdot 1_{G_{t+1}} k \phi(t) \frac{(1+\delta)^2}{(1+w)^2} \leq 1_{G_{t+1}} \frac{c_3}{w} \phi(t) .
\]

Hence, substituting Equation (3.17) with the above calculations yields, for \( c_3, c_4 > 0 \),

\[
\text{Var}(1_{G_{t+1}} \phi(t + 1) - 1_{G_t} \phi(t) | F_t) = 1_{G_{t+1}} \frac{c_3}{w} \phi(t) + c_4 k .
\]

Finally, for Item 3, recall that \( q_{ij} = q_i - q_j \) and \( d_{ij} = d_i - d_j \), and then from Equation (3.15) we get

\[
1_{G_{t+1}} \phi(t + 1) - E[1_{G_{t+1}} \phi(t + 1) | F_t] \\
= 1_{G_{t+1}} \left| \sum_{i=1}^{k} \sum_{j=1}^{k} (2q_{ij} (d_{ij} - E[d_{ij} | F_t]) + d_{ij}^2 - E[d_{ij}^2 | F_t]) \right| \\
\leq 1_{G_{t+1}} \left( 2 \sqrt{\phi(t) \sqrt{6k + 4k}} \right) \leq c_5 \sqrt{1_{G_t} \phi(t) k} + c_6 k ,
\]

81
where in the first inequality we used the Cauchy-Schwarz inequality, and the fact that \(|d_{ij}| \leq 1\).

We show Lemma 19 using a similar idea as the proof of Lemma 18.

**Proof of Lemma 19.** Let \(a_i = a_i(t), a'_i = a_i(t + 1), \bar{q}_i = a_i/w_i, \bar{d}_i = (a'_i - a_i)/w_i\).

Define \(\bar{Q}_j = \sum_{i=1}^{k}(\bar{q}_i)^j\) and \(\bar{D}_j = \sum_{i=1}^{k}(\bar{d}_i)^j\). Recall that \(A = \sum_{i=1}^{k} A_i\) and we denote \(A = A(t)\). We also consider all the notation defined in the proof of Lemma 18.

We begin by discussing some properties of \(\bar{d}_i\). Since \(a_i\) increases by 1 with probability \(A_i(A_i - 1)/n(n - 1)\) and decreases by 1 with probability \(a_iA_i/n(n - 1)\) we have that

\[
E[\bar{d}_i|F_t] = \frac{A_i(A_i - 1)}{n(n - 1)} \left( - \frac{A_i}{n(n - 1)} \right) = \frac{1}{n^2} \left( q_i^2 - \bar{q}_i A_i \right) - \frac{q_i}{w_i} \frac{n_i}{w_i n^2},
\]

which implies that

\[
\left| E[\bar{d}_i|F_t] - \frac{1}{n^2} (q_i^2 - \bar{q}_i A_i) \right| \leq \frac{3}{n}.
\]

Let’s prove the first item of the lemma. Denote \(\bar{q}_{ij} = \bar{q}_i - \bar{q}_j\) and \(\bar{d}_{ij} = \bar{d}_i - \bar{d}_j\).

Then \(\psi(t) = \sum_{i=1}^{k} \sum_{j=1}^{k} (\bar{q}_{ij} + \bar{d}_{ij})^2\). Note that \(\bar{d}_i \bar{d}_j = 0\) for \(i \neq j\) and \(|\bar{d}_i|' \leq |\bar{d}_i| \leq 1\) for \(r \geq 1\), then

\[
E[\psi(t + 1)|F_t] = E \left[ \sum_{i=1}^{k} \sum_{j=1}^{k} (\bar{q}_{ij} + \bar{d}_{ij})^2 \middle| F_t \right]
\]

\[
\leq \psi(t) + 2 \sum_{i=1}^{k} \sum_{j=1}^{k} \bar{q}_{ij} E[\bar{d}_{ij}|F_t] + 2k \sum_{i=1}^{k} E[\bar{d}_i^2|F_t]
\]

\[
\leq \psi(t) + \frac{2}{n^2} \sum_{i=1}^{k} \sum_{j=1}^{k} \bar{q}_{ij} (q_i^2 - q_j^2 - A\bar{q}_{ij}) + \frac{6}{n} \sum_{i=1}^{k} \sum_{j=1}^{k} |\bar{q}_{ij}| + 2k \sum_{i=1}^{k} E[|d_i||F_t]
\]

\[
\leq \left( 1 - \frac{2A}{n^2} \right) \psi(t) + \frac{2}{n^2} \sum_{i=1}^{k} \sum_{j=1}^{k} \bar{q}_{ij} q_i + q_j + \frac{6}{n} \sum_{i=1}^{k} \sum_{j=1}^{k} |\bar{q}_{ij}|
\]

\[
+ 2k \sum_{i=1}^{k} \left( \frac{q_i^2 + q_i A_i}{n^2} + \frac{3}{n} \right).
\] (3.18)
CHAPTER 3. DIVERSITY, FAIRNESS, AND SUSTAINABILITY IN POPULATION PROTOCOLS

If we work on the event \( \{ \xi(t) \in \mathcal{E} \} \), then \( q_i \leq c_1 n/k \) for some \( c_1 > 0 \). Applying the Cauchy-Schwarz inequality yields

\[
\sum_{i=1}^{k} \sum_{j=1}^{k} \overline{q}_{ij} q_{ij} (q_i + q_j) \leq \sqrt{\psi(t) \phi(t)} c_1 n/k , \quad \text{and}
\]

\[
\sum_{i=1}^{k} \sum_{j=1}^{k} |\overline{q}_{ij}| \leq k \sqrt{\psi(t)} .
\]

It holds that \( a \leq c_2 n/k \) for some \( c_2 > 0 \) on the event \( \{ \xi(t) \in \mathcal{E} \} \). Then for some \( c_3 > 0 \),

\[
2k \sum_{i=1}^{k} \left( \frac{q_i^2 + q_i A}{n^2} + \frac{3}{n} \right) \leq 2k \sum_{i=1}^{k} \left( \frac{q_i^2 + a_i A}{n^2} + \frac{3}{n} \right) \]

\[
= 2k \left( \sum_{i=1}^{k} q_i^2 + a A + 3 kn \right) \]

\[
\leq 2k \left( \sum_{i=1}^{k} \left( \frac{c_1 n}{k} \right)^2 + \frac{c_2 n}{k} A + 3 kn \right) \leq c_3 .
\]

Replacing the previous three equations in Equation (3.18) yields

\[
\mathbb{E}[\mathbf{1}_{G_{t+1}} \psi(t+1) | \mathcal{F}_t] \leq \mathbf{1}_{G_{t+1}} \psi(t) \left( 1 - \frac{2A}{n^2} \right) + \frac{2}{n^2} \sqrt{\psi(t) \phi(t)} c_1 n/k + \frac{6k \sqrt{\psi(t)}}{n} + c_3
\]

\[
= \mathbf{1}_{G_{t+1}} \psi(t) \left( 1 - \frac{2A - (2c_1 n/k) \sqrt{\phi(t)/\psi(t)} - 6kn/\sqrt{\psi(t)}}{n^2} \right) + c_3
\]

\[
\leq \mathbf{1}_{G_{t+1}} \psi(t) \left( 1 - \frac{(1 + o(1))n - O(n/k) - (1 - o(1))n}{n^2} \right) + c_3
\]

\[
= \mathbf{1}_{G_{t+1}} \psi(t) \left( 1 - \frac{C_1}{n} \right) + C_2 ,
\]

for some \( c_1, c_2 > 0 \), where in the previous to last equation we used that in the event \( G_{t+1} \) it holds that \( \psi(t) \geq \max\{16\phi(t), k^2\} \), and that \( 2A \geq 2(1 - \delta)nw/(w + 1) \geq (1 + c)n \) as \( \delta \) is small enough and \( w \geq k \geq 2 \).

We continue with Item 2. Note that

\[
\psi(t+1) - \psi(t) = 4k \sum_{i=1}^{k} \overline{q}_i \bar{d}_i - 4\overline{Q}_1 \bar{D}_1 + 2k \sum_{i=1}^{k} (\bar{d}_i)^2 - 2(\bar{D}_1)^2
\]

\[
= 4k \sum_{i=1}^{k} \overline{q}_i \bar{d}_i - 4\overline{Q}_1 \bar{D}_1 + (2k - 2)\bar{D}_2 .
\]

83
CHAPTER 3. DIVERSITY, FAIRNESS, AND SUSTAINABILITY IN POPULATION PROTOCOLS

Following the same reasoning of Item 2 of Lemma 18 yields

\[
\text{Var}(1_{G_{t+1}} \psi(t+1) - 1_{G_t} \psi(t) \mid F_t) = 1_{G_{t+1}} \text{Var}(\psi(t+1) - \psi(t) \mid F_t)
\]

\[
\leq 21_{G_{t+1}} \left( \mathbb{E} \left( \left( \sum_{i=1}^{k} (4k \bar{q}_i - 4 \bar{Q}_1) d_i \right)^2 \mid F_t \right) + \mathbb{E} \left( \left( (2k - 2) \bar{D}_2 \right)^2 \mid F_t \right) \right).
\]

(3.19)

For the first term, we use \( \mathbb{E}[d_i^2 \mid F_t] \leq (1 + \delta)^2/w \), thus

\[
\mathbb{E} \left( \left( \sum_{i=1}^{k} (4k \bar{q}_i - 4 \bar{Q}_1) d_i \right)^2 \mid F_t \right) = \sum_{i=1}^{k} (4k \bar{q}_i - 4 \bar{Q}_1)^2 \mathbb{E}[d_i^2 \mid F_t]
\]

\[
\leq \frac{(1 + \delta)^2}{w} 16k(k \bar{Q}_2 - \bar{Q}_1^2) = \frac{8(1 + \delta)^2 k}{w} \psi(t) = c \psi(t).
\]

(3.20)

As for the second term, we have

\[
1_{G_{t+1}} \mathbb{E} \left( \left( (2k - 2) \bar{D}_2 \right)^2 \mid F_t \right) \leq 4k^2 1_{G_{t+1}} \mathbb{E} \left( \sum_{i=1}^{k} (\bar{d}_i)^4 \mid F_t \right) \leq 4k^2 1_{G_{t+1}} \mathbb{E} \left( \sum_{i=1}^{k} |\bar{d}_i| \mid F_t \right)
\]

\[
\leq 1_{G_{t+1}} \frac{4k^2}{n^2} (Q_2 + \bar{Q}_1 A) \leq 1_{G_{t+1}} \frac{4k^2}{n^2} (Q_2 + a A) \leq c k,
\]

(3.21)

where in the penultimate step we used that \( \bar{Q}_1 = \sum_{i=1}^{k} a_i/w_i \leq a \), and in the last step that \( q_i \leq cn/k \) and \( a \leq cn/k \) in the event \( \{\xi(t) \in \mathcal{E}\} \subseteq G_{t+1} \). By combining Equation (3.19), Equation (3.20) and Equation (3.21) we obtain Item 2.

Finally, for Item 3,

\[
|1_{G_{t+1}} \psi(t+1) - \mathbb{E}[1_{G_{t+1}} \psi(t+1) \mid F_t]| = 1_{G_{t+1}} \left| \sum_{i=1}^{k} \sum_{j=1}^{k} (2 \bar{q}_{ij} (\bar{d}_{ij} - \mathbb{E}[\bar{d}_{ij} \mid F_t]) + (\bar{d}_{ij})^2 - \mathbb{E}[\bar{d}_{ij}^2 \mid F_t]) \right|
\]

\[
\leq 1_{G_{t+1}} \left( 2 \sum_{i=1}^{k} \sum_{j=1}^{k} \bar{q}_{ij}^2 \left( \sum_{i=1}^{k} \sum_{j=1}^{k} (\bar{d}_{ij} - \mathbb{E}[\bar{d}_{ij} \mid F_t])^2 + \sum_{i=1}^{k} \sum_{j=1}^{k} (\bar{d}_{ij}^2 + \mathbb{E}[\bar{d}_{ij}^2 \mid F_t]) \right) \right)
\]

\[
\leq 1_{G_{t+1}} (2 \sqrt{\psi(t) 6k} + 4k) \leq c_5 \sqrt{1_{G_t} \psi(t) k} + c_6 k,
\]

84
where in the first inequality we used the Cauchy-Schwarz inequality, and in the second we used that at most one agent changes its colour and thus $|\tilde{d}_{ij}| \leq 1$.

Proof of Lemma 15 and Lemma 16

The proofs of Lemma 15 and Lemma 16 are essentially the same as Lemma 18 and Lemma 19, showing that both potential functions satisfy the same properties (with different constants).

Proof of Lemma 15. Set $T = \lfloor qwn \rfloor$ where $q$ is some constant that we will determine later. Let $c_1, \ldots, c_6$ be the constants as in Lemma 18.

Define the event $B_j$ by

$$B_j = \left\{ 1_{B_s}\phi(s) \leq n^2/2^j, \forall s \in \{jT, jT + 1, \ldots, n^8\} \right\},$$

and note that $P(B_0) = 1$. Let $J = \min\{j \geq 0 : n^2/2^j \leq C'wn \log n\}$. We will prove that the event $B_J \cap B_{J-1}^c$ holds with high probability.

For any $j \in \{1, \ldots, J\}$ we have that

$$P(B_j^c) \leq 1 - P(B_j \cap B_{j-1}) = 1 - [P(B_{j-1}) - P(B_{j-1} \cap B_j^c)]
= P(B_j^c \cap B_{j-1}) + P(B_{j-1}^c).$$

For the first term on the right-hand side we have

$$P(B_j^c \cap B_{j-1}) \leq \sum_{s=jT}^{n^8} P(1_{G_s}\phi(s) > n^2/2^j),$$

where $G_s = B_s \cap \{\phi((j-1)T) \leq n^2/2^{j-1}\}$ for $s \geq (j-1)T$, and note that $G_s+1 \in \mathcal{F}_s$ for $s \geq (j-1)T$.

Now, we iterate Item 1 of Lemma 18, and choose $q$ sufficiently large such that for $T = \lfloor qwn \rfloor$ we have $(1 - c_1/(nw))^T \leq 1/8$. Then, letting $m = n^2/2^{j-1}$ we obtain for
any $t \geq jT$ that

$$
\mathbb{E}[\mathbb{I}_{G_t}\phi(t)] \leq (1 - c_1/(nw))^T \mathbb{E}[\mathbb{I}_{G_{(j-1)T}}\phi((j-1)T)] + c_2wn/c_1
$$

$$
\leq \frac{m}{8} + \frac{c_2}{c_1}wn \leq \frac{m}{4} .
$$

The last inequality holds due to the definition of $J$, and specifically that $m \geq n^2/2^{J-1} > Cwn \log n$.

Now, define the process $M(t) = \mathbb{I}_{G_t+jT}\phi(t+jT)$ for $t \geq 0$, and apply Lemma 18 and Lemma 20 with

$$
\alpha = c_1/(wn), \beta = c_2, \gamma = c_5\sqrt{mk} + c_6k, \delta^2 = c_3m/w + c_4k, \text{ and } \lambda = m/4 ,
$$

then we obtain for $t \geq 0$,

$$
P(M(t) \geq m/2) \leq P(M(t) \geq \mathbb{E}[M(0)] + m/4)
$$

$$
\leq \exp \left( -\frac{\lambda^2/2}{2\alpha - \alpha^2} + \frac{\lambda^2}{3} \right) \leq \exp \left( -c \min \left\{ \frac{m}{n}, \sqrt{\frac{m}{k}} \right\} \right), \quad (3.22)
$$

where $c$ is a small constant that only depends on $c_1, c_3, c_4, c_5$ and $c_6$. Now let $r > 0$, then by choosing $c$ sufficiently large we get

$$
\exp \left( -c \min \left\{ \frac{m}{n}, \sqrt{\frac{m}{k}} \right\} \right) \leq \frac{1}{n^r} .
$$

Therefore, for all $s \geq jT$,

$$
P(\mathbb{I}_{G_s}\phi(s) > n^2/2^j) = P(M(t) \geq m/2, t \geq 0) \leq n^{-r} .
$$

Thus we conclude that for any $j \in \{1, \ldots, J\}$ we have that

$$
P(\mathcal{B}_j \leq n^8 \mathbb{I}_{G_{(j-1)T}}\phi((j-1)T)) \leq n^{8}n^{-r} + P(\mathcal{B}_{j-1}^c)
$$

$$
\leq Jn^8n^{-r} + P(\mathcal{B}_0^c) \leq n^{10}n^{-r} ,
$$

since $P(\mathcal{B}_0) = 1$ and $J = O(\log n)$ by definition. \hfill\Box
As for the proof of Lemma 16, we follow the same steps but replacing $\mathds{1}_{B_t}\phi(t)$ by $\mathds{1}_{B'_t}\psi(t)$, and the definition of $J$ by $J = \min\{j \geq 0 : n^2/2^j \leq C'wn\log n\}$, with $C' > 32C$ (where $C$ is the constant chosen in the definition of the set of configurations $E'$ in Equation (3.10)), and the constants $\alpha, \beta, \gamma$ and $\delta$ by

$$\alpha = c_1/n, \beta = c_2, \gamma = c_3k\sqrt{m} + c_4k, \delta^2 = c_3km + c_4, \text{ and } \lambda = m/4.$$ 

With the above changes, the right-hand side of Equation (3.22) becomes

$$P(M(t) \geq m/2) \leq P(M(t) \geq E[M(0)] + m/4) \leq \exp\left(-c\min\{m/n, \sqrt{m/k}\}\right),$$

where $M(t) = \mathds{1}_{B'_t}(t + jT)$ and $t \geq 0$. Note, however, that $m/n \geq C'\log n$, so we can still choose $C'$ sufficiently large such that $\exp\left(-c\min\{m/n, \sqrt{m/k}\}\right) \leq n^{-r}$.

**Proof of Lemma 20**

We give a proof of Lemma 20, which follows a similar method as the one used in the proofs Theorems 7.3 and 7.5 of [28].

Let $s \geq 0$ to be chosen later. A simple computation shows for $t \geq 1$,

$$E[e^{sM(t)}|\mathcal{F}_{t-1}] = e^{sE[M(t)|\mathcal{F}_{t-1}]}E[e^{sM(t)} - E[e^{sM(t)}|\mathcal{F}_{t-1}]]|\mathcal{F}_{t-1}|$$

$$= e^{sE[M(t)|\mathcal{F}_{t-1}]} \sum_{j=0}^{\infty} \frac{s^j}{j!}E\left[(M(t) - E[M(t)|\mathcal{F}_{t-1}])^2|\mathcal{F}_{t-1}\right]. \tag{3.23}$$

Let $G(x) = 2\sum_{j=2}^{\infty} x^{j-2}/j!$. For $0 \leq x \leq 3$ we have that the function $G(x)$ is
increasing and that \( G(x) \leq 1/(1 - x/3) \). Then,

\[
\sum_{j=0}^{\infty} \frac{s^j}{j!} E \left[ (M(t) - E[M(t)|\mathcal{F}_{t-1}])^j | \mathcal{F}_{t-1} \right] = 1 + \sum_{j=2}^{\infty} \frac{s^j}{j!} E \left[ (M(t) - E[M(t)|\mathcal{F}_{t-1}])^j | \mathcal{F}_{t-1} \right]
\]

\[
\leq 1 + \sum_{j=2}^{\infty} \frac{s^j}{j!} \gamma^{j-2} \text{Var}(M(t)|\mathcal{F}_{t-1}) \quad \text{(by Lemma (ii))}
\]

\[
= 1 + G(s\gamma) \frac{s^2}{2} \text{Var}(M(t)|\mathcal{F}_{t-1})
\]

\[
\leq \exp \left( \frac{(s\delta)^2}{2} G(s\gamma) \right) \quad \text{(by Lemma (iii))}.
\] (3.24)

Substituting Equation (3.23) with Equation (3.24), and by Item (i) of the lemma, we get

\[
E[e^{sM(t)}|\mathcal{F}_{t-1}] \leq \exp (sE[M(t)|\mathcal{F}_{t-1}]) \cdot \exp \left( \frac{(s\delta)^2}{2} G(s\gamma) \right)
\]

\[
\leq \exp \left( s(1 - \alpha)M(t - 1) + s\beta + \frac{(s\delta)^2}{2} G(s\gamma) \right) \quad \text{(by Item (i))}.
\]

Taking expectation on both sides of the above inequality gives

\[
E[e^{sM(t)}] \leq E \left[ \exp (s(1 - \alpha)M(t - 1)) \right] \exp \left( s\beta + \frac{(s\delta)^2}{2} G(s\gamma) \right) .
\]

By iterating and using that \( G(x) \) is increasing for \( x \leq s\gamma < 3 \) (we will choose \( s \) to ensure \( s\gamma < 3 \)), we obtain that

\[
E[e^{sM(t)}] \leq E \left[ \exp (s(1 - \alpha)^2M(t - 2)) \right] \exp ([(1 - \alpha) + 1]s\beta)
\]

\[
\exp \left( [(1 - \alpha)^2 + 1] \frac{(s\delta)^2 G(s\gamma)}{2} \right)
\]

\[
\leq E \left[ \exp (s(1 - \alpha)^tM(0)) \right] \exp \left( s\beta \frac{\sum_{i=0}^{t-1} (1 - \alpha)^i + \frac{(s\delta)^2 G(s\gamma)}{2} \sum_{i=0}^{t-1} (1 - \alpha)^{2i}}{\alpha} \right)
\]

\[
= E \left[ \exp (s(1 - \alpha)^tM(0)) \right] \exp \left( s\beta \frac{1 - (1 - \alpha)^t}{\alpha} + \frac{(s\delta)^2 G(s\gamma)}{2} \frac{1 - (1 - \alpha)^{2t}}{2\alpha - \alpha^2} \right)
\]

\[
\leq \exp \left( \frac{C(s\delta)^2 G(s\gamma)}{2\alpha - \alpha^2} \right) \quad \text{(for some constant \( C \))}.
\]
Then by Markov’s inequality, we have

\[ P(M(t) \geq E[M(t)] + \lambda) = P(e^{sM(t)} \geq e^{sE[M(t)] + s\lambda}) \leq E[e^{sM(t)} e^{-s\lambda}] \]

\[ \leq \exp \left( -\lambda s + \frac{C(s\delta)^2 G(s\gamma)}{2\alpha - \alpha^2} \right) . \]

Setting \( s = \frac{\lambda}{\alpha^2/(2\alpha - \alpha^2) + \lambda \gamma/3} \), which ensures that \( s \gamma \leq 3 \) (i.e. we are working in the part where \( G \) is increasing), and using that \( G(x) \leq 1/(1 - x/3) \) for \( x \leq 3 \), we get

\[ P(M(t) \geq E[M(t)] + \lambda) \leq \exp \left( -\frac{\lambda^2/2}{\frac{\delta^2}{2\alpha - \alpha^2} + \lambda \gamma/3} \right) . \]

### 3.5.3 Proof of Section 3.3.3

The proof of Lemma 23 is rather similar to the ones of Lemma 15 and Lemma 16. The proof is a corollary of the following lemma, which is the analogue of Lemma 18 and Lemma 19 for the potential function:

\[ \sigma(t) = \frac{A(t)}{w} - a(t). \]

**Lemma 24.** Let \( T \geq 0 \) be an integer. Let \( G_t \) be an event such that \( \hat{B}_t \subseteq G_{t+1} \in \mathcal{F}_t \) for all \( t \geq T \), and such that \( G_t \subseteq G_{t+1} \). Then there exist constants \( C_1, C_2, C_3, C_4, C_5, C_6 \) (independent of \( T, k, \) etc.) such that for any \( t \geq T \)

1. \( E[1_{G_{t+1}} \sigma^2(t + 1)|\mathcal{F}_t] \leq 1_{G_t} \sigma^2(t) \left( 1 - \frac{C_1}{n} \right) + C_2 \),

2. \( \text{Var}(1_{G_{t+1}} \sigma^2(t + 1)|\mathcal{F}_t) \leq C_3 1_{G_t} \sigma^2(t) + C_4 \),

3. \( |1_{G_{t+1}} \sigma^2(t + 1) - E[1_{G_{t+1}} \sigma^2(t + 1)|\mathcal{F}_t]| \leq C_5 |\sigma(t)| + C_6 \).

With the above lemma, the proof of Lemma 23 follows exactly the same steps of the proofs of Lemma 15 and Lemma 16, and thus omitted.

**Proof of Lemma 24.** In the event \( \hat{B}_t \) we have that

\[ \sum_{i=1}^{k} \sum_{j=1}^{k} \left( \frac{A_i(t)}{w_i} - \frac{A_j(t)}{w_j} \right)^2 \leq \hat{C} wn \log n . \]
Dividing by $2k^2$ on both sides of the above inequality and rearranging the terms, we get

$$
\frac{1}{2k^2} \sum_{i=1}^{k} \sum_{j=1}^{k} \left( \frac{A_i(t)}{w_i} - \frac{A_j(t)}{w_j} \right)^2 = \frac{1}{k} \sum_{i=1}^{k} \left( \frac{A_i(t)}{w_i} \right)^2 - \left( \frac{1}{k} \sum_{i=1}^{k} \frac{A_i(t)}{w_i} \right)^2 \\
= \frac{1}{k} \sum_{i=1}^{k} \left( \frac{A_i(t)}{w_i} - X(t) \right)^2 \\
\leq \frac{C_w n \log n}{2k^2},
$$

where $X(t) = \frac{1}{k} \sum_{i=1}^{k} \frac{A_i(t)}{w_i}$. Notice that

$$
|A(t) - wX(t)| \leq \left| \sum_{i=1}^{k} (A_i(t) - w_i X(t)) \right| = \left| \sum_{i=1}^{k} \left( \frac{A_i(t)}{w_i} - X(t) \right) w_i \right| \\
\leq \sqrt{\sum_{i=1}^{k} \left( \frac{A_i(t)}{w_i} - X(t) \right)^2} \sqrt{\sum_{i=1}^{k} w_i^2} \\
\leq w \sqrt{\frac{C_w n \log n}{2k}} = c_1 \sqrt{n \log n},
$$

for some constant $c_1 > 0$ (this constant depends on the values of $w$ and $k$, which are also assumed to be constants).

We also have that

$$
\sum_{i=1}^{k} \frac{A_i(t)^2}{w_i} - wX(t)^2 = \sum_{i=1}^{k} \left( \frac{A_i(t)^2}{w_i^2} - X(t)^2 \right) w_i \\
\leq \sqrt{\sum_{i=1}^{k} \left( \frac{A_i(t)}{w_i} - X(t) \right)^2} \sqrt{\sum_{i=1}^{k} \left( \frac{A_i(t)}{w_i} + X(t) \right)^2} w_i^2 \\
\leq \sqrt{\frac{C_w n \log n}{2k}} \sqrt{n^2 \left( \sum_{i=1}^{k} w_i \right)^2} \\
\leq c_2 n^{3/2} \sqrt{\log n},
$$

where the last but one step is due to $A_i(t) \leq n$ and $X(t) \leq n$. Thus, combining the two inequalities gives

$$
\left| \sum_{i=1}^{k} \frac{A_i(t)^2}{w_i} - \frac{A(t)^2}{w} \right| \leq c_3 n^{3/2} \sqrt{\log n}.
$$

(3.25)
CHAPTER 3. DIVERSITY, FAIRNESS, AND SUSTAINABILITY IN POPULATION PROTOCOLS

Since $\sigma(t) = A(t)/w - a(t)$, then

$$
\sigma(t + 1) = \frac{A(t + 1)}{w} - a(t + 1) = \sigma(t) + \frac{\delta_A}{w} - \delta_a = \sigma(t) + \left(1 + \frac{1}{w}\right)\delta_A,
$$

where $\delta_A$ and $\delta_a$ are the changes of dark and light colours in one step respectively, and $\delta_A + \delta_a = 0$. Therefore, as at most one agent changes its colour in one step, we obtain that

$$
|\sigma(t + 1) - \sigma(t)| \leq 1 + \frac{1}{w}.
$$

Note that the probability that a dark-coloured agent changes its colour to light at time $t$ is given by $\sum_{i=1}^{k} \frac{A_i(A_i(t) - 1)}{n(n-1)w_i}$, and the probability that it changes from light to dark is $\frac{a(t)A(t)}{n(n-1)}$. Therefore, denoted by $A = A(t)$ and $a = a(t)$, we have

$$
E[\delta_A] = \frac{aA}{n(n-1)} - \sum_{i=1}^{k} \frac{A_i(A_i - 1)}{n(n-1)w_i} = \frac{1}{n(n-1)} \left(aA - \sum_{i=1}^{k} \frac{A_i^2 - A_i}{w_i}\right) \\
\leq \frac{1}{n(n-1)} \left(aA - \sum_{i=1}^{k} \frac{A_i^2}{w_i} + A\right).
$$

Thus, by the above computations and the bound given in Equation (3.25), we get

$$
E \left[ \mathbb{1}_{G_{t+1}} \sigma(t + 1)^2 | \mathcal{F}_t \right] = E \left[ \mathbb{1}_{G_{t+1}} \left( \sigma(t) + (1 + 1/w)\delta_A \right)^2 | \mathcal{F}_t \right] \\
= \mathbb{1}_{G_{t+1}} \left( \sigma(t)^2 + 2\sigma(t) \left(1 + \frac{1}{w}\right) E[\delta_A] + \left(1 + \frac{1}{w}\right)^2 E[\delta_A^2] \right) \\
\leq \mathbb{1}_{G_{t+1}} \left( \sigma(t)^2 + 2\sigma(t) \frac{1 + 1/w}{n(n-1)} \left(a(t)A(t) - \sum_{i=1}^{k} \frac{A_i(t)^2}{w_i} + A(t)\right) \right) + 4 \\
\leq \mathbb{1}_{G_{t+1}} \left( \sigma(t)^2 + c_4 \frac{\sigma(t)}{n^2} \left(a(t)A(t) - \frac{A(t)^2}{w} + c_3n^{3/2}\sqrt{\log n} + c_5n\right) \right) + 4 \\
= \mathbb{1}_{G_{t+1}} \sigma(t)^2 \left(1 - \frac{c_4}{n^2} A(t) + \frac{c_6\sqrt{\log n}}{\sigma\sqrt{n}}\right) + 4 \quad \text{(since } \sigma(t) = A(t)/w - a(t),) \\
= \mathbb{1}_{G_{t+1}} \sigma(t)^2 \left(1 - \frac{c_4}{n}\right) + c_2,
$$

where in the last step we used that in the event $\widehat{B}_t$ (which is contained in $G_{t+1}$) we have that $A(t) \geq cn$ for some constant $c > 0$. 

91
For the variance, notice that

\[
\text{Var}(1_{G_{t+1}}(\sigma(t+1) + \sigma(t)2|\mathcal{F}_t) \\
= 1_{G_{t+1}} \left[ \text{Var}(2\sigma(t)(\sigma(t+1) - \sigma(t))|\mathcal{F}_t) + \text{Var}((\sigma(t+1) - \sigma(t))^2|\mathcal{F}_t) \right] \\
+ \text{Cov}(2\sigma(t)(\sigma(t+1) - \sigma(t)), (\sigma(t+1) - \sigma(t))^2)] \\
\leq \frac{3}{2} 1_{G_{t+1}} \left[ 4\sigma(t)^2 \text{Var}(\sigma(t+1) - \sigma(t)|\mathcal{F}_t) + \text{Var}((\sigma(t+1) - \sigma(t))^2|\mathcal{F}_t) \right] \\
\leq 6\sigma(t)^2 1_{G_{t+1}} \text{Var}(\sigma(t+1) - \sigma(t)|\mathcal{F}_t) + \frac{3}{2} 1_{G_{t+1}} \text{Var}((\sigma(t+1) - \sigma(t))^2|\mathcal{F}_t) ,
\]

where in the first inequality we used that \(2\text{Cov}(X, Y) \leq \text{Var}(X) + \text{Var}(Y)\) for any pair of random variables \(\{X, Y\}\). Furthermore, note that \(|\sigma(t+1) - \sigma(t)| \leq 1+1/w \leq 2\) as at most only one agent changes its colour from a light one to dark one, or vice versa. Then,

\[
\text{Var}(1_{G_{t+1}}(\sigma(t+1) + \sigma(t)2|\mathcal{F}_t) \leq C_3 1_{G_{t}} \sigma^2(t) + C_4 .
\]

Finally, since \(|\sigma(t+1) - \sigma(t)| \leq 2\), we have

\[
|E[1_{G_{t+1}}(\sigma(t+1) + \sigma(t)2|\mathcal{F}_t) - 1_{G_{t+1}}(\sigma(t+1) + \sigma(t)2]| \\
= 1_{G_{t+1}} \left| E\left[ (\sigma(t) + \sigma(t+1) - \sigma(t))^2|\mathcal{F}_t \right] - (\sigma(t) + \sigma(t+1) - \sigma(t))^2 \right| \\
\leq 1_{G_{t+1}} \left| E\left[ 2\sigma(t)(\sigma(t+1) - \sigma(t)) + (\sigma(t+1) - \sigma(t))^2|\mathcal{F}_t \right] \\
+ 1_{G_{t+1}} \left| 2\sigma(t)(\sigma(t+1) - \sigma(t)) + (\sigma(t+1) - \sigma(t))^2 \right| \\
\leq 1_{G_{t}} C_5 |\sigma(t)| + C_6 .
\]

\[\square\]

### 3.6 Auxiliary results

**Theorem 25** (Chapter XIV.2, XIV.3 in [53]). Let \(p \in (0, 1) \setminus \{1/2\}\) and \(b, s \in \mathbb{N}\).

Consider a discrete time Markov chain \((Z_t)_{t \geq 0}\) with state space \(\Omega = [0, b]\) where

- \(Z_0 = s \in [0, b]\)


92
Let $T = \min\{t \geq 0 \mid Z_t \in \{0, b\}\}$. Then,

$$P(Z_T = b) = \frac{\left(\frac{1-p}{p}\right)^s - 1}{\left(\frac{1-p}{p}\right)^b - 1}, \quad \text{and} \quad P(Z_T = 0) = \frac{\left(\frac{1-p}{p}\right)^b - \left(\frac{1-p}{p}\right)^s}{\left(\frac{1-p}{p}\right)^b - 1}.$$  

Moreover,

$$\mathbb{E}[T] = \frac{s}{1-2p} - \frac{b}{1-2p} \left(1 - \left(\frac{1-p}{p}\right)^s\right).$$  

The following Chernoff bound for Markov Chains is from [29], however, we simplified their general statement for the purposes of our proof. Let $M$ be a Markov chain with transition matrix $P$ and stationary distribution $\pi$. The $\varepsilon$-mixing time $T_{\text{mix}}$ of $M$, where $0 < \varepsilon < 1$, is defined as the first time such that the total variation distance from stationarity is at most $\varepsilon$, i.e.

$$T_{\text{mix}}(\varepsilon) = \min\{t : \max_x ||P^t(x, \cdot) - \pi||_{TV} \leq \varepsilon\}.$$  

**Theorem 26 ([29]).** Let $M$ be an ergodic Markov Chain with stationary distribution $\pi$ on a finite state space. Let $T_{\text{mix}}$ represent the $(1/8)$-mixing time. Let $N_i$ denote the number of times that $M$ hits the state $i$ in the first $t$ steps, then for $0 < \delta < 1$, it holds that

$$P(|N_i - \pi(i)t| \geq \delta \pi(i)t) \leq c \exp \left(-\delta^2 \pi(i)t/(72T_{\text{mix}})\right),$$  

where $c > 0$ is a constant independent of $\delta, \pi$, and $t$.  

93
Chapter 4

Triangle-Switch Processes on 2–Regular Graphs

4.1 Introduction

There has been a substantial amount of interest in the processes and algorithms of generating random graphs in graph theory and theoretical computer science. Researchers are usually interested in those random graphs generation models or protocols that favour graphs with few triangles, in which each pair of vertices has a mutual neighbour in the third vertex. For instance, an efficient algorithm is introduced for generating random simple graphs with large girth in [11]. Note that the girth of an undirected graph is the length of its shortest cycle if there is any. For any undirected graph with no cycles, which is a forest, its girth is defined to be infinite. Therefore, if a graph has girth 4, it is triangle-free. Particularly, generating random graphs with a given degree sequence has been one of the central topics and has been extensively studied (see [21, 27, 102]). In the work of [19], a faster algorithm that improves the best known running time is analysed to generate random bipartite graphs with a general degree distribution. A sequential importance sampling (SIS) method is discussed to produce bipartite graphs with a given degree sequence in [26].
However, in the application of modelling social networks, triangles play an important role in structuring the process of mutual friendship formation, as in any set of three persons containing friendship between two pairs of people will tend to form the third according to [60]. In other words, one’s friend’s friend is likely to become one’s friend, thus creating a relational triangle. Furthermore, in the work of [70], the authors modelled the evolution of the structure of social networks over time and showed that within communities it contains a high density of triangles of friendship. Moreover, triangles are of great significance in citation networks as they generate many triangles. See [103, 24] for empirical study on the number of triangle formations of evolving models of citation networks.

Since lots of attention has been attracted for simulating real world networks such as social networks and citation networks, where triangles are an intriguing structure that needs to be considered, we study two random processes in this chapter using local edge transformations known as switches, which show tendency to produce many triangles.

A switch is a local edge transformation in a graph and is a probabilistic approach for generating simple graphs with fixed nodes and fixed degree sequences. It was proved to be irreducible (all graphs in the space can be obtained with this transformation) by [80, 98]. In a switch, a pair of distinct edges \(xy\) and \(wz\) of graph \(G\) are chosen u.a.r.\(^1\) and replaced with a uniformly chosen perfect matching of the vertices \(\{x, y, w, z\}\). (We use the notation \(xy\) as a shorthand for the edge \(\{x, y\}\).) If the resulting graph \(G'\) is not simple then the move is rejected. See Figure 4.1 for an example of a switch move. Clearly, a switch move preserves the degree of each node but does not necessarily preserve connectivity.

The mixing time of switch Markov chains has been extensively studied for sampling the realisations of prescribed degree sequences of simple graphs. With different mechanisms, several results have been proved on the rapid mixing nature of the switch

\(^1\)We use ‘u.a.r.’ to mean ‘uniformly at random’ for brevity
CHAPTER 4. TRIANGLE-SWITCH PROCESSES ON 2-REGULAR GRAPHS

Figure 4.1: A switch, in which edges $xy$ and $wz$ are replaced by $xw$ and $yz$.

Markov chain on graphs with different types of degree sequences, such as bipartite, regular and usual degree sequences.

Kannan, Tetali, and Vempala [73] proposed a simple Markov chain Monte Carlo (MCMC) method for uniformly sampling realisations of a given degree sequence. Their algorithm is based on a series of randomly chosen switch transformations, which ‘swap’ a pair of nonadjacent edges at each time-step to generate a random walk on the family of all possible realisations. They conjectured that the process is rapidly mixing. In other words, if starting from an arbitrary realisation of the degree sequence, the process reaches a random realisation (any possible graph with the degree sequence) after polynomial number of steps (polynomial in the number of nodes). However, they only analysed the mixing time of the switch Markov chain in the bipartite case and proved that it is rapidly mixing. Cooper, Dyer, and Greenhill [31] provided a rapid mixing proof of this switch Markov chain for generating arbitrary regular undirected graphs using a canonical path argument; and subsequently Greenhill [61] extended the result of [31] and proved an analogous result for the regular directed case. Later, Miklós, Erdős and Soukup [52] showed that the switch Markov chain is rapid mixing for semi-regular bipartite graphs where nodes have the same degree on only one side of the bipartition; and recently Erdős, Miklós and Toroczkai [51] proved the rapid mixing for the almost semi-regular case.

Allen-Zhu et al. [7] considered the local edge-switches on $d$-regular graphs for $d = \Omega(\log n)$. They proved that such a dynamic converges to an algebraic expander, in the
sense that the non-zero eigenvalues of the Laplacian matrix is $\Omega(d)$, in $O(n^2d^2\sqrt{\log n})$ steps of the switch chain with high probability. While in our work we analyse the spectral properties of the Markov chain defined by the switch transformation, their analysis is based on tracking the evolution of the eigenvalues of the graph over time by a potential function.

By further restricting to a subset of simple switches, we can ensure that every switch will change the set of triangles in the graph. A triangle-switch is a graph transformation based on edge switches which is used to make or break triangles. The triangle-switch is illustrated in Figure 4.2. When looking at the figure from left to right the triangle is broken or merged; and when looking from right to left the triangle is made or formed. Cooper, Dyer, and Greenhill [32] considered triangle-switch moves on 3-regular (cubic) graphs and proved that any Markov chain that performs random triangle-switches is irreducible. For two particular chains they also gave an estimation on the number of triangles in the long run.

![Figure 4.2: A triangle-switch.](image)

In this chapter we investigate several variants of the triangle-switch process on a class of $n$-vertex 2-regular graphs $\mathcal{G}(n, 2)$. A graph $G$ in the set $\mathcal{G}(n, 2)$ consists of a union of disjoint cycles, and is not connected by triangle-switches, which can only alter the cycle lengths by multiples of 3. So, for example, a graph on $n = 12$ vertices with 3 cycles of length 4 cannot be transformed into 4 cycles of length 3. To avoid this, we restrict our attention to the simplest case, the subset $\mathcal{G}^+(n, 2)$ of graphs in which all cycles have length $3i$, and $n = 3k$ for $i, k \in \mathbb{N}^+$. 97
CHAPTER 4. TRIANGLE-SWITCH PROCESSES ON 2-REGULAR GRAPHS

Considering a graph \( G \) of \( n = 3k \) vertices for \( k \in \mathbb{N}^+ \), which contains only cycles of length divisible by 3, i.e., \( G = \bigcup \{C_{3i}\} \) for \( i \in \{1, \ldots, k\} \), we introduce two main triangle-switch protocols: basic triangle-switch process and alternative triangle-switch process, that create or remove triangles (if there exists one) at each step. The basic process is defined below in the next section, and the alternative process in Section 4.3.

4.2 Basic triangle-switch process on \( \mathcal{G}^*(n, 2) \)

Basic triangle-switch rules. Let \( G \in \mathcal{G}^*(n, 2) \) be a graph consisting of cycles, such that the length of each cycle is divisible by 3. The simplest case of the basic process is as follows. In each time-step, choose a vertex \( v \) uniformly at random. Let \( a, b \) be the neighbours of vertex \( v \). Provided the resulting graph is simple, apply the basic triangle-switch process as follows to transform the graph \( G \).

A: (Make triangles) If \( a \) and \( b \) are not connected by an edge, then these three vertices \( a, b \) and \( v \) are on a cycle with length greater than 3. Suppose the other neighbours of \( a \) and \( b \) besides \( v \) are \( c \) and \( d \) respectively. If the result is graphic (i.e., \( v \) is on a cycle of length at least 6), switch the edges \( ac \) and \( bd \). The switch \( \{ac, bd\} \rightarrow \{ab, cd\} \) is shown in Figure 4.2 when we look at the two subfigures from right to left.

B: (Break triangles) If there is an edge between the two neighbours \( a \) and \( b \) of \( v \), in which case these three vertices form a triangle \( abv \), we choose a random edge \( cd \in E(G) \) whose endpoints are not \( a, b \) or \( v \), and then switch the two edges \( ab \) and \( cd \) by swapping one end of both edges. The switch \( \{ab, cd\} \rightarrow \{ac, bd\} \) is shown in Figure 4.2 when we look at the two subfigures from left to right. Formally, to remove ambiguity, we can choose a random orientation of \( ab \) and \( cd \) for the switch.

Thus, in the case that vertices \( a, b \) and \( v \) form a triangle, we break this triangle and merge it into a random cycle \( C_{3j} \) with \( j \geq 1 \) by applying the triangle-switch
process and obtain a larger cycle \( C_{3j+3} \); while in the case that \( a, b \) and \( v \) are on a cycle \( C_{3i} \) with \( i \geq 2 \), we separate a triangle from the previous cycle \( C_{3i} \) to obtain \( C_{3i-3} \) and \( C_3 \) forming by vertices \( a, b \) and \( v \).

The number of triangles is considered as an important quantity to measure, which helps understand some of the patterns in the structure of the graphs. In [10] and [101], the authors have looked at the density of triangles (ratio of triangles in connected triples) in a network as the network transitivity. In the next four sub-sections we consider three variants of the basic triangle-switch process and analyse the expected time for triangles to converge.

1. The vertex \( v \) is chosen uniformly at random (u.a.r. process). (See Section 4.2.1.)
2. The moves A:(make triangles) and B:(break triangles) are chosen with probability \( p \) and \( q \) respectively. (See Section 4.2.2.)
3. The u.a.r. process is extended to include ignoring cycles of length 4 and 5 when encountered. (See Section 4.2.3.)
4. The expected number of cycles and the time for triangles to converge in the u.a.r process. (See Section 4.2.4.)

A typical graph generated by the basic u.a.r. triangle-switch process on \( n = 300 \) vertices is shown in Figure 4.3.

### 4.2.1 Basic triangle-switch, u.a.r. variant

**Recurrences for the expected number of triangles.** Let \( G \) be a graph of \( n \) vertices, where \( n = 3k \) for \( k \in \mathbb{N}^+ \), consisting of simple cycles of length divisible by 3. Denote by \( N_j(t) \) the number of vertices on cycle of length \( 3j \) at time \( t \) for \( 1 \leq j \leq k \). Then for any \( t \geq 0 \) we have \( \sum_{j=1}^{k} N_j(t) = n \).

To compute the expected value of \( N_j \) at time \( t + 1 \), we consider three cases in the proof. The first case is \( j = 1 \), the next \( 2 \leq j \leq k - 1 \), as breaking a \( C_6 \) gives two triangles and merging a triangle into another triangle results in the loss of two
CHAPTER 4. TRIANGLE-SWITCH PROCESSES ON 2–REGULAR GRAPHS

Figure 4.3: A typical output of the basic u.a.r. triangle-switch on the subset $G^*$ of 2–regular graphs. Here $n = 300$ vertices, and the simulation was run for 10,000 steps.

$C_3$’s; while for other cycles with length larger than 3 there will be only one triangle involved. The final case is $j = k$. We obtain the expectations for step $t+1$ conditional on the values at step $t$. Denote by $\mathcal{F}_t$ the filtration generated by the values of $N_j$ up to step $t$.

For $j = 1$, the expected number of vertices $N_1$ on $C_3$ at time $t + 1$ consists of: 1) vertices on $C_3$ at time $t$; 2) vertices of two new triangles generated by breaking a $C_6$ at time $t$; 3) vertices of one new triangle if a $C_{3j}$ with $j \geq 3$ is broken at time $t$; 4) the loss of two triangles if a $C_3$ is chosen to merge with another $C_3$ at time $t$; 5) the loss of one triangle if a $C_3$ is merged with a $C_{3j}$ with $2 \leq j \leq k - 1$ at time $t$.

Therefore, we have the following recurrence according to the above reasoning, and
we use \( \sum_{j \geq 1} N_j = n \) in the computation.

\[
E[N_1(t+1)|\mathcal{F}_t] = N_1(t) + 6 \frac{N_2(t)}{n} + 3 \sum_{i=3}^{k} \frac{N_i(t)}{n} - 6 \frac{N_1(t)}{n} \frac{N_1(t)}{n} - 3 \frac{N_1(t)}{n} \sum_{i=2}^{k-1} \frac{N_i(t)}{n} \\
= N_1(t) + 6 \frac{N_2(t)}{n} + 3 \left( 1 - \frac{N_1(t)}{n} - \frac{N_2(t)}{n} \right) - 6 \left( \frac{N_1(t)}{n} \right)^2 \\
- 3 \frac{N_1(t)}{n} \left( 1 - \frac{N_1(t)}{n} - \frac{N_1(t)}{n} \right) \\
= N_1(t) + 3 \frac{N_2(t)}{n} + 3 - 6 \frac{N_1(t)}{n} - 3 \left( \frac{N_1(t)}{n} \right)^2 + 3 \frac{N_1(t) N_k(t)}{n^2} .
\] (4.1)

As for the case \( 2 \leq j \leq k-1 \), the expected number of vertices \( N_j \) on \( C_{3j} \) at time \( t+1 \) is made up of 1) the vertices \( N_j \) at time \( t \); and 2) the vertices of a new \( C_{3j} \) generated by either separating a triangle from a \( C_{3(j+1)} \) at time \( t \), or 3) merging a triangle with a \( C_{3(j-1)} \) at time \( t \); and 4) the loss of vertices of a \( C_{3j} \) due to either breaking a \( C_{3j} \) at time \( t \), or 5) merging a triangle with a \( C_{3j} \) at time \( t \). In each case, \( 3j \) vertices are added to or removed from \( N_j(t) \). Thus, for any \( 2 \leq j \leq k-1 \), the recurrence of the expected number of vertices on \( C_{3j} \) at time \( t+1 \) is:

\[
E[N_j(t+1)|\mathcal{F}_t] = N_j(t) + 3 \frac{N_{j+1}(t)}{n} + 3j \frac{N_1(t) N_{j-1}(t)}{n} - 3j \frac{N_j(t)}{n} - 3j \frac{N_1(t) N_j(t)}{n} .
\] (4.2)

The case \( j = k = n/3 \) is special, as \( N_k \) can only increase if a \( C_{3(k-1)} \) merges with a triangle and decrease if a triangle is separated from a \( C_{3k} \). Therefore,

\[
E[N_k(t+1)|\mathcal{F}_t] = N_k(t) + 3k \frac{N_1(t) N_{k-1}(t)}{n} - 3k \frac{N_k(t)}{n} .
\] (4.3)

The last line follows because \( N_k = 0 \) identically for \( n < 3k \), and if \( N_k = 1 \) then \( N_1 = 0 \).

**Equivalent asymptotic system.** Taking expectations again to remove the conditioning on \((N_1, N_2, ..., N_k)\) and letting \( n \to \infty \), we get the following asymptotic
system.
\[
\begin{align*}
\mathbb{E}N_1(t+1) &= \mathbb{E}N_1(t) + \frac{3\mathbb{E}N_2(t)}{n} + 3 - 6\frac{\mathbb{E}N_1(t)}{n} - 3\frac{\mathbb{E}[N_1(t)^2]}{n^2} + 3\frac{\mathbb{E}[N_1(t)N_k(t)]}{n^2} \\
\mathbb{E}N_j(t+1) &= \mathbb{E}N_j(t) + \frac{3j\mathbb{E}N_{j+1}(t)}{n} + 3\frac{j\mathbb{E}[N_1(t)N_{j-1}(t)]}{n^2} - 3\frac{j\mathbb{E}N_j(t)}{n} \\
&\quad - 3\frac{j\mathbb{E}[N_1(t)N_j(t)]}{n^2} \\
\mathbb{E}N_k(t+1) &= \mathbb{E}N_k(t) + \frac{3k\mathbb{E}[N_1(t)N_{k-1}(t)]}{n^2} - 3\frac{k\mathbb{E}N_k(t)}{n} \\
\end{align*}
\] (4.4)

Quadratic expectations, a problem. At this point we are stuck as we cannot evaluate the quadratic terms such as \(\mathbb{E}[N_1(t)^2]\), \(\mathbb{E}[N_1(t)N_j(t)]\) etc.

Intuitively, in the long run, the expected number of vertices on any cycle \(C_{3j}\) for \(j \geq 1\) will reach some stationary value and will not change on continuing the triangle-switch process. In other words, \(\mathbb{E}N_j(t+1) \sim \mathbb{E}N_j(t)\) for any \(j \geq 1\), i.e., in the limit \(\mathbb{E}N_1(t+1) - \mathbb{E}N_1(t) = 0\). Similarly the values of the variables \(N_j(t)\) should become concentrated and thus \(\mathbb{E}N_iN_j \sim \mathbb{E}N_i\mathbb{E}N_j\).

Deterministic asymptotic system. We replace the random variables \(N_j(t)\) in Recurrences (4.1)-(4.3) by equivalent deterministic ones. Divide by \(n\) to give \(P_j(t) = \frac{N_j(t)}{n}\) and let \(n \to \infty\). In the limiting system equivalent to the (4.4) system, we will have \(P_j(t+1) = P_j(t) = P_j\), basically cancelling \(\mathbb{E}N_1(t+1)\) with \(\mathbb{E}N_1(t)\). The equation for \(N_k(t)\) disappears as \(k \to \infty\). Thus, cancelling the \(3j\) for \(j \geq 1\), we have
\[
\begin{align*}
0 &= P_2 + 1 - 2P_1 - P_1^2 + P_1P_k , \\
0 &= P_{j+1} + P_1P_{j-1} - P_j - P_1P_j , \quad \text{for } 2 \leq j \leq k-1 , \\
0 &= P_1P_{k-1} - P_k .
\end{align*}
\] (4.5)-(4.7)

By summing up the above recurrences, we can cancel out all terms consisting of \(P_i\) for \(i \in \{2, \ldots, k-1\}\), leaving only terms of \(P_1\) and \(P_k\). Therefore,
\[
0 = 1 - 2P_1 + P_1P_k \quad \Rightarrow \quad P_1 = \frac{1}{2 - P_k} \quad \text{or} \quad P_k = 2 - \frac{1}{P_1} .
\] (4.8)
CHAPTER 4. TRIANGLE-SWITCH PROCESSES ON 2-REGULAR GRAPHS

Notice that $0 \leq P_k = 2 - 1/P_1 \leq 1$ gives $1/2 \leq P_1 \leq 1$. We calculate $P_j$ by considering two cases: $P_1 > 1/2$ and $P_1 = 1/2$, and demonstrate that $P_1 = 1/2$ by contradiction. As for the first case we suppose $P_1 > 1/2$, then by Recurrence (4.5) we have $P_2 > 1/4$. Iterating in Recurrence (4.6) gives $P_j > 1/(2^j)$, and by Recurrence (4.7) we obtain $P_k > 1/(2^k)$. Therefore, the sum of these probabilities will be $\sum_{j \geq 1} P_j > \sum_{j \geq 1} 1/(2^j) > 1$. However, note that $\sum_{j \geq 1} N_j / n = 1$ gives $\sum_{j \geq 1} P_j = 1$, which is the contradiction as required. We next check the validity of the case $P_1 = 1/2$. By substituting for Recurrence (4.5) and iterating in Recurrence (4.6) we know that $P_j = 1/(2^j)$, and $P_k = 1/(2^k)$ by Recurrence (4.7). Therefore $P_k \to 0$ for $k = \Omega(\log n)$, in which case, we confirm $P_1 \to 1/2$ by Recurrence (4.8). Hence, we conclude that for any $j \in \{1, \ldots, k\}$ we have

$$P_j = \frac{1}{2^j}.$$

As $\sum_{j \geq 1} P_j = 1$, the system is normalised to a deterministic stationary process. This suggests that the expected number of vertices on cycle of length $C_3j$ is $N_j \to n/2^j$ for $j \in \{1, \ldots, k\}$ in the long run.

**Conjecture 27.** Given a graph $G$ with $n = 3k$ vertices consisting of cycles of length $3j$, where $1 \leq j \leq k$, if the u.a.r. basic triangle-switch process is applied to the graph space $G^\ast(n, 2)$, then for sufficiently large $k$, the expected number of vertices $N_j$ on cycle of length $C_3j$ for $1 \leq j \leq k$ tends to $n/2^j$.

Below we give an edited simulation output for the basic triangle-switch process on $n = 30,000$, which helps confirm this. The simulation was written in R and run for 1,000,000 steps.
4.2.2 Variant $P$(make $C_3$) = $p$, $P$(break $C_3$) = $q$

Let $G$ be a graph consisting of only cycles, where the length of each cycle is divisible by 3. In each time-step, following the triangle-switch rule, we create a triangle with probability $p$ according to Rule A:(make triangles) and break one with probability $q$ according to Rule B:(break triangles).

Conjecture 28. Consider a graph $G$ with $n$ vertices consisting of only cycles $C_{3j}$, where $1 \leq j \leq k$ and $k = n/3$. If the triangle-switch process is applied to the graph such that a triangle is made with probability $p$ and broken with probability $q$, then after the process reaches a steady state the probability of picking a vertex $v$ which is on any cycle $C_{3j}$ for $1 \leq j \leq k$ tends to

$$ P(v \in C_{3j}) = \frac{p \cdot q^{j-1}}{(p + q)^j} . $$

Proof. Let $N_j(t)$ be the number of vertices on cycle of length $3j$ at time $t$ where $1 \leq j \leq k$ and $k = n/3$. Then for any $t \geq 0$ we have $\sum_{j=1}^{k} N_j = n$. We compute the
expected number of \( N_j \) at time \( t + 1 \) in the following three cases. For simplicity we denote \( N_j(t + 1) \) by \( N'_j \) and \( N_j(t) \) by \( N_j \). The recurrences are:

\[
\begin{align*}
\mathbb{E}N'_1 &= N_1 + 6 \frac{N_2}{n} p + 3 \sum_{i=3}^{k} \frac{N_i}{n} p - 6 \frac{N_1}{n} \frac{N_j}{n} q - 3 \frac{N_1}{n} \sum_{i=2}^{k-1} \frac{N_i}{n} q, \\
\mathbb{E}N'_j &= N_j + 3j \frac{N_{j+1}}{n} p + 3j \frac{N_{j-1}}{n} q - 3j \frac{N_j}{n} p - 3j \frac{N_1}{n} \frac{N_j}{n} q, \quad 2 \leq j \leq k - 1, \\
\mathbb{E}N'_k &= N_k + 3k \frac{N_1}{n} \frac{N_{k-1}}{n} q - 3k \frac{N_k}{n} p.
\end{align*}
\]

Let \( P_j = N_j/n \) be the probability of picking a vertex that on a cycle of length \( 3j \) at time \( t \) for \( j \geq 1 \). (Recall that \( N_j \) is the simplified notation of \( N_j(t) \).) Since the expected number of vertices on any cycle \( C_{3j} \) is settled at the steady state and will not change if continuing the triangle-switch process and notice that \( \sum_{j \geq 1} P_j = 1 \), then the above recurrences can be expressed by

\[
\begin{align*}
0 &= p - (p + q)P_1 + pP_2 - qP_1^2 + qP_1P_k, \\
0 &= pP_{j+1} + qP_1P_{j-1} - pP_j - qP_1P_j, \\
0 &= qP_1P_{k-1} - pP_k.
\end{align*}
\]

We derive an equation for \( P_1 \) and \( P_k \) by summing up the above recurrences where all other terms \( P_i \) for \( i \in \{2, \ldots, k - 1\} \) are cancelled out.

\[
0 = p - (p + q)P_1 + qP_1P_k \quad \Rightarrow \quad P_1 = \frac{p}{p + q - qP_k}.
\]

The recurrences can be solved with the same reasoning as when we demonstrated Conjecture 27 in section 4.2.1. We obtain that for \( j \geq 1 \),

\[
P_j = \frac{p \cdot q^{j-1}}{(p + q)^j}.
\]

\[
\square
\]

### 4.2.3 Variant u.a.r. triangle-switch, ignore \( C_4 \) and \( C_5 \)

In this subsection, we extend the u.a.r triangle-switch process to include cycles of any length (not necessarily divisible by 3 as in section 4.2.1) and ignore \( C_4 \) and \( C_5 \).
when encountered in the process. Let \( G \) be a graph consisting of a group of cycles \( \cup \{C_x\} \) for \( x \geq 3 \). In each time-step, we choose a random vertex \( v \) whose neighbours are \( a, b \in N(v) \),

- if \( v, a, b \) are not on a cycle of length either 4 or 5 we apply the triangle-switch rules A:(make triangles) and B:(break triangles);
- otherwise, we keep choosing another random vertex and check its neighbours until they are not on a \( C_4 \) or \( C_5 \), and then apply the triangle-switch rules.

**Conjecture 29.** Let \( G \) be a graph with \( n \) vertices consisting of a group of cycles \( C_x \), where \( x \geq 3 \). Denote by \( P_x \) the probability of picking a vertex that belongs to a cycle of length \( x \). If the triangle-switch process is applied to the graph, then after the process reaches a steady state the probability \( P_x \) tends to

\[
P_{3j} = \left[ \frac{1}{2} (1 - P_4 - P_5) \right]^j, \ j \geq 1,
\]

\[
P_{3j+1} = P_4 \cdot \left[ \frac{1}{2} (1 - P_4 - P_5) \right]^{j-1}, \ j \geq 2,
\]

\[
P_{3j+2} = P_5 \cdot \left[ \frac{1}{2} (1 - P_4 - P_5) \right]^{j-1}, \ j \geq 2.
\]

**Proof.** Let \( N_x(t) \) be the number of vertices on cycles of length \( x \) at time \( t \) where \( x \geq 3 \). Then for any \( t \geq 0 \) we have \( \sum_{x \geq 3} N_x = n \). Denote \( N_x(t + 1) \) by \( N'_x \) and \( N_x(t) \) by \( N_x \) for simplicity. Note that \( C_{n-2} \) and \( C_{n-1} \) cannot exist during the process, otherwise there will be two vertices or a single vertex left in the graph. We compute
CHAPTER 4. TRIANGLE-SWITCH PROCESSES ON 2-REGULAR GRAPHS

the recurrences of the expected number of $N_x$ at time $t + 1$ in the following cases.

$$E_{N_3'} = N_3 + 6 \frac{N_6}{n} + 3 \sum_{i=7}^{n} \frac{N_i}{n} - 6 \frac{N_3}{n} \frac{N_3}{n} - 3 \frac{N_3}{n} \sum_{i=4}^{n-3} \frac{N_i}{n},$$

$$E_{N_4'} = N_4 + 4 \frac{N_7}{n} - 4 \frac{N_3}{n} \frac{N_4}{n},$$

$$E_{N_5'} = N_5 + 5 \frac{N_8}{n} - 5 \frac{N_3}{n} \frac{N_5}{n},$$

$$E_{N_x'} = N_x + x \frac{N_{x+3}}{n} + x \frac{N_3}{n} \frac{N_{x-3}}{n} - x \frac{N_x}{n} - x \frac{N_3}{n} \frac{N_x}{n}, \quad x = \{6, \ldots, n-6\} \cup \{n-3\},$$

$$E_{N_{n-5}'} = N_{n-5} + (n-5) \frac{N_3}{n} \frac{N_{n-8}}{n} - (n-5) \frac{N_{n-5}}{n},$$

$$E_{N_{n-4}'} = N_{n-4} + (n-4) \frac{N_3}{n} \frac{N_{n-7}}{n} - (n-4) \frac{N_{n-4}}{n},$$

$$E_{N_n'} = N_n + n \frac{N_3}{n} \frac{N_{n-3}}{n} - n \frac{N_n}{n}. \quad (4.9)$$

Let $P_x = N_x/n$ be the probability of picking a vertex that on a cycle of length $x$ at time $t$ for $x \geq 3$. Since the expected number of vertices on cycles of any length is settled at the steady state and will not change if continuing the triangle-switch process and notice that $\sum_{x \geq 3} P_x = 1$, then the above recurrences become

$$0 = 1 - 2P_3 - P_3^2 - P_4 - P_5 + P_6 + P_3P_n ,$$

$$0 = P_7 - P_3 \cdot P_4 ,$$

$$0 = P_8 - P_3 \cdot P_5 ,$$

$$0 = P_{x+3} + P_3P_{x-3} - P_x - P_3P_x , \quad x = \{6, \ldots, n-6\} \cup \{n-3\},$$

$$0 = P_3P_{n-8} - P_{n-5} ,$$

$$0 = P_3P_{n-7} - P_{n-4} ,$$

$$0 = P_3P_{n-3} - P_n . \quad (4.9)$$

We sum up all recurrences in (4.9) so that all terms containing $P_i$ for $i = \{6, \ldots, n-1\}$ cancell out, which gives

$$P_3 = \frac{1 - P_4 - P_5}{2 - P_n} \rightarrow \frac{1 - P_4 - P_5}{2} .$$

107
Replacing Recurrences (4.9) with $P_3 = (1 - P_4 - P_5)/2$, we can solve all $P_x$ by iteration. Therefore,

\[
P_{3j} = \left[ \frac{1}{2} (1 - P_4 - P_5) \right]^j, \quad j \geq 1, \\
P_{3j+1} = P_4 \cdot \left[ \frac{1}{2} (1 - P_4 - P_5) \right]^{j-1}, \quad j \geq 2, \\
P_{3j+2} = P_5 \cdot \left[ \frac{1}{2} (1 - P_4 - P_5) \right]^{j-1}, \quad j \geq 2.
\]

\[
\square
\]

4.2.4 Triangles and the total number of cycles, u.a.r variant

An equation for drift. Denote by $X_j(t)$ the number of cycles of length $3j$ at time-step $t$. Let $C(t)$ be the sum of the number of all cycles of lengths $3j$ at time-step $t$ for $j \geq 1$, such that $C(t) = \sum_{j \geq 1} X_j(t)$. The following lemma shows a correlation between the drift of the number of cycles and the change of the number of cycles in the basic u.a.r triangle-switch process (see section 4.2.1 for the rules).

Lemma 30. Let $G$ be a graph of $n$ vertices consisting of cycles $C_{3j}$, for $j = \{1, \ldots, k\}$ and $k = n/3$. If the basic u.a.r. triangle-switch process is applied to the graph, then

\[
\text{EC}(t+1) = C(t) + 1 - \frac{6X_1(t)}{n}.
\]

The above lemma indicates that in one step the expected number of cycles increases if $X_1(t) < n/6$; and decreases if $X_1(t) > n/6$. In order to prove the lemma, we will analyse the relation between the drift in the number of cycles and the change of the number of cycles in one step.

Proof. Recall that $C(t) = X_1(t) + \sum_{j \geq 2} X_j(t)$. According to the triangle-switch rules, there are four cases in one step depending on whether a chosen node $u$ belongs to a triangle or a cycle of length larger than 3. Suppose a node $u \in V(G)$ is chosen, then
CHAPTER 4. TRIANGLE-SWITCH PROCESSES ON 2-REGULAR GRAPHS

1) Node $u$ is on a $C_6$:

\[ X_1(t + 1) = X_1(t) + 2, \quad \sum_{j \geq 2} X_j(t + 1) = \sum_{j \geq 2} X_j(t) - 1 \]

\[ \implies C(t + 1) = C(t) + 1. \]

2) Node $u$ is on a $C_{3j}$ with $j \geq 3$:

\[ X_1(t + 1) = X_1(t) + 1, \quad \sum_{j \geq 2} X_j(t + 1) = \sum_{j \geq 2} X_j(t) + 1 - 1 \]

\[ \implies C(t + 1) = C(t) + 1. \]

3) Node $u$ is on a triangle and an edge from another triangle is chosen:

\[ X_1(t + 1) = X_1(t) - 2, \quad \sum_{j \geq 2} X_j(t + 1) = \sum_{j \geq 2} X_j(t) + 1 \]

\[ \implies C(t + 1) = C(t) - 1. \]

4) Node $u$ is on a triangle and an edge from a $C_{3j}$ with $j \geq 2$ is chosen:

\[ X_1(t + 1) = X_1(t) - 1, \quad \sum_{j \geq 2} X_j(t + 1) = \sum_{j \geq 2} X_j(t) + 1 - 1 \]

\[ \implies C(t + 1) = C(t) - 1. \]

In cases 1) and 2), the probability of choosing a vertex $u$ which is not on a triangle is $1 - 3X_1(t)/n$. In cases 3) and 4), the probability of choosing a vertex $u$ which is on a triangle is $3X_1(t)/n$. Thus

\[ \mathbb{E}C(t + 1) = C(t) + \left( 1 - \frac{3X_1(t)}{n} \right) \cdot (+1) + \frac{3X_1(t)}{n} \cdot (-1), \]

completing the proof of the lemma.

In summary, the only way for the number of all cycles to increase by one is if a triangle is ‘broken off’ a cycle of length larger than 3; and the only way for the number of all cycles to decrease by one is if a triangle is ‘merged with’ another cycle (including a triangle). Thus, the drift up and down in the number of all cycles corresponds directly to the increase or decrease in the number of triangles.
CHAPTER 4. TRIANGLE-SWITCH PROCESSES ON 2-REGULAR GRAPHS

Time to converge for triangles. From Lemma 30 we know that $E[C(t + 1) | X_1(t)] = C(t) + 1 - 6X_1(t)/n$. If $X_1(t) < n/6$ there is an expected increase in the number of cycles, and $X_1(t) > n/6$ an expected decrease. We will investigate this drift of the total number of cycles to show that the number of triangles converges in the long run.

Lemma 31. Let $G$ be a graph of $n$ vertices consisting of cycles $C_{3j}$, with $j \in \{1, \ldots, k\}$ and $k = n/3$. If the basic u.a.r. triangle-switch process is applied to $G$, then after $O(n/\varepsilon)$ steps, where $0 < \varepsilon < 1$, with high probability, the number of triangles $C_3$ is $(1 + O(\varepsilon))n/6$, irrespective of the initial value of $C_3$.

Proof. Denote by $\Delta_t$ the drift of the total number of cycles in one step, then by Lemma 30,
$$E[\Delta_t] = E[C(t + 1) - C(t)] = 1 - \frac{6X_1(t)}{n}. \quad (4.10)$$

Assume for now that $X_1(t) < (1 - \varepsilon)n/6 = X_1^*$ for some sequence of steps $t \leq \tau$. For example if $X_1(0) = (1 - \delta)n/6$ where $\delta > \varepsilon$, then $\tau \geq (\delta - \varepsilon)n/6$. Let $p(t) = 1 - 3X_1(t)/n$ be the probability that the total number of cycles increases by one in one step, and $q(t) = 3X_1(t)/n$ the probability that it decreases by one, then $p(t) - q(t) \geq \varepsilon$. Let $p = (1 + \varepsilon)/2$, and $q = (1 - \varepsilon)/2$. Then as long as $X_1(t) \leq X_1^*$, we have $p(t) \geq p$. An alternative way to compute a lower bound on the expected drift in one step is:

$$E[\Delta_t] = P(C(t + 1) = C(t) + 1) \times 1 + P(C(t + 1) = C(t) - 1) \times (-1)$$
$$= p(t) - q(t) \geq p - q = \varepsilon.$$

From now on we work with the lower bound on the drift towards $X_1^*$, in which case the expectation of the total drift $\Delta = \sum_{t=1}^{T} \Delta_t$ of the number of all cycles in $T$ steps is:

$$E[\Delta] \geq (p - q) \cdot T = \varepsilon T.$$
The values $\Delta_1, \ldots, \Delta_T$ are independent bounded random variables, with $\Delta_i \in [-1, 1]$ for any $i \in \{1, \ldots, T\}$, and $E\Delta_i \geq p - q = \varepsilon$. Applying the Hoeffding bound gives that for $\delta > 0$:

$$P(\Delta \leq E\Delta - T\delta) \leq P(\Delta \leq T\varepsilon - T\delta) \leq \exp\left(-\frac{2T\delta^2}{(1+1)^2}\right)$$

$$\Rightarrow P(\Delta \leq (\varepsilon - \delta)T) \leq \exp\left(-\frac{1}{2}T\delta^2\right)$$

$$\Rightarrow P(\Delta \leq (1 - \frac{\delta}{\varepsilon})E\Delta) \leq \exp\left(-\frac{1}{2}T\delta^2\right).$$

If we take $\delta = \varepsilon/2$, then the above bound becomes

$$P\left(\Delta \leq \frac{1}{2}E\Delta\right) \leq \exp\left(-\frac{1}{8}T\varepsilon^2\right).$$

In order to get the above inequality, we set $\varepsilon^2 T/8 = \varepsilon n/12$, which gives $T = 2n/(3\varepsilon)$. Thus, when $t \geq T = 2n/(3\varepsilon)$ we have $P(\Delta \leq E\Delta/2) \leq \exp(-\varepsilon n/12)$ w.h.p..

In the following part, we shall prove by contradiction to demonstrate that the number of triangles will approach $n/6$ in the long run, and as long as it gets close to $n/6$, it is not likely to go below a small range around $n/6$.

From Equation (4.10), we notice that $E\Delta_t$ increases if $X_1(t) < n/6$ and decreases if $X_1(t) > n/6$. Suppose $X_1(t) < (1 - \varepsilon)n/6$ for any $t \leq T$. By Equation (4.10),

$$E\Delta_t > 1 - \frac{6}{n}(1 - \varepsilon)\frac{n}{6} = \varepsilon,$$

$$\Rightarrow E\Delta = E\left[\sum_{t=1}^{T} \Delta_t\right] > \varepsilon T.$$

Since $P(\Delta \leq E\Delta/2) \to 0$, we have

$$\Delta > \frac{1}{2}E\Delta > \frac{1}{2}\varepsilon T = \frac{1}{2}\varepsilon \cdot \frac{2n}{3\varepsilon} = \frac{n}{3},$$

which is the contradiction as required since the maximum number of cycles is $n/3$. Hence, $\exists t \leq T$ such that $X_1 \geq (1 - \varepsilon)n/6$. 

111
Suppose for any $t \geq T$, we have $X_1(t) > (1 + \varepsilon)n/6$. Then, by Equation (4.10):

$$E\Delta_t < 1 - \frac{6}{n}(1 + \varepsilon)\frac{n}{6} = -\varepsilon,$$

$$\implies E\Delta < -\varepsilon T'.$$

Applying Chernoff-Hoeffding bound gives:

$$P\left(\Delta > \frac{3}{2}E\Delta\right) \leq \exp\left\{-\frac{1}{8}\varepsilon^2 T\right\} = \exp\left\{-\frac{\varepsilon n}{12}\right\} \to 0.$$

Therefore,

$$\Delta \leq \frac{3}{2}E\Delta < -\frac{3}{2}\varepsilon T = -\frac{3}{2}\varepsilon\frac{2n}{3\varepsilon} = -n.$$

This is the required contraction as $|\Delta| < n/3$, and thus $\exists t \leq T$, such that $X_1(t) < (1 + \varepsilon)n/6$.

Hence, for any $t \geq T = 2n/(3\varepsilon)$, the number of triangles $X_1(t) \in [(1-\varepsilon)n/6, (1+\varepsilon)n/6]$. In other words, $X_1$ will hit the range $[(1-\varepsilon)n/6, (1+\varepsilon)n/6]$ after any interval of $T$.

**Remark 32.** In some ways it looks as if, when we apply the ideas of the lemma to decrease the number of triangles to $n/6$ from above, there is a catch. Consider a graph with $n/2$ vertices on triangles, and $n/2$ vertices on hexagons. Initially the probability that we pick a hexagon and create two triangles is $1/2$. The expected number of triangles lost is $(-1)(1/2)(1/2) + (-2)(1/2)(1/2)$, with the first case being to pick a hexagon and a triangle, the second to pick two triangles. Thus the drift per step is $2(1/2) - (1/4) - (1/2) = 1/4$, and the number of triangles must increase for a linear number of steps.

Indeed, this must be the case for the first $\Omega(n)$ steps, but after this the number of triangles will drift back towards $n/6$. Lemma 31 needs $O(n/\varepsilon)$ steps for convergence within $\varepsilon n/6$ of the long run value. This is $O(n^2)$ starting from $n/6 + 1$ triangles.

A typical plot of the number of triangles, starting with $n/2$ vertices on triangles and $n/2$ vertices on hexagons, is shown below in Figure 4.4. The number of triangles
increases for the first 1,000 steps and then drifts back to fluctuate around \( n/6 = 5,000 \) (here \( n = 3k = 30,000 \)).

Figure 4.4: The number of triangles during 100,000 simulation steps for \( k = 10,000 \), sampled every 100 steps. The starting mix was \( n/2 \) vertices on triangles, and \( n/2 \) on hexagons.

### 4.3 Linear triangle-switch process, an alternative triangle process

In this section, we study an alternative triangle-switch process. Different from the basic triangle-switch, in which case the recurrences of the expected number of triangles contain quadratic terms and difficult to be evaluated, this alternative process gives linear recurrences of the expected number of cycles and allows us to have an insight of the mixing time and other properties.
Let $G$ be a graph of $n$ vertices consisting of cycles, each of which has length divisible by 3. In each step, we choose a vertex $v \in V(G)$ uniformly at random whose neighbours are $a, b \in N(v)$, and apply a triangle-switch process as follows:

A: With probability $p$ we make a triangle-increasing switch at $v$. If $v$ is in a triangle we do nothing, otherwise we ‘snip’ a triangle off from the cycle at $v$. Suppose that $v$ is on a cycle of length greater than 3, and that the other neighbours of $a$ and $b$ besides $v$ are $c$ and $d$ respectively. We switch the edges $ac$ and $bd$ by swapping one of the endpoints of both edges as shown in Figure 4.5.

B: With probability $q = 1 - p$ we make a triangle-reducing switch at $v$. If there is no triangle left we do nothing, otherwise we ‘merge’ a triangle with the cycle at $v$. We switch an edge $ef$ of a random triangle $\triangle ef g$ in the graph with an edge (say $va$) of the cycle at $v$ as shown in Figure 4.6. In the case that $v$ is in a triangle, this swapping merges two triangles into a hexagon.

![Figure 4.5: Rule A: make a triangle with probability $p$.](image)

![Figure 4.6: Rule B: merge a triangle with probability $q$.](image)
If we let \( n/3 = k \), the state space for the triangle-switch process corresponds to integer partitions of \( k \). We give an example for \( n = 12, k = 4 \). On the left are the partitions of 4, and on the right are the possible numbers of cycles length \( 3j \), \( j = 1, 2, 3, 4 \), and the number of vertices on cycles of that size.

| 1+1+1+1 | (4,0,0,0) | (12,0,0,0) |
| 1+1+2  | (2,1,0,0) | (6,6,0,0) |
| 1+3    | (1,0,1,0) | (3,0,9,0) |
| 2+2    | (0,2,0,0) | (0,12,0,12) |
| 4      | (0,0,0,1) | (0,0,0,12) |

Let \( N_j(t) \) be the number of vertices on cycles of length \( 3j \) at step \( t \), where \( j \in \{1, \ldots, k\} \). On the assumption there are enough triangles (we are in a stable phase), we can write down the conditional recurrences for \( N_j(t+1) \). For simplicity, denote \( N_j = N_j(t) \) and \( N'_j = N_j(t+1) \). Let \( k = n/3 \) and \( \rho = \rho(t) = P(N_1 > 0) \).

\[
\begin{align*}
E N'_1 &= N_1 + 3p \sum_{i=3}^{k} \frac{N_i}{n} + 6p \frac{N_2}{n} - 3\rho q \sum_{i=2}^{k-1} \frac{N_i}{n} - 6\rho q \frac{N_1}{n} \\
E N'_j &= N_j + 3j \rho q N_{j-1} + 3j \rho q N_j - 3j \rho q N_j + 3j \rho q N_j, \; j = \{2, \ldots, k-1\} \\
E N'_k &= N_k + 3k \rho q N_{k-1} - 3k \rho q N_k, \; k = \frac{n}{3}
\end{align*}
\]

(4.11)

Provided \( N_1 \gg 0 \), and assuming \( \rho(t) = 1 - o(1) \) (for sufficiently large \( t \)) we can consider the deterministic system given by (4.12), which should be a close approximation.

\[
\begin{align*}
N'_1 &= \left( 1 - \frac{3}{n} \right) N_1 + \frac{3p}{n} N_2 + \frac{3q}{n} N_k + 3(p - q) \\
N'_j &= \left( 1 - \frac{3j}{n} \right) N_j + \frac{3j\rho q}{n} N_{j-1} + \frac{3j\rho p}{n} N_{j+1}, \; j = \{2, \ldots, k-1\} \\
N'_k &= (1 - p) N_k + q N_{k-1}, \; k = \frac{n}{3}
\end{align*}
\]

(4.12)

We briefly outline the justification for assuming \( \rho(t) = 1 - o(1) \). If \( p > q \), and \( n \to \infty \) it will be the case (see Equation (4.14)) that the equilibrium solution for \( N_1 \)
tends to $N_1 = n(p - q)/p$. From here it would take around $(1/q)^{N_1}$ steps to reduce the number of triangles to zero, which helps justify this approximation.

We do not formally justify this claim for reasons we next explain. Intuitively, the number of triangles $C_1$ in the process is similar to a random walk on the line $\{0, 1, ..., k\}$, with position dependent transitions of size $-2, -1, 0, 1, 2$, and with loops at either end; at $C_1 = 0$ with probability $q$, and at $C_1 = k$ with probability $p$. To justify the claim, it seems we need to study a walk on the full state space $S = \{(N_1, ..., N_k)\}$, rather than a simple random walk on the line $\{0, 1, ..., k\}$.

In vector notation, the system (4.12) becomes $\mathbf{N}(t + 1) = A \cdot \mathbf{N}(t)$ where $\mathbf{N} = (N_j : j = 1, ..., k)$ is a column vector, and the matrix $A$ is given by

$$A = \begin{pmatrix}
1 - \frac{3}{n} + 3\frac{p-q}{n} & 3\frac{p}{n} + 3\frac{p-q}{n} & 3\frac{p-q}{n} & \cdots & 3\frac{p-q}{n} & 3\frac{p}{n} \\
6\frac{2}{n} & 1 - \frac{6}{n} & 6\frac{p}{n} & 0 & \cdots & 0 \\
0 & \frac{9q}{n} & 1 - \frac{9}{n} & \frac{2p}{n} & 0 & \cdots & 0 \\
0 & 0 & \frac{12q}{n} & 1 - \frac{12}{n} & \cdots & 0 & \\
\vdots & \vdots & \vdots & \ddots & \ddots & \ddots & \vdots & \\
\vdots & \vdots & \vdots & \ddots & \ddots & \ddots & \ddots & \\
0 & \cdots & 1 - \frac{3i-1}{n} & 3\frac{p_i-1}{n} & 0 & \cdots & 0 \\
0 & \cdots & 3q\frac{i}{n} & 1 - \frac{3i}{n} & 3\frac{p_i}{n} & \cdots & 0 \\
0 & \cdots & 0 & 3q\frac{i+1}{n} & 1 - \frac{3i+1}{n} & \cdots & 0 \\
\vdots & \vdots & \vdots & \ddots & \ddots & \ddots & \ddots & \\
\vdots & \vdots & \vdots & \ddots & \ddots & \ddots & \ddots & \\
0 & 0 & \cdots & \cdots & 0 & q & q
\end{pmatrix}
$$

(4.13)
The row sums of matrix $A$ are:

1st row $= 1 - \frac{3}{n} + \frac{3p}{n} + \frac{3q}{n} + 3(p - q) = 4 - 6q$,

$i$-th row $= 1 - \frac{3i}{n} + \frac{3ip}{n} + \frac{3iq}{n} = 1$, for $i = \{2, \ldots, k-1\}$,

$k$-th row $= 2q$.

The column sums of matrix $A$ all add up to one:

1st column $= 1 - \frac{3}{n} + \frac{3(p - q)}{n} + \frac{6q}{n} = 1$,

$i$-th column $= 1 - \frac{3i}{n} + \frac{3(p - q)}{n} + \frac{3(i-1)p}{n} + \frac{3(i+1)q}{n} = 1$, for $i = \{2, \ldots, k-1\}$,

$k$-th column $= \frac{3p}{n} + \frac{3(k-1)p}{n} + q = 1$.

Note that all column sums of matrix $A$ (4.13) are 1, therefore matrix $A$ is column-stochastic, and it is doubly stochastic if $p = q = 1/2$. For $p \geq q$, all entries are non-negative, and the largest row sum is 1.. This confirms that 1 is an upper bound on the spectral radius, since that the maximum row sum of a non-negative matrix is an upper bound on its spectral radius (see Horn and Johnson, Matrix Analysis [65], 8.1.22).

Assuming $p \geq q$, the entries are all non-negative and the column sums add to one. Thus $A^\top$ is row stochastic, and is the transition matrix of a random walk on some edge weighted digraph. An example of the digraph for $p = q$ is given in Figure 4.7. The matrix $A$ is aperiodic (loop at $N_1, p$), irreducible as can get from any state to any other, and thus has a unique stationary distribution.

We consider two cases $p > q$ and $p = q$, and discuss the properties of matrix $A$ in the following subsections. In subsection 4.3.1, we give an explicit solution to the recurrences of the number of vertices on cycles of length $3j$ for the $p > q$ case. As for the case when $p = q$, we discuss certain spectral properties in subsection 4.3.2, and investigate the mixing time of the associated random walk in subsections 4.3.3 and 4.3.4. In subsection 4.3.5, we will show the simulations of the process on the convergence.
4.3.1 Case $p > q$

The Recurrences (4.12) of $N_j$ have an explicit solution (4.14) for $j = \{1, \ldots, k\}$. This is obtained starting from $N_k$ and working backwards.

$$N_j = \begin{pmatrix} \left(\frac{q}{p}\right)^{j-1} \frac{p-q}{p-q\left(\frac{q}{p}\right)^{k-1}} \end{pmatrix} n .$$ (4.14)

**Remark 33.** On the assumption that $\rho$ is time invariant, the exact Recurrences (4.11) have the following solution, where $\rho$ is unknown,

$$N_j = \begin{pmatrix} \left(\frac{\rho q}{p}\right)^{j-1} \frac{p-\rho q}{p-q\left(\frac{\rho q}{p}\right)^{k-1}} \end{pmatrix} n .$$ (4.15)

For $p > 1/2$, considering solutions to $A \mathbf{y} = \lambda \mathbf{y}$, with $\mathbf{y} = (N_j : j = 1, \ldots, k)$ as given by Equation (4.14), we see that it is the principal eigenvector with eigenvalue $\lambda = 1$. The left eigenvector of this is $\mathbf{x}^\top A = \mathbf{x}^\top$ where $\mathbf{x} = (1, 1, \ldots, 1)$. Thus $A^\top$ has stationary distribution $\pi = ((q/p)^{j-1} (p-q)/(p-q(q/p)^{k-1}) : j = 1, \ldots, k)$.

Iterating the system (4.12) as $\mathbf{N}(t+1) = A \cdot \mathbf{N}(t)$ we get $\mathbf{N}(t) = A^t \cdot \mathbf{N}(0)$, whose limiting solution is $n\pi$ as given in Equation (4.14).

An interesting question is the rate of convergence of this system. As the system corresponds to a weighted digraph, it is not obvious how to approach this question. In Section 4.3.5 we investigate the convergence experimentally, and in Section 4.3.3 we derive a theoretical upper bound for the special case $p = q$.

4.3.2 Case $p = q$

If $p = q = 1/2$, the Recurrences (4.12) of $N_j$ become:

$$\begin{cases} 
N'_1 &= \left(1 - \frac{3}{n}\right) N_1 + \frac{3}{2n} N_2 + \frac{3}{2n} N_k \\
N'_j &= \left(1 - \frac{3j}{n}\right) N_j + \frac{3j}{2n} N_{j-1} + \frac{3j}{2n} N_{j+1}, \quad j = \{2, \ldots, k-1\} \\
N'_k &= \frac{1}{2} N_k + \frac{1}{2} N_{k-1}, \quad k = \frac{n}{3}
\end{cases}$$ (4.16)

118
For a long term steady state, a possible solution for $N_i$ based on $N'_i = N_i$ is:

$$N_i = \frac{n}{k} = 3, \text{ for } i = \{1, \ldots, k\}.$$ 

Let $X_i$ be the number of cycles with length $3i$ for $i \in \{1, \ldots, k\}$. Thus the expected number of cycles of length $3i$ in equilibrium is $EX_i = N_i/3i = 1/i$, and as the inverse proportion to cycle size. However this solution ignores the fact that we assumed for sufficiently large $t$, $\rho(t) \sim 1$, and that triangle-reducing moves were never rejected. As $X_1 = 1$, there is only one triangle in expectation in the long run, it must be that for a constant proportion of moves $\rho(t) < 1$, and the assumption $\rho \sim 1$ is not valid.

Consider the system (4.16), where matrix $A$ of its vector notation is given in (4.17). Matrix $A$ is doubly stochastic (rows and columns both add to one) with a very simple form - nearly tri-diagonal, and has an interesting structure. We will further investigate its properties, and the convergence of the random walks associated with $A$ as well as its transpose.
The following summarises the discussion so far.

**Theorem 34.** In the case of \( p = q = 1/2 \), matrix \( A \) given by (4.17) is irreducible, aperiodic and doubly stochastic. The principal eigenvalue of \( A \) is 1, with principal eigenvector \((1/k)\), the stationary distribution of both \( A \) and \( A^\top \).

**Proof.** Since there is a path in \( A \) from any state to any other state, it follows that \( A \) is irreducible and aperiodic. Matrix \( A \) is doubly stochastic as both row and column sums of matrix \( A \) are 1.

It is demonstrated in Lemma 35 and Lemma 37 that the principal eigenvalue of matrix \( A \) is 1, and \( A \) has only one eigenvalue of maximum modulus. For the eigenvector \( \mathbf{x} = (1/k, 1/k, \ldots, 1/k)^\top \) corresponding to eigenvalue 1, we have \( \mathbf{x}^\top A = \mathbf{x}^\top \) and \( A^\top \mathbf{x} = \mathbf{x} \), thus \( \mathbf{x} = (1/k, 1/k, \ldots, 1/k)^\top \) is the stationary distribution of both \( A \) and \( A^\top \).

**Lemma 35.** In the case of \( p = q = 1/2 \), matrix \( A \) given in (4.17) has principal eigenvalue 1.

**Proof.** Let \( \lambda_1 \geq \lambda_2, \ldots, \lambda_k \) be the eigenvalues of \( A \) listed in decreasing order. Notice that 1 is an eigenvalue of matrix \( A \) given in (4.17) corresponding to the eigenvector \((1/k, 1/k, \ldots, 1/k)^\top \). We prove by contradiction that 1 is the largest eigenvalue, and thus \( \lambda_1 = 1 \). Suppose \( A = (a_{ij}) \) and \( A\mathbf{x} = \lambda\mathbf{x} \) for some \( \lambda > 1 \). Since all entries of matrix \( A \) are nonnegative and each row sums to 1, then for any \( i \in \{1, \ldots, k\} \) we have \( \sum_j a_{ij} = 1 \) and \( a_{ij} \geq 0 \). This implies that each arbitrary element \( (A\mathbf{x})_i = \sum_j a_{ij}x_j \) is a convex combination of the components of \( \mathbf{x} \), which cannot be greater than the largest component of \( \mathbf{x} \), say \( x_{\max} \). However, at least one element of \( \lambda\mathbf{x} \) is greater than \( x_{\max} \) since \( \lambda > 1 \). This is the contradiction as required.

**Corollary 36.** Suppose that matrix \( A \) is irreducible and nonnegative. If \( A \) has \( \ell > 1 \) many eigenvalues of maximum modulus, then every main diagonal entry of \( A \) is zero.
Proof. According to Horn & Johnson [65] Corollary 8.4.6, the irreducible and non-negative matrix $A$ is similar to $e^{i\phi} \cdot A$, where $\phi = 2\pi/\ell$, and thus has the same eigenvalues. Hence, if the trace of $A$ is real, we have

$$
\text{tr}(A) = e^{i\phi} \cdot \text{tr}(A) = e^{2\pi i/\ell} \cdot \text{tr}(A) = (-1)^{2/\ell} \cdot \text{tr}(A)
$$

$$
\implies \ell = 1 \text{ or } \text{tr}(A) = 0.
$$

\hfill \square

Lemma 37. In the case of $p = q = 1/2$, 1 is the unique eigenvalue of maximum modulus of matrix $A$ given in (4.17).

Proof. Since the main diagonal entries of $A$ are positive, and Corollary 36 says if there is more than one eigenvalue of max modulus then the main diagonal is all zeroes, then we can ensure that 1 is the only eigenvalue of maximum modulus, i.e.,

$$
\max_{i \geq 2} |\lambda_i| < 1,
$$

where $\lambda_i$ is any other eigenvalue of $A$. \hfill \square

4.3.3 An investigation of the spectral gap and mixing time of matrix $A$ in the case $p = q$

Preliminary remarks. The matrix $A$ given in (4.17) is not symmetric but its structure is rather simple, and it is tri-diagonal with the exception of the entry $A_{1,k} = 3/2n$. Tri-diagonal matrices can be brought into symmetric form.

The matrices $A$ and $A^\top$ can be seen as transition matrices of weighted random walks and the underlying digraph $D$ of $A^\top$ is shown in Figure 4.7. With the exception of the edge $(k,1)$, for every edge $(i,j)$ there is a corresponding edge $(j,i)$. As the underlying graph of the digraph $D$ is a cycle, one might expect a second eigenvalue of size something like $1 - \Theta(1/k^2)$. In the following section, we use a method from Fiedler [54] to lower bound the absolute spectral gap $\min |1 - \lambda_j|$, where $j \geq 2$, by $(1 - o(1))27\pi^2/n^3$. 

121
From Lemma 37, we know that the spectral radius $\rho(A) = \lambda_1 = 1$ is the unique eigenvalue of maximum modulus. Unfortunately $A$ cannot be made symmetric, the problem being $A_{1,k} = 3/2n$. This directed edge is $(k, 1)$ in the diagram of Figure 4.7. Thus some eigenvalues of $A$ may be complex, and occur in conjugate pairs. This occurs for $A$ with $n \geq 18$, as can be checked by direct computation.

The eigenvalue separation in Lemma 37 suggests some possibility of bounding the convergence time of the walks associated with $A$. Although convergence to stationarity for random walks no longer follows directly from the spectral gap, we can obtain estimates for the mixing time of the random walk represented by $A$ based on a method of J. Fill [55]. This method uses the symmetrised matrix $M = A^t A$. Interestingly the absolute second eigenvalue of $M$ is $\sqrt{1 - \frac{9}{2n^2}}$, but we haven’t found a way to map this back to $A$.

**Fiedler’s bound for $|1-\lambda|$**. In what follows, Fiedler’s bound, the absolute spectral gap $|1-\lambda|$ refers to the absolute value of a possibly complex number. Thus $|a+ib| = \sqrt{a^2 + b^2}$.
\[ \sqrt{a^2 + b^2} \]. A non-stochastic eigenvalue, is any eigenvalue other than the one which achieves the spectral radius. The following lemma gives a bound on \(|1 - \lambda|\).

**Lemma 38.** In the case of \( p = q = 1/2 \), the distance between 1 and any non-stochastic eigenvalue of \( A \) is lower bounded by \( \frac{27\pi^2}{n^3} (1 - o(1)) \).

**Proof.** Let \( \lambda_1 = 1, \lambda_2, \lambda_3, \ldots \) be the eigenvalues of matrix \( A \). Following the same method as in Fiedler [54] Theorem 3.4, we give a bound on the distance between any non-stochastic eigenvalue and 1. Denote \( N = \{1, 2, \ldots, k\} \), and let \( W \subset N \) be a non-void proper subset of \( N \). If \( A \) is a doubly stochastic \( k \times k \) matrix where \( k = n/3 \), then for any \( i \geq 2 \),

\[ |1 - \lambda_i| \geq \Psi_k [\mu(A)] , \]

where \( \Psi_k(x) = 2(1 - \cos \frac{\pi}{k})x \), for \( 0 \leq x \leq \frac{1}{2} \), and \( \mu(A) = \min_{\emptyset \neq W \subset N} \sum_{i \in W, j \not\in W} a_{ij} \).

By Taylor expansion, we obtain

\[ \cos \frac{\pi}{k} = 1 - \frac{\pi^2}{2k^2} + \frac{\pi^4}{4!k^4} - \ldots = 1 - \frac{\pi^2}{2k^2} + O \left( \frac{1}{k^4} \right) . \]

The minimum sum of the entries of some sub-matrix of \( A \) is:

\[ \mu(A) = \min_{\emptyset \neq W \subset N} \sum_{i \in W, j \not\in W} a_{ij} \geq \frac{3}{n} , \quad (\text{equality is satisfied when } W = \{1\}) . \]

Thus,

\[ |1 - \lambda_i| \geq 2 \left[ 1 - \left( 1 - \frac{\pi^2}{2k^2} + O \left( \frac{1}{k^4} \right) \right) \right] \frac{3}{n} = \left[ \frac{\pi^2}{k^2} - O \left( \frac{1}{k^4} \right) \right] \frac{3}{n} = \frac{27\pi^2}{n^3} - O \left( \frac{1}{n^5} \right) . \]

\[ \square \]

### 4.3.4 A theorem of Fill, and its application to the mixing time of \( A \) in the case \( p = q \)

We use an approach given by Fill [55] to estimate the spectral gap of the multiplicative reversibilisation of the transition matrix \( A^\top \), where matrix \( A \) is given in (4.17). This gives us the following result.
**Theorem 39.** Let \( P = A^\top \). The stationary distribution is given by \( \pi_P = \pi_P \cdot P \). For \( i = 1, \ldots, k \), \( \pi_P(i) = 1/k \). The following holds for \( t \geq 0 \),

\[
4 ||P^t(i, \cdot) - \pi_P||^2 \leq \left( 1 - \frac{9}{2n^2} \right)^t \cdot \frac{n}{3} \tag{4.18}
\]

From the variation distance given in the above theorem, we have the following conclusion.

**Corollary 40.** In \( T = O(n^2 \log n) \) steps, the distribution of the Markov chain with transition matrix \( P = A^\top \) is \( \pi(1 + o(1)) \), and thus close to stationarity. It follows that the system given by (4.16), with iteration \( N(s) = A^s \cdot N(0) \) at step \( s \), has value \( N(T) \sim n\pi \), where \( N_i = n/k = 3 \) for \( i \in \{1, \ldots, k\} \).

In order to prove Theorem 39, we will introduce a theorem from Fill [55] which establishes a bound on the variation distance of \( P = A^\top \) from stationary in terms of the second eigenvalue of \( M = P \cdot \tilde{P} \), where \( \tilde{P} \) is the time reversal of \( P \); and demonstrate how to estimate the second eigenvalue of \( M \) using the Bottleneck ratio property, which relates the mixing time of a chain to the spectral gap of the transition matrix.

**The theorem of Fill (1991).** Let \( P \) be a transition matrix such that \( P = A^\top \), where \( \pi_P \) is the stationary distribution of \( P \) such that \( \pi_P = \pi_P \cdot P \). Denote by \( \tilde{P}(x, y) = \frac{\pi_P(y) \cdot P(y, x)}{\pi_P(x)} \) the time reversal of \( P \). Then, \( \tilde{P}(x, y) = P(y, x) = A^\top(y, x) = A(x, y) \), and \( \tilde{P} \) has stationary distribution \( \pi_{\tilde{P}} = 1/k = \pi_P \). Define \( M = P \cdot \tilde{P} = A^\top \cdot A \).

**Theorem 41.** (Fill, 1991) Let \( P \) be an ergodic transition matrix on a finite state space \( \Omega \) with stationary distribution \( \pi \). Denote by \( \lambda_2(M) \) the second largest eigenvalue of the matrix \( M = P \cdot \tilde{P} \), where \( \tilde{P} \) is the time reversal of \( P \). Then for any \( i \in \Omega \),

\[
4 ||P^t(i, \cdot) - \pi||^2 \leq \left( \frac{\lambda_2(M)^t}{\pi(i)} \right) \cdot \frac{n}{3} \tag{4.19}
\]
The matrix $M = A^T \cdot A$. Note that the entries $M(i, j)$ of $M$ can be computed as follows, for $i, j \in \{1, \ldots, k\}$:

$$M(1, k) = M(k, 1) = \frac{3}{2n} \left( 1 - \frac{3}{n} \right)$$

$$M(i - 2, i) = \left[ \frac{3(i - 1)}{2n} \right]^2$$

$$M(i - 1, i) = \left[ 1 - \frac{3(i - 1)}{n} \right] \frac{3(i - 1)}{2n} + \left( 1 - \frac{3i}{n} \right) \frac{3i}{2n}$$

$$M(i, i) = \left[ \frac{3(i - 1)}{2n} \right]^2 + \left( 1 - \frac{3i}{n} \right)^2 + \left[ \frac{3(i + 1)}{2n} \right] \left[ 1 - \frac{3(i + 1)}{n} \right]$$

$$M(i + 1, i) = \frac{3i}{2n} \left( 1 - \frac{3i}{n} \right) + \left[ \frac{3(i + 1)}{2n} \right] \left[ 1 - \frac{3(i + 1)}{n} \right]$$

$$M(i + 2, i) = \left[ \frac{3(i + 1)}{2n} \right]^2 , \text{ and}$$

$$M(i, j) = 0 \text{ otherwise.} \quad (4.20)$$

Note that the stationary distribution of matrix $M$ is $\pi_M = 1/k = \pi_P = \pi_B$. Moreover, $M$ is symmetric since $M^T = (A^T \cdot A)^T = A^T \cdot A = M$, then $\pi_M(x)M(x, y) = \pi_M(y)M(y, x)$. In other words, $M$ is a reversible transition matrix. Thus, by the Perron-Frobenius theorem, the eigenvalues of $M$ are real and nonnegative, and the second largest eigenvalue $\lambda_2(M) \in [0, 1]$. To see this, we make use of the fact that $M$ is symmetric and thus orthogonally diagonalisable (by the spectral theorem).

Therefore, $M$ can be written as $SDS^{-1}$, where $S$ is invertible and $D$ is diagonal ($S, D$ not unique). Then, for diagonal $D$ we define $\sqrt{D} = (\sqrt{d_{ij}})_{i,j}$ (it is well-defined since all $d_{ij} \geq 0$). Hence,

$$M = SDS^{-1} = \left( S\sqrt{D}S^{-1} \right) \left( S\sqrt{D}S^{-1} \right) = B^2$$

$$\implies \lambda_M \mathbf{v} = B^2 \mathbf{v} = B(\lambda_B \mathbf{v}) = \lambda_B (B \mathbf{v}) = \lambda_B^2 \mathbf{v}$$

$$\implies \lambda_M = \lambda_B^2 \geq 0 .$$

**Bottleneck ratio definition.** Next we will introduce a generalised definition of conductance, which is related to the mixing time of the chain and the spectral gap.
of the transition matrix. An essential point of conductance is to formalise the idea that bottlenecks in a state space of a Markov chain make some regions of the graph difficult to enter and leave, and thus taking longer for the chain to converge. It is obvious that if a chain has some states which can only be reached via some restricted set of states, then the chain will mix slower and need longer to reach the stationary distribution.

We compute the bottleneck ratio following a method from Levin, Peres and Wilmer [77], Chapter 7.

Let the edge measurement $Q(x, y) = \pi(x)P(x, y)$ be the equilibrium flow through edge $(x, y)$. Denote by $Q(A, B) = \sum_{x \in A, y \in B} \pi(x)P(x, y)$ the ergodic flow between two subsets $A$ and $B$ of the state space. Then the bottleneck ratio of a set $S$ is defined as:

$$\Phi(S) = \frac{Q(S, \bar{S})}{\pi(S)} ,$$

where $\pi(S) = \sum_{x \in S} \pi(x)$. Note that $Q(A, B)$ is the probability that a Markov chain crosses the cut from $A$ to $B$ in one step under the stationary distribution, then $\Phi(S)$ can be viewed as the conditional probability of moving from $S$ to $\bar{S}$ in one step under the stationary distribution, given that the chain starts from a position in $S$.

The bottleneck ratio of the whole chain, i.e., conductance of the underlying graph, is defined to be

$$\Phi^* = \min_{S \in \Omega, S \neq \emptyset, \pi(S) \leq \frac{1}{2}} \Phi(S) .$$

**Bottle neck ratio of $M$.** As for the bottleneck ratio $\Phi^*_M$ of the reversible transition matrix $M$, we are seeking the minimum quantity $\min \Phi_M(S)$:

$$\Phi^*_M = \min_{S \in \Omega, S \neq \emptyset, \pi(S) \leq \frac{1}{2}} \Phi_M(S) = \min_{S \in \Omega, S \neq \emptyset, \pi(S) \leq \frac{1}{2}} \frac{Q(S, \bar{S})}{\pi(S)} = \min_{S \in \Omega, S \neq \emptyset, \pi(S) \leq \frac{1}{2}} \frac{\sum_{x \in S, y \in \bar{S}} \pi_M(x)M(x, y)}{\sum_{x \in \bar{S}} \pi_M(x)} = \min_{S \in \Omega, S \neq \emptyset, \pi(S) \leq \frac{1}{2}} \frac{\sum_{x \in S, y \in \bar{S}} M(x, y)}{|S|} .$$
By the entries of $M$ given in (4.20) we can see that only the ‘quint-diagonal’ entries, $M(1,k)$ and $M(k,1)$ are non-zero. Let $S = \{1, \ldots, i\}$, where $i \leq k/2$, then $\pi_M(S) = i/k \leq 1/2$, and $\bar{S} = \{i+1, \ldots, k\}$. Therefore the bottleneck ration of $S$ is

$$
\Phi_M(S) = \frac{Q(S, \bar{S})}{\pi(S)} = \frac{\sum_{x \in S, y \in \bar{S}} M(x, y)}{i} \\
\leq \frac{1}{i} [M(i-1, i+1) + M(i, i+1) + M(i, i+2) + M(1,k)] \\
= \frac{1}{i} \left\{ \left(\frac{3i}{2n}\right)^2 + \left(1 - \frac{3i}{n}\right) \frac{3i}{2n} + \left[1 - \frac{3(i+1)}{n}\right] \frac{3(i+1)}{2n} \right. \\
\left. + \left[\frac{3(i+1)}{2n}\right]^2 + \frac{3}{2n} \left(1 - \frac{3}{n}\right) \right\} \\
= \frac{1}{i} \left[ \frac{3i}{n} + O\left(\frac{1}{n}\right) \right] \sim \frac{3}{n}.
$$

Hence, the bottleneck ratio of the matrix $M$ is

$$
\Phi_M^* = \min_{S \in \Omega, S \neq \emptyset, \pi(S) \leq \frac{1}{2}} \Phi_M(S) \sim \frac{3}{n}.
$$

Relating bottleneck ratio to eigenvalue gap of $M$. We will apply Cheeger’s Inequality given in Theorem 42 to bound the second eigenvalue of $M$ in terms of its bottleneck ratio. Cheeger’s Inequality establishes an interesting connection between the second eigenvalue and the conductance, and it was originally from Cheeger (1969)[25] and was modified and proved by Jerrum and Sinclair (1989) [77].

**Theorem 42.** (Cheeger’s Inequality) Let $\lambda_2$ be the second largest eigenvalue of a reversible Markov chain, and $\Phi^*$ be the bottleneck ratio of the whole chain. Then the eigenvalue gap satisfies:

$$
\frac{\Phi^*^2}{2} \leq 1 - \lambda_2 \leq 2\Phi^*.
$$

Now we are ready to prove Theorem 39 using all the above analysis and the computation of $\lambda_2(M)$ and $\Phi_M^*$. 

127
Proof of Theorem 39. From the bound on the variation distance (4.19) in Theorem 41 (Fill, 1991) we have

\[
4\|P^t(i, \cdot) - \pi\|^2 \leq \frac{(\lambda_2(M))^t}{\pi(i)}.
\]

We obtain an upper bound \(\lambda_2(M) \leq 1 - \Phi_M^* / 2\) from Theorem 42 (Cheeger, 1969). Recall that the bottleneck ratio of \(M\) is \(\Phi_M^* \sim 3/n\). This gives an upper bound on the variation distance claimed in (4.18):

\[
4\|P^t(i, \cdot) - \pi\|^2 \leq \left(1 - \frac{\Phi_M^*}{2}\right)^t k \sim \left(1 - \frac{9}{2n^2}\right)^t \frac{n}{3},
\]

thus completing the proof of Theorem 39.

\[\square\]

4.3.5 Experimental results

We give the results of some experiments on the convergence, of lack of it, to the predicted asymptotic expected number \(N_j\) of vertices on cycles length \(3j\). For \(p > q\) and \(k = n/3\), the expected number \(N_j(p, n)\), of vertices on cycles length \(3j\), where \(j = \{1, \ldots, k\}\), is given by

\[
N_j(p, n) = \left[\left(\frac{q}{p}\right)^{j-1} \frac{p-q}{p-q\left(\frac{q}{p}\right)^{k-1}}\right] n.
\]

Let \(C_j(p)\) be the expected number of cycles of length \(3j\). As \(n\) increases, \((q/p)^k \to 0\), so the asymptotic expected number is

\[
C_j(p) = \frac{N_j(p, n)}{3j} \to \frac{k p - q}{j} \left(\frac{q}{p}\right)^{j-1}, \quad (4.21)
\]

which is monotone decreasing in \(j\) and \(p\). As \(p\) increases, the values of \(C_j\) separate clearly in the plots, and the number of long cycles decreases. The experiments confirm this.

As \(p\) decreases to \(p = 1/2\), the influence of \(k\) in the denominator of \(N_j(p, n)\) increases. The formal limit of \(N_j(p, n)\) tends to \(n/k = 3\), which is confirmed by subsection 4.3.2 (case \(p = q\)) of this chapter. Thus \(C_i = N_i/(3i) = 1/i\). However
this value ignores the fact that we assumed $\rho \sim 1$ whereas $\rho < 1$ if there is only one triangle in expectation. Thus the triangle-break move of the simulation must fail on occasions, and the resulting output is conditioned by this. We can measure the number of fails as part of the simulation.

The experiments are initialised with a single Hamilton cycle of length $n = 3k$, so we cannot expect any convergence to the final answer in $o(n)$ steps even for large $p$, as we need to 'break off' order $n$ triangles. We might expect convergence in $O(n \log n)$ steps in best case. The case $p = 0.75$ seems to confirm this, as the data plot is stable after 10,000 steps for $n = 3,000$. As $p$ decreases to $p = 0.5$ the experimentally observed quality of the convergence degrades visibly. When $p = 0.5$, there is a positive probability that a $q$-move 'fails', as there is no triangle with which to merge. This should bias the effective value of $p$ upwards, and the final experiment, running an $n = 300$ system to stationarity in $O(n^2 \log n)$ steps seems to support this.

On average the runs took less than an hour elapsed time each. The main experimental outputs are as follows.

1. For $p = 0.5, 0.51, 0.53, 0.55, 0.75$.

2. They are initialised with a cycle of $n = 3000$ vertices and run for $t = 50,000$ steps for $p = 0.75$, $t = 100,000$ steps for $p = 0.51, 0.53, 0.55$, and $t = 500,000$ steps for $p = 0.5$.

3. The data plot over time is given for cycles of length 3 (red), 6 (green), 9 (blue).

4. The horizontal lines are

   (i) Coloured line: experimental average taken over the last 1,000 steps,

   (ii) Black line: the value of $C_j$ as given by Equation (4.21).

5. A data summary including the numeric averages for $j = 1, \ldots, 10$ for the last 1,000 steps; and the list of component sizes for $p = 0.5, 0.51$.

6. The final graph for visual inspection of large cycles.
Probability $p = 0.75$, $n = 3000$

This is the most convincing example, with good experimental evidence of convergence. The value of $C_j$ (black line) coincides with the experimental averages (coloured line). No components of size $3j$ were observed for $j \geq 6$ in the last 1,000 steps.

Probability $p = 0.55$, $n = 3000$

The convergence is still quite reasonable, but the experimental variance in the number of cycles of a given size is evident. The final graph shows the emergence of longer cycles.
Probability \( p = 0.53, \ n = 3000 \)

The length of the experiment was doubled to 100,000 steps but the convergence for cycles length 3 (red curve) is poor. The expected value of \( C_j \sim 113 \), but the observed is 138. The number of long cycles in the final graph is increasing.
Probability $p = 0.51$, $n = 3000$

The run time was increased to 100,000 steps. The convergence of the average is still reasonable, but the variability in the number of triangles has increased relative to the mean. The expected value of $C_j = 39$, and the data average for this experiment is close at 42. The graph now has many larger cycles. The number of components of each size at the final step is given in full. The largest, at 333 vertices, is more than one third of the graph size (see last line of data summary).
CHAPTER 4. TRIANGLE-SWITCH PROCESSES ON 2-REGULAR GRAPHS

Probability $p = 0.5$, $n = 3000$

What to expect here is unclear. The value $C_j = 1/j$ is far from the observed values after 500,000 steps (see below). The earliest part of the simulation would be mainly spent breaking triangles off the initial cycle. The percentage of triangle-break failures is around 1% (5718 fails in 500,000 steps) which is not very high. The final graph has 5 triangles.

The inferred mixing time upper bound of $T = O(n^2 \log n)$ is very long (see subsection 4.3.4 (The method of Fill)). An examination of the constants in the Inequality (4.18) shows that after less than $T = n^2 \log n$ steps the system is close to
stationarity. We ran the simulation on a smaller graph with \( n = 300 \) vertices. The results are given in the next subsection.

For \( n = 3000 \), \( T \) is around \( T = 70,000,000 \), much longer than the 500,000 steps of the simulation with output shown below.

**Probability \( p = 0.5, n = 300 \). Running the simulation to stationarity**

The number of fails in 500,000 steps is 18,800, about 3.8%. The data plot is similar to the \( n = 3,000 \) case, and similarly unrevealing. How to interpret \( N_i = 3 \), where \( i = 1, \ldots, k \)? The value of \( C_i = 1/i \) goes down in rank order, so maybe we should see
the number of 6-cycles in the simulation as half the number of 3-cycles, the number of 9-cycles as 1/3-rd, and so on? Another approach is to expect 3i vertices on cycles length 3, 6, 9, ..., 3i. Neither seems to work and the averages data looks more like the \( p > q \) case. In the end, the most probable interpretation is that the small but positive chance of fails reduces \( q \) and increases \( p \) slightly, so that the matrix \( A \) with \( p = q \) is not valid. As a guess, we might suppose \( p = \frac{1}{2} \left(1 + \frac{18,800}{500000}\right) = 0.5188 \approx 0.52 \).

Equation (4.21) predicts 7, 7, 3, 6, 2, 2, 1, 4 for \( C_i \) where \( i = 1, \ldots, 4 \), to be compared with 8, 3, 8, 2, 1, 4 from the data. Not too unreasonable.
Chapter 5

Permutation Forests

5.1 Introduction

Consider a class of random processes for producing a spanning forest from an undirected graph as follows. Given a graph $G$, we produce our forest by first randomly ordering the vertices of $G$; say $1, 2, ..., n$. We process the vertices in the given order starting with 1 and so on. For each vertex $x$ we examine its neighbours that appear after it in the ordering, and select an edge that connects $x$ to a neighbour $y$ according to different rules in the three protocols: Pick-First, Pick-Last and Pick-UAR, if such an edge exists. In the Pick-First model, we choose the first neighbour appearing after $x$ in the ordering; while in Pick-Last model we select the last neighbour of $x$ that appearing after it in the ordering; and for Pick-UAR model, we choose a neighbour uniformly at random which appears after $x$ in the ordering. The selected edge for each vertex $x$ are retained, and all other edges $xy, y > x$ are discarded. Hence, for each vertex, at most one ‘forward’ edge, to a vertex higher in the order is preserved from the underlying graph, and vertex with no forward edge forms the root of a component of the forest. We refer to such a forest as a permutation forest of the underlying graph $G$. For instance, in the complete graph $K_n$ under Pick-First neighbour rule, the process generates a Hamilton path in the permutation order.
We will give three simple examples and see how edges are selected, in order to visualise how a forest is generated by this process, under the Pick-First, Pick-Last, and Pick-UAR protocols, respectively. For instance, given a graph of 5 vertices as shown on the left-hand side of Figure 5.1, we first randomly order the five vertices of the graph. To generate a permutation forest under Pick-First rule, we copy all vertices from the underlying graph. As for the edges, we then examine the neighbours of each vertex one by one according to the ordering. Vertex \( v_1 \) is connected to vertices \( v_2 \) and \( v_4 \) in the underlying graph, so we choose the first neighbour appearing after \( v_1 \), which is vertex \( v_2 \), and retain the edge \((v_1, v_2)\) in the permutation forest. While vertex \( v_2 \) connects to vertices \( v_1, v_3 \) and \( v_4 \) in the underlying graph, we retain the edge \((v_2, v_3)\) as \( v_3 \) is the first neighbour of \( v_2 \) appearing after \( v_2 \) in the ordering. Since \( v_3 \) does not have any neighbour that appears after it in the ordering, it fails to connect and it is a root of a tree (component) in the forest. Similarly, we retain the edge \((v_4, v_5)\) for vertex \( v_4 \), and \( v_5 \) is another root of a tree in the forest. The graph on the right-hand side of Figure 5.1 shows the resulting permutation forest generated by Pick-First process. There are two trees rooted at \( v_3 \) and \( v_5 \) in this forest.

Figure 5.1: Permutation forest obtained from the Pick-First process. Underlying graph \( G \) shown on left-hand side, with vertices ordered 1, 2, 3, 4, 5.

Figure 5.2 gives a simple example of how a permutation forest is generated from a given underlying graph under the Pick-Last protocol. We use the same underlying graph as in the example of Pick-First model (see Figure 5.1) to see how choices of different rules when choosing the edges to retain would affect the forests they produce. The graph on the left-hand side in Figure 5.2 is the given graph with a
random ordering on the five vertices. To obtain the generated forest $F$ we add all vertices and then reveal the edges by checking the forward neighbours of each vertex from $v_1$ to $v_5$ in the given order. For vertex $v_1$ we add the edge $(v_1, v_4)$ to $F$ since $v_1$ has two neighbours $v_2$ and $v_4$ in the underlying graph and $v_4$ is the last one appearing after it in the ordering. Vertex $v_2$ connects to $v_1, v_3$ and $v_4$, so the edge $(v_2, v_4)$ should be added to $F$ as $v_4$ is the last neighbour of $v_2$ that appearing after it in the ordering. Following the same procedure of selecting edges to present in $F$ which coming from vertices $v_3, v_4$ and $v_5$, we finally obtain the trees as shown on the right-hand side of Figure 5.2.

![Figure 5.2: Permutation forest obtained from the Pick-Last process.](image)

As for the example to visualise how a permutation forest is generated by Pick-UAR process, suppose we are given a graph shown on the left-hand side in Figure 5.3 with an ordering on the vertices, which is the same graph as in the examples of Pick-First and Pick-Last models. To construct the permutation forest, we first copy all the vertices and then check the neighbours of each vertex based on the ordering to choose the edges to present in the forest. Starting with vertex $v_1$, which is connected to $v_2$ and $v_4$ in the underlying graph, we choose a random vertex between $v_2$ and $v_4$, say $v_4$, and add the edge $(v_1, v_4)$ to the forest. For vertex $v_2$ we can see that its neighbours appearing after it are $v_3$ and $v_4$. Choose a neighbour of $v_2$ uniformly at random, say $v_3$, and we add the edge $(v_2, v_3)$ to the forest. Since vertex $v_3$ does not have any neighbour appearing after it, no edge will be added, and it will be a root in the forest. For $v_4$ the only neighbour that appears after it in the ordering is $v_5$, so we add the edge $(v_4, v_5)$ to the forest. Under the Pick-UAR rules, the permutation
forest generated from the given underlying graph with an ordering on the vertices consists of two trees, each of which is a path, as shown on the right-hand side in Figure 5.3.

![Figure 5.3: Permutation forest obtained from the Pick-UAR process.](image)

We study these random processes to produce a spanning forest from an underlying Erdős-Rényi random graph $G(n, p)$, where each edge occurs with probability $p$ uniformly and independently of each other edge. We examine various properties the forest inherits under the choices of three different protocols: Pick-First, Pick-Last and Pick-UAR. Aspects of the Pick-First protocol were previously studied by Vu [100]. The emergence of a giant component in the permutation forest generated by Pick-First process was studied by Cooper, Kang, McDowell, Radzik and Vu [36].

The permutation forest subgraphs generated by these processes can be seen as a type of random recursive forest. A random recursive forest is generated by adding vertices one at a time to an existing forest. Each new vertex chooses a single vertex of the existing forest as a parent to which it attaches. To generate the permutation forest in this way, we would have to start with the last vertex $n$ in the order, and work backwards to $n - 1$, and so on, vertex 1 being added last.

Random recursive trees have attracted considerable attention, with most work focusing on two models in particular: $i)$ the uniform model, where a vertex uniformly chooses a parent among all existing vertices (see e.g. [8, 43, 96, 97]), and $ii)$ the preferential attachment model, where a vertex chooses a parent with probability proportional to the degree of the parent vertex (see e.g. [9, 74, 75, 82, 83]).
In a more recent work [43], Devroye, Fawzi and Fraiman considered a model in which vertex \( i \) connects to vertex \( \lfloor iX_i \rfloor \) in the ordering, where \( X_i \)'s are identically distributed random variables on \([0, 1)\). They studied different choices for \( X_i \)'s. Moreover, if \( X_i \)'s are uniform, it is identical to the uniform recursive forest we mentioned above. Our three models are distinct from this case since if using their model it would require \( X_i \)'s to be non-identically distributed.

In another recent work [75], Korenblit, Tails and Levin proposed a preferential attachment model which bears some similarities to ours. In their model, a vertex attaches to a vertex of the highest degree with probability \( p \). If failing to attach, it attempts to attach to a vertex with the next highest degree with probability \( p \) and so on. Since the underlying graph of our three models is drawn from the Erdős-Rényi random graph distribution, where each edge exists independently with probability \( p \), our models follow a similar structure. However, the order of attempted connections are determined entirely by the order of the vertices in our Pick-First and Pick-Last models and uniformly at random in our Pick-UAR model with respected to the existence or non-existence of the random edges, rather than the degree of vertices as in the model of [75].

We are interested in the case where the initial random ordering of the vertices is chosen uniformly at random from all permutations on \([n]\). For \( G(n, p) \) the edges are i.i.d.\(^1\) with probability \( p \), so this is equivalent to the identity permutation on the vertices, which simplifies matters. The permutation forest generated from \( G(n, p) \) has some features analogous to \( G(n, p) \) graphs. One of the key properties of \( G(n, p) \) that make the model tractable, is the independence of the existence of each edge. For example, under the Pick-First rule, the probability that any two vertices form a descendant-ancestor pair is exactly \( p \) (this is proved in Proposition 46). However, if a vertex \( x \) in the permutation forest is a descendant of a vertex \( y \) then it implies e.g., that \( x \) is the descendant of all ancestors of \( y \). Therefore, we cannot expect

\(^1\)We use ‘i.i.d.’ to denote ‘independent and identically distributed’.
independence of edges for all pairs of vertices.

Finally, several alternative models of random forests have been considered related to \( G(n, p) \) and \( G(n, m) \) random graphs in different ways, where \( G(n, m) \) is a random graphs with \( n \) vertices and \( m \) edges. For instance, in [3], edges of the underlying graph are added repeatedly, which are selected uniformly at random from the set of all remaining edges which do not create a cycle; while in [81], a forest is chosen uniformly at random from all vertex-labelled forests with \( n \) vertices and \( m \) edges.

## 5.2 General results

Given a graph \( G = G(V, E) \), we produce our permutation forest \( F \) by first giving a random ordering \( \pi \) on the vertices of the underlying graph \( G \). Then, we copy all vertices of \( G \) to \( F \), and for each vertex we examine its neighbours according to the ordering to choose the edges that would present in \( F \). To do this, for each vertex we select the edge that connects it to the first or last neighbour in the ordering, or a uniformly chosen neighbour, appearing after it in the ordering, by Pick-First, Pick-Last or Pick-UAR process, respectively, if such an edge exists. A vertex fails to connect and becomes the root of a tree component if no neighbour appears after it in the ordering. A permutation forest is generated by retaining the selected edges in the process and discarding all other edges.

We can make a number of immediate observations on the properties of \( F \) which are true for every graph and choice of permutation. Firstly, we prove that \( F \) is indeed what we claim - a forest.

**Proposition 43.** For any graph \( G \) and choice of \( \pi \), the permutation forest (subgraph) \( F = \pi,G \) is a forest, independently of the model used to choose the retained edges.

**Proof.** Suppose \( F \) contains a cycle. Every vertex in that cycle must have degree at least 2 in \( F \). If we examine the vertex in this cycle, which appears first in the permutation \( \pi \), then as it must have two edges in this cycle, both must be forward
edges, since all the other vertices are ordered after it. This is a contradiction, and thus $F$ does not contain a cycle.

Observe that the number of components (trees) of $F$ is precisely the number of vertices which fail to connect. This follows from observing that we have added one edge to $F$ for each vertex that did not fail to connect. Thus the number of edges in a forest is the number of vertices subtracted by the number of vertices which fail to connect. Each component is a tree, consisting of forward edges, terminating at a single vertex which failed to connect. We find it is helpful to visualise this as a rooted tree with the root as the vertex that failed to connect. These vertices are the roots of the trees. The leaves of the trees are vertices for which no other vertex connected to them with a forward edge.

A natural question is how many components we should expect to see in the final graph $F$. The expected number of components is independent of the model used to choose the retained edges, although of course the distribution of sizes of the components may differ between models.

**Proposition 44.** Let $G$ be a graph with a given degree sequence $d_1, d_2, \cdots, d_n$. Denote by $R$ the number of roots (components) in the permutation forest generated from $G$. Then, regardless of the model used to choose the retained edges,

$$\mathbb{E}[R] = \sum_{i=1}^{n} \frac{1}{d_i + 1}.$$  

**Proof.** Notice that the probability of any given vertex $v_i$ with degree $d_i$ being a root is the probability that it appears after all of its neighbours in the permutation. Given any set of vertices, the probability that a particular vertex appears last amongst them is uniform. Thus the probability of $v_i$ being a root is the probability that it appears last amongst the union of its neighbourhood and itself, a set of size $d_i + 1$. By linearity of expectation, the expected number of roots is therefore $\mathbb{E}[R] = \sum_{i=1}^{n} \frac{1}{d_i + 1}$ as required.
Since $F$ is a permutation forest, it has at most $n$ edges, so the total sum of the degrees is less than $2n$. It tells us that the expected total degree of any given vertex, wherever it appears in the order is strictly less than 2.

5.2.1 Expected number of components in $G(n, p)$

Considering the case where $G = G(n, p)$, we prove that the number of roots exhibits reasonably tight concentration for a wide range of values of $p$.

**Theorem 45.** Let $F$ be a permutation forest generated from an Erdős-Rényi random graph $G(n, p)$ with a fixed permutation $\pi$ on the vertices. Denote by $R$ the number of components of $F$. Let $\omega/n \leq p \leq 1$, where $\omega = \omega(n) \to \infty$. Then, regardless of the model used to choose the retained edges (Pick-First or Pick-Last or Pick-UAR protocols), the expected number of components $E[R] \sim 1/p$. Moreover for any $\omega/\sqrt{n} \leq \varepsilon \leq 1$, the number of components is concentrated in the interval $(1 \pm O(\varepsilon))E[R]$ with high probability.

**Proof.** Let $d^+_i$ be the forward degree of vertex $v_i$, which has only binary value 0 or 1, since each vertex in the permutation forest connects to at most one forward neighbour (neighbour with higher index in the ordering) by the definition of Pick-First, or Pick-Last, or Pick-UAR protocol.

Consider the number of roots as a sum of $n$ independent indicator variables $R_i$ for $i \in \{1, \ldots, n\}$, where $R_i = 1$ if vertex $i$ is a root and 0 otherwise. Then, we have $R = \sum_{i=1}^n R_i$, and $P(R_i = 1) = P(d^+_i = 0)$.

A component is formed by a vertex with no forward edges in the permutation order, regardless of the model used to choose retained edges. Thus the expected number of roots $R$, is

$$E[R] = \sum_{i=1}^n P(d^+_i = 0) = \sum_{i=1}^n (1 - p)^{n-i} = \frac{1 - (1 - p)^n}{p}.$$ 

Let $p = c/n$ where $c \geq \omega$, then $(1 - p)^n = e^{-np}(1 - o(1)) \sim e^{-\omega} = o(1)$, in which case the expected number of components in $F$ is $E[R] \sim (1 - e^{-\omega})/p = (1 - o(1))/p$. 

143
By Chernoff-Hoeffding bounds we have the following result: for any $0 < \varepsilon < 1$,\
\[
\Pr \left( R \geq (1 + \varepsilon) \mathbb{E}[R] \right) \leq \exp \left\{ -\frac{\varepsilon^2 \mathbb{E}[R]}{2 + \varepsilon} \right\} = \exp \left\{ -\frac{\varepsilon^2 (1 - O(e^{-np}))}{(2 + \varepsilon)p} \right\} \to e^{-\varepsilon^2/(2 + \varepsilon)p},
\]
\[
\Pr \left( R \leq (1 - \varepsilon) \mathbb{E}[R] \right) \leq \exp \left\{ -\frac{\varepsilon^2 \mathbb{E}[R]}{2} \right\} = \exp \left\{ -\frac{\varepsilon^2 (1 - O(e^{-np}))}{2p} \right\} \to e^{-\varepsilon^2/(2p)}.
\]

Thus, for $\omega/n \leq p \leq 1$ and $\varepsilon \geq \sqrt{\omega p}$, there exists concentration of the number of roots in the permutation forest of a random graph around $1/p$, i.e.

\[
\Pr \left( (1 - \varepsilon) \mathbb{E}[R] \leq R \leq (1 + \varepsilon) \mathbb{E}[R] \right) = 1 - O \left( e^{-\varepsilon^2/(3p)} \right) \to 1.
\]

\[\square\]

### 5.2.2 Remarks on Theorem 45

The result of Theorem 45 is valid for all three processes: Pick-First, Pick-Last and Pick-UAR; and indeed any other such process we might devise. For $\omega/n \leq p \leq 1$, the number of components in the permutation forest generated by the Pick-First, Pick-Last or Pick-UAR models is concentrated around $1/p$ with high probability.

Given the three models all form a forest $F$ of around $1/p$ trees, what structural differences do these forests exhibit, if any? There are various possible measures of difference such as maximum tree height from the root, maximum component size, distribution of component sizes etc., which we can investigate theoretically or experimentally. Tables 5.4 and 5.5 summarise theoretical findings; for experimental results see Section 5.6.
Figure 5.4: Order of magnitude results for maximum component size in the three models.

In Pick-First process, it seems likely that as \( p \) increases the forward paths will begin to intersect, and thus components begin to merge, producing increases in the largest component size (see Figure 5.11 for result of simulations). For \( p \geq \omega/\sqrt{n} \), there exists a linear component in the permutation forest with high probability [36]. In Pick-Last and Pick-UAR processes, the size of the largest component grows as \( p \) increases, which is shown in the simulations in Section 5.6.2. For \( 0 \leq p < 1 \), the expected size of the largest component is \( 1 + \Theta(np) \) for both processes.

Figure 5.5: Order of magnitude results for maximum tree height \( H \) in the three models.

In Pick-First process, the independence provided by the forthcoming Proposition 46 enables us to show why a forward path from the first vertex in the ordering
has expected length $O(np)$ for $p \geq \omega/n$. Compared to Pick-First process, the maximum height of trees produced by Pick-Last process is much shorter in expectation, which is $O(\log n/\log \log n)$ for $n^{-\alpha} \leq p \leq o(1/\log n)$. This is because almost all the first $n - s$ vertices are linked directly to the last $s$ vertices with high probability, where $c\log n \leq s \leq cn/\omega$ for constant $c > 1$ (see Theorem 55). As for Pick-UAR, the expected maximum tree height is $O(\log n)$. More randomness has been added when revealing the forward neighbour of each vertex, resulting in more randomness for the increase of path lengths.

5.3 Pick-First model

We give a formal definition of the Pick-First process as follows. Let $G(n, p) = G(V, E)$ be an Erdős-Rényi random graph, and let $\pi(V) = (v_1, \ldots, v_n)$ be an initial random ordering (permutation) on the vertices of $G$. Note that in a $G(n, p)$ random graph, each vertex has the same local environment, so that no vertex can be distinguished from any other based on the vertices and edges surrounding it. Therefore, we can just let this permutation $\pi$ be the identity permutation.

Define a forest $F = F_{\pi,G} \subseteq G$ as follows. Let $F$ contain all vertices of $G$, i.e. $V(F) = V(G)$. For each vertex $v_i \in V$, where $i = \{1, \ldots, n\}$, we add an edge between $v_i$ and $v_j$ in $F$ where $j = \min\{k > i : (v_i, v_k) \in E(G)\}$. If there is no edge between $v_i$ and vertices after $v_i$ in the ordering, then no edge will be added to $F$. We say such a vertex ‘fail to connect’. We call a path $(v_1, v_2, \ldots, v_k)$ increasing if $\pi(v_1) < \pi(v_2) < \cdots < \pi(v_k)$.

Note that the permutation forest $F$ generated from a random graph $G(n, p)$ with a random ordering $\pi$ is also a random structure, which is vertex-transitive in distribution. In an vertex-transitive graph, every pair of vertices is equivalent under some element of its automorphism group, i.e. for any two vertices $u$ and $v$ of a vertex-transitive graph $G$, there exists an automorphism $f : G \to G$, such that $f(u) = v$. If we first fix the permutation $\pi$ on the vertices of $G$ and then determine which edges
in the underlying graph are present in $F$ by gradually revealing neighbours of each vertex according to this order, we see that the choice of permutation does not impact on the unlabelled structure of $F$ in this case.

In a permutation forest $F$, every vertex will have at most one ‘forward’ edge, to a vertex appearing after it in the ordering, but may have more than one ‘back’ edge, where other vertices have connected to it. While both the underlying graph $G$ and its subgraph $F$ are undirected, it will be useful to view each edge $(v_i, v_j)$ of $F$ as directed from vertex $v_i$ connecting to another vertex $v_j$ when we analyse the structure of $F$.

We consider the following topics for the Pick-First model.

1. Monotone paths
2. Expected degree of a vertex
3. Size of the largest component

### 5.3.1 Monotone increasing paths

A monotone increasing path $x_1, ..., x_k$ in a permutation forest is a path such that $x_1 < x_2 < \cdots < x_k$ in the permutation order of the vertices. The example shown in Figure 5.1 consists of two monotone paths 1, 2, 3 and 4, 5. We will show that the probability that a vertex $v_j$ is in an monotone increasing path starting at $v_i$, where $i < j$, in the Pick-First permutation forest is exactly $p$.

**Proposition 46.** Denote by $F = F_{\pi, G(n, p)}$ the permutation forest generated by Pick-First process from an underlying graph $G(n, p)$ and a permutation $\pi$ on the the vertices of $G$. For any two vertices $x, y \in V(G)$ with $\pi(x) < \pi(y)$, the probability that there exists an increasing path between $x$ and $y$ in $F$ is exactly $p$.

**Proof.** For $1 \leq i \leq n$, let $v_i$ be the vertex $v$ such that $\pi(v_i) = i$. The probability that $v_i$ connects to $v_{i+1}$ is $p$. We consider the forward path from $v_i$ formed by taking the vertex (if any) to which $v_i$ connects, say vertex $u$, then taking the vertex to which $u$ connected by that vertex and so on. We claim that for any fixed $j > i$, the probability
that $v_j$ is in the forward path $Q_i$ starting from $v_i$ is exactly $p$, independently of the event of any other fixed vertex being in $Q_i$.

We can construct the path $Q_i$ starting from $v_i$ by examining one by one vertices $v_{i+1}, v_{i+2}, \ldots$. Suppose we have reached vertex $v_{j-1}$ in this examination. Either $v_{j-1}$ is a vertex of $Q_i$ in which case $v_j$ is in the path with probability $p$, or some earlier vertex $v_k$, $i \leq k < j - 1$ was the last vertex in the path, and the edges $(v_k, v_{k+1}), \ldots, (v_k, v_{j-1})$ failed to be added. With probability $p$, $(v_k, v_j)$ is a path edge independently of the previous history of the path construction, and this holds for every vertex $v_j$ with $j > i$.

This forward path is clearly an increasing path, and in addition, since each vertex has at most one forward edge in the permutation forest $F$, there is an increasing path from $x$ to $y$, if and only if, $y$ is in the forward path from $x$, which we have shown occurs with probability $p$. \hfill \Box

If we examine only the forward path from a single vertex, at each step we only consider forward edges from the most recent vertex in the path. Therefore we have not examined and hence exposed the randomness of any forward edges from any vertex that is not in the forward path. As each edge in the forward path of a vertex occurs independently with probability $p$, we have the following corollary.

**Corollary 47.** The forward path from vertex $\pi(1)$ has expected length $(n - 1)p$ and this path length is concentrated provided $np \rightarrow \infty$.

### 5.3.2 Expected degree of a vertex

Since $F$ is a permutation forest, it has at most $n$ edges, and so the total sum of the degrees is less than $2n$. It tells us that the expected total degree of any given vertex, wherever it appears in the order is strictly less than 2.

In the case where $G = G(n, p)$, for a vertex $v_i$ that appears in the position $i$ in the permutation, we call a forward edge of $v_i$ an edge which connects $v_i$ to a vertex
with \( j > i \) in the generated forest \( F \), if such an edge exists; while for those edges that connect any vertex \( v_k \) to \( v_i \) for any \( k < i \), we call them backward edges of \( v_i \). Denote by \( d_{+}^i \) the forward degree of \( v_i \) and \( d_{-}^i \) the backward degree of \( v_i \).

The following proposition gives a result on the symmetry between forward and backward degrees.

**Proposition 48.** Let \( G(n, p) \) be a random graph and \( \pi = (v_1, \ldots, v_n) \) a random permutation on the vertices of \( G \). Denote by \( F = F_{\pi,G} \) the permutation forest generated by Pick-First process. For a vertex \( v_i \) in \( F \), its forward and backward degrees satisfy:

\[
E[d_{+}^i] = E[d_{-}^{n-i+1}].
\]

**Proof.** Notice that for a vertex \( v_i \) that appears in position \( i \) in the ordering, its forward \( d_{+}^i \) degree is a \( 0-1 \) indicator variable which takes value 1 with probability:

\[
P(d_{+}^i = 1) = E[d_{+}^i] = p + (1-p)p + (1-p)^2p + \cdots + (1-p)^{n-i-1}p
\]

\[
= p \left( \sum_{j=0}^{n-i-1} (1-p)^j \right) = p \left( \frac{1 - (1-p)^{n-i}}{p} \right)
\]

\[
= 1 - (1-p)^{n-i}.
\]

This can be alternatively formulated by observing that it takes value zero, if and only if there exist no edges from the vertex to another vertex after it in the permutation, which occurs with probability \( (1-p)^{n-i} \).

For \( 1 < i \leq n \) the backward \( d_{-}^i \) degrees are a sum of indicator variables, with expectation \( p, (1-p)p, (1-p)^2p, \ldots \), since there exists an edge in \( F \) from a vertex \( v \) to \( v_i \), if and only if there is an edge between them in \( G \), and no edge between \( v \) and the intermediate vertices in the permutation between them, giving:

\[
E[d_{-}^i] = p + (1-p)p + (1-p)^2p + \cdots + (1-p)^{i-2}p
\]

\[
= p \left( \frac{1 - (1-p)^{i-1}}{p} \right) = 1 - (1-p)^{i-1}.
\]
Interestingly, this is symmetric, so \( E[d_1^+] = E[d_n^-] \), and in general,

\[
E[d_i^+] = E[d_{n-i+1}^-].
\]

### 5.3.3 Size of the largest component in Pick-First model

We will prove that if \( p \) is at least \( \sqrt{\log n/\varepsilon n} \) for some small \( \varepsilon \), then all vertices except possibly for the final \( \varepsilon n \) vertices in the permutation order belong to the same component.

**Theorem 49.** Denote by \( F = F_{\pi,G} \) the permutation forest generated from a random graph \( G(n, p) \) and a permutation \( \pi(V) = (v_1, \cdots, v_n) \) by Pick-First process. Given \( \delta > 0 \), if \( p \geq (1+\delta)\sqrt{\log n/\varepsilon n} \), then with probability at least \( 1-O(n^{-\delta}) \), the size of the largest component of \( F \) is \((1-\varepsilon)n = \Theta(n)\), where \( 0 < \varepsilon < 1 \) and \( \varepsilon = \omega((\log n)/n) \).

**Proof.** From Proposition 46, we observe that the forward path from the first vertex contains each other vertex independently with probability \( p \). This allows us to demonstrate that this forward path contains reasonably many vertices in the final \( \varepsilon n \) vertices in the permutation. We show this result in a way that does not expose too many of the random edges of the graph when examining a forward path, and then use these remaining unexposed random edges to show that any vertices appearing before this final block that have not already been added to this forward path, will with high probability connect to at least one of these vertices in the path, putting them in the same component.

We start by examining the forward path \( P_1 \) beginning at \( v_1 \). For each vertex \( v_j \) where \( j \geq (1-\varepsilon)n \), the probability that \( v_j \) is in this path is \( p \). Consider the final \( \varepsilon n \) vertices in the ordering, each of which is present in path \( P_1 \) with probability \( p \). Therefore the number of the vertices in the last \( \varepsilon n \) block that are contained in this path is the sum of the \( \varepsilon n \) independent random indicator variables, each equal to 1 with probability \( p \).
Denote by $X$ the number of vertices in the final $\varepsilon n$ vertices in the ordering which are present in path $P_1$. Then $X = \sum_{i=\varepsilon n}^{n} x_i$, where $x_i = 1$ with probability $p$ and 0 otherwise, and $EX = \varepsilon np$.

Let $\delta > 0$ and $\eta = \delta/(1 + \delta)$. If $p \geq (1 + \delta)\sqrt{\log n}/\sqrt{\varepsilon n}$, then by Hoeffding-Chernoff inequality, we have

$$P(X \leq (1 - \eta)EX) \leq \exp\left\{ -\frac{\eta^2EX}{2} \right\} = \exp\left\{ -\frac{\delta^2p\sqrt{\varepsilon n \log n}}{2(1 + \delta)} \right\} = n^{-\delta^2/2} = o\left(\frac{1}{n}\right).$$

Therefore, with high probability, the forward path $P_1$ from $v_1$ contains at least $\sqrt{\varepsilon n \log n}$ vertices appearing within the last $\varepsilon n$ vertices in the ordering.

Let $A$ be the collection of the vertices in the final $\varepsilon n$ block in the ordering which are present in path $P_1$. Then $|A| = X \geq \sqrt{\varepsilon n \log n}$. We construct path $P_1$ beginning from $v_1$ by checking the forward connections of each vertex according to the ordering. Suppose we now reach a vertex $v_i$ for $1 < i \leq (1 - \varepsilon)n$, and any previous vertex that has been added to forward path $P_1$ does not connect to $v_i$. Since we have not exposed the forward edges of $v_i$, then with probability $p$ there will be a path connecting $v_i$ to the first vertex of $A$ in the ordering, thus connecting this path to path $P_1$. If the path from $v_i$ does not connect to the first vertex in $A$, then it will again connect to the next vertex of $A$ with probability $p$, and so on. The probability that this path from $v_i$ does not connect to any vertex of $A$ is

$$(1 - p)^X \leq e^{-pX} \leq e^{-(1+\delta)\log n} = n^{-(1+\delta)}.$$

Thus, with probability at least $1 - n^{-(1+\delta)}$, the path from $v_i$ will connect to path $P_1$. Taking a union bound, we conclude that with high probability all of the first $(1 - \varepsilon)n$ vertices in the ordering are in the same component as required.  

**Remark 50.** In [36], Cooper, Kang, McDowell, Radzik and Vu analysed the threshold behaviour of $p$ for the emergence of a linear component. They proved that for $p \geq \ldots$
\( \omega / \sqrt{n} \), where \( \omega = \omega(n) \rightarrow \infty \), there exists a linear component in the permutation forest with high probability.

### 5.4 Pick-Last model

In the Pick-Last process, for each vertex \( v \) we choose the last neighbour that appears after it in the ordering, rather than the first forward neighbour as in Pick-First model; and then we retain the edge between them if such an edge exists and discard all other edges. A vertex that fails to connect becomes a root of a tree (component) if no neighbour appears after it in the ordering. Same as in Pick-First process, there is at most one ‘forward’ edge from each vertex (to another vertex higher in the ordering) in the permutation forest produced by Pick-Last process, and there may be more than one ‘back’ edge.

We define the Pick-Last process formally as follows. Given a random graph \( G(n, p) = G(V, E) \) with a random ordering \( \pi(V) = (v_1, \ldots, v_n) \) on the vertices of \( G \), we construct a permutation forest \( F = F_{\pi,G} \subseteq G \) generated by \( G \) and \( \pi \) in the following steps. First we copy all the vertices of \( G \) to \( F \), such that \( V(F) = V(G) \). Then, for each vertex \( v_i \), for \( i = \{1, \ldots, n\} \), an edge between \( v_i \) and \( v_j \) is added to \( F \), where \( j = \max\{k > i : (v_i, v_j) \in E(G)\} \). If for a vertex \( v_i \) there is no such a vertex \( v_j \) such that \( j > i \) and \( (v_i, v_j) \in E(G) \), then no edge will be added to \( F \).

We consider the following topics for the Pick-Last model.

1. Expected size of the largest component
2. Expected maximum tree height

We first compute the probability of two vertices directly connecting to each other.

**Lemma 51.** Let \( \pi = (v_1, \ldots, v_n) \) be a random permutation on the vertices of a given random graph \( G(n, p) \). Denote by \( F = F_{\pi,G} \) the permutation forest generated by Pick-Last process. Then a vertex \( v_i \) in \( F \) attaches directly to its last forward neighbour \( v_j \),
where \( j > i \) in the permutation, with probability \( p(1 - p)^{n-j} \) for \( 0 \leq p \leq 1 \).

**Proof.** Notice that if vertex \( v_i \) connects to \( v_j \) in \( F \), where \( 1 \leq i < j \leq n \), it indicates that there is an edge from \( v_i \) to \( v_j \), but no edge from \( v_i \) to \( v_k \), for \( j + 1 \leq k \leq n \). Since in a random graph \( G(n, p) \) an edge occurs with probability \( p \) uniformly and independently of any other edge, then

\[
\psi_j = P(v_j \text{ is the last forward neighbour of } v_i) = p(1 - p)^{n-j}.
\]

Recall that \( d^+_i \) and \( d^-_i \) are the forward and backward degrees of \( v_i \) in the permutation forest \( F \), respectively, which is introduced in the Pick-First process before Proposition 48.

Before continuing we note an identity relating the expected forward and backward degrees. This differs from the one obtained for the Pick-First model where \( \mathbb{E}[d^+_i] = \mathbb{E}[d^{n-i+1}_n] \) (see Proposition 48).

**Proposition 52.** Let \( F = F_{\pi,G} \) be the permutation forest generated by Pick-Last process from an underlying random graph \( G(n, p) \) and a permutation \( \pi = (v_1, \ldots, v_n) \) on the vertices of \( G \). For a vertex \( v_i \) in \( F \), its forward degree \( d^+_i \) and its backward degree \( d^-_i \) satisfy the following equation:

\[
\mathbb{E}d^+_i + \frac{\mathbb{E}d^-_i}{(i-1)p} = 1.
\]

**Proof.** Since the forward degree \( d^+_i \) of a vertex \( v_i \) can be viewed as an indicator variable which takes value 0 with probability \( P(d^+_i = 0) = (1 - p)^{n-i} \), and 1 with probability \( P(d^+_i = 1) = p + (1 - p)p + \cdots + (1 - p)^{n-i-1}p = 1 - (1 - p)^{n-i} \), then \( \mathbb{E}[d^+_i] = 1 - (1 - p)^{n-i} \).

By Lemma 51 we know that the probability that a vertex \( v_j \) appearing before \( v_i \) in the ordering connects to \( v_i \) is \( p(1-p)^{n-i} \). Notice that \( d^-_i \) follows Binomial distribution \( \text{Bin}(i - 1, p(1-p)^{n-i}) \), then \( \mathbb{E}[d^-_i] = (i-1)p(1-p)^{n-i} \). \(\square\)
5.4.1 Expected size of the largest component in Pick-Last model

Intuitively, due to the Pick-Last rule, it is more likely that the last few vertices are the roots of large components. Since for any vertex we choose its last neighbour appearing after it in the ordering when constructing the forest $F$, the last few vertices have more chance to have more children. For a vertex $v_i$ that appears at position $i$ in the permutation, the probability that $v_i$ connects to root $v_n$ in $F$ is $p$, while the probability that $v_i$ connects to $v_k$ for $i < k < n$ is $p(1-p)^{n-k}$.

Denote by $C_i$ the component rooted at $i$. We analyse the sizes of components rooted at the last few vertices in the ordering. The following Lemma 53 gives the probability that a vertex $v_i$ in the ordering belongs to component $C_j$, where $1 \leq i < j \leq n$, based on which we are able to compute the expected size of $C_j$.

**Lemma 53.** Let $F = F_{\pi,G}$ be the permutation forest generated from a random graph $G(n,p)$ and a permutation $\pi = (v_1, \cdots, v_n)$ on the vertices under Pick-Last rule. Let $Q_i^{(j)}$ be the probability that vertex $v_i$ at position $i$ in the ordering belongs to the component $C_j$ rooted at $j$ in $F$, where $1 \leq i < j \leq n$. Then,

$$Q_i^{(j)} = \prod_{k=i+1}^{j-1} \left[ 1 + p(1-p)^{n-k} \right] p(1-p)^{2(n-j)}.$$

**Proof.** Suppose we reveal the forward edges of each vertex one by one in the (backwards) order $v_j, v_{j-1}, \ldots, v_i$. It is a fundamental observation that given $\pi$, the forward edges of any vertex are disjoint from the forward edges of any other vertex; and thus the distributions of these edges are independent.

We know that $P(v_j \in C_j) = (1-p)^{n-j}$ since if $v_j$ is a root in the forest it fails to connect to all vertices appearing after it in the ordering. Also, $P(v_{j-1} \in C_j) = p(1-p)^{n-j} = \psi_j$ by Lemma 51, since $v_{j-1}$ connect to $v_j$ with probability $p$ in $G(n,p)$ and that $v_{j-1}$ does not have any neighbour appearing after $v_j$ in the ordering. The event that $v_j$ is a root is independent of this. As for the probability
that $v_{j-2} \in C_j$, either $v_{j-2}$ connects to $v_j$ directly and $v_j$ is a root, or it connects to $v_{j-1}$ and $v_{j-1} \in C_j$. Therefore,

$$Q^{(j)}_j = (1 - p)^{n-j} ,$$
$$Q^{(j)}_{j-1} = \psi_j Q^{(j)}_j ,$$
$$Q^{(j)}_{j-2} = \psi_j Q^{(j)}_j + \psi_{j-1} Q^{(j)}_{j-1} = (1 + \psi_{j-1}) \psi_j Q^{(j)}_j .$$

Inductively, the event that $v_i$ belongs to $C_j$ happens if $v_i$ connects to $v_j$ directly and $v_j$ is a root; or $v_i$ connects to $v_{j-1}$ and $v_{j-1} \in C_j$; or $v_i$ connects to $v_{j-2}$ and $v_{j-2} \in C_j$ and so on. Thus, for $1 \leq i \leq j - 2$,

$$Q^{(j)}_i = \psi_j Q^{(j)}_j + \psi_{j-1} Q^{(j)}_{j-1} + \cdots + \psi_{i+1} Q^{(j)}_{i+1} = \prod_{k=i+1}^{j-1} (1 + \psi_k) \psi_j Q^{(j)}_j .$$

By Lemma 51, we know that $\psi_k = p(1 - p)^{n-k}$. The probability of $v_i$ in component $C_j$ is therefore $Q^{(j)}_j = (1 - p)^{n-j}$, $Q^{(j)}_{j-1} = p(1 - p)^{2(n-j)}$, and for $1 \leq i \leq j - 2$,

$$Q^{(j)}_i = \prod_{k=i+1}^{j-1} [1 + p(1 - p)^{n-k}] p(1 - p)^{2(n-j)} . \tag{5.1}$$

Having given the probability that a vertex belongs to a component produced by the permutation forest of a random graph, we turn to providing bounds on the size of largest components.

**Theorem 54.** Denote by $F = F_{\pi,G}$ the permutation forest generated by Pick-Last process from a random graph $G(n, p)$ and a permutation $\pi = (v_1, \cdots, v_n)$ on the vertices. Provided $p = o(1)$, the largest component is rooted at $n$ in expectation. The expected size of $C_n$ is bounded by

$$1 + (n - 2)p(1 - p) \leq E[C_n] \leq 1 + (n - 2)pe .$$
CHAPTER 5. PERMUTATION FORESTS

Proof. Lemma 5.3 gives the probability $Q_i^{(j)}$ that a vertex $v_i$ belongs to a component $C_j$ rooted at $j$. It allows us to compute the expected size of component $C_j$ by considering $j$ many indicator variables, each of which takes value 1 with probability $Q_i^{(j)}$ and 0 otherwise, for $i \in \{1, \ldots, j\}$. Thus, by the linearity of expectation,

$$
E[C_j] = \sum_{i=1}^{j} Q_i^{(j)} = \sum_{i=1}^{j-2} Q_i^{(j)} + Q_{j-1}^{(j)} + Q_j^{(j)}
$$

$$
= \sum_{i=1}^{j-2} \prod_{k=i+1}^{j-1} \left[ 1 + p(1-p)^{n-k} \right] p(1-p)^{2(n-j)} + p(1-p)^{2(n-j)} + (1-p)^{n-j}.
$$

Notice that $E[C_j]$ is maximised when $j = n$. Therefore, the largest component is rooted at $n$. Replacing $j$ by $n$ in Equation (5.1) gives for $1 \leq i \leq n-2$,

$$
P(v_i \in C_n) = Q_i^{(n)} = \prod_{k=1}^{n-i-1} \left[ 1 + p(1-p)^k \right] p.
$$

Thus, the expected size of component $C_n$ is

$$
E[C_n] = \sum_{i=1}^{n} Q_i^{(n)} = Q_1^{(n)} + Q_{n-1}^{(n)} + \sum_{i=1}^{n-2} Q_i^{(n)}
$$

$$
= 1 + p + \sum_{i=1}^{n-2} \prod_{k=1}^{n-i-1} \left[ 1 + p(1-p)^k \right] p.
$$

(5.2)
We first seek an upper bound on the expected size of $C_n$ in the above equation.

\[
E[C_n] \leq 1 + p + p \sum_{i=1}^{n-2} \prod_{k=1}^{n-i-1} \exp\{p(1-p)^k\}
\]

\[
= 1 + p + p \sum_{i=1}^{n-2} \exp\left\{ \sum_{k=1}^{n-i-1} (1-p)^k \right\}
\]

\[
= 1 + p + p \sum_{i=1}^{n-2} \exp\left\{ (1-p) \left[ 1 - (1-p)^{n-i-1} \right] \right\}
\]

\[
= 1 + p + p e^{-p} \sum_{l=2}^{n-1} \exp\left\{ -(1-p)^l \right\}
\]

\[
= 1 + p + p e^{-p} \sum_{l=2}^{n-1} [1 - O((1-p)^l)]
\]

\[
\sim 1 + p + p e^{-p} \left[ (n-2) - \frac{\omega}{p} \right]
\]

\[
\leq 1 + (n-2)pe, \text{ for } p = o(1).
\]

As for the lower bound of the expected size of $C_n$ in Equation (5.2), we apply the inequality $e^{x+y} \leq 1 + x$ for any $|x| < 1$. To see this, we deduce from $1 - x < e^{-x}$.

\[
1 - x < e^{-x}
\]

\[
\implies e^{x} > \frac{1}{1-x} = 1 + \frac{x}{1-x}, \text{ and } 1 + \frac{x}{1-x} < e^{\frac{x}{1-x}}
\]

\[
\implies e^{\frac{x}{1-x}} < 1 - x < e^{-x}.
\]

Substituting this inequality to the expression of the expected size of $C_n$ allows us to obtain a lower bound on the component size.

\[
E[C_n] \geq 1 + p \sum_{i=1}^{n-2} \prod_{k=1}^{n-i-1} \left( 1 + p \cdot e^{-kp/(1-p)} \right)
\]

\[
\geq 1 + p \sum_{i=1}^{n-2} \left( 1 + p \cdot \sum_{k=1}^{n-i-1} e^{-kp/(1-p)} \right)
\]

\[
= 1 + (n-2)p + p^2 \sum_{i=1}^{n-2} \frac{1 - e^{-(n-i-1)p/(1-p)}}{e^{p/(1-p)} - 1}
\]

\[
= 1 + (n-2) \left( p + \frac{p^2}{e^{p/(1-p)} - 1} \right) - \frac{p^2}{(e^{p/(1-p)} - 1)^2} \left( 1 - e^{-(n-2)p/(1-p)} \right).
\]
Note that $1 - \frac{p}{1-p} < e^{p/(1-p)} < 1 - p$, then $\frac{1}{1-p} < e^{p/(1-p)} < \frac{1-p}{1-2p}$. Therefore,

$$E[C_n] \geq 1 + (n - 2) \left( p + \frac{p^2}{p/(1-2p)} \right) - \frac{p^2}{(p/(1-p))^2} \left( 1 - \left( \frac{1-2p}{1-p} \right)^{n-2} \right)$$

$$= 1 + (n - 2)2p(1-p) - (1-p)^2 \left[ 1 - \left( \frac{1-p}{1-p} \right)^{n-2} \right]$$

$$\geq 1 + (n - 2)2p(1-p) - (1-p)^2 \frac{(n-2)p}{1-p}$$

$$= 1 + (n - 2)p(1-p).$$

Hence, the expected size of component $C_n$ is bounded by $1 + (n - 2)p(1-p) \leq E[C_n] \leq 1 + (n - 2)p e$.

5.4.2 Maximum height of trees in Pick-Last model

Before analysing the maximum height of the trees in the permutation forest, we first introduce the following theorem which shows that reasonably many vertices are connected directly to the final block of vertices in the permutation.

**Theorem 55.** Let $F = F_{\pi,G}$ be the permutation forest generated from a random graph $G(n,p)$ and a permutation $\pi = (v_1, \cdots, v_n)$ on the vertices of $G$ under Pick-Last rule. Denote by $X(s)$ the number of vertices that attach directly to the last $s$ vertices in the ordering. Then,

$$E[X(s)] = (n - s) \left[ 1 - (1-p)^s \right].$$

Let $s = (1 + \alpha) \frac{\log n}{p}$, where $\alpha > 0$. For $\omega \log n/n \leq p \leq 1$, the first $(1-o(1))(n-s)$ vertices in the ordering attach directly to the last $s$ vertices in the forest $F$ with probability at least $1 - \exp \{-cn\}$, where $c = c(\alpha) > 0$.

**Proof.** Denote by $X_j$ the number of vertices that attach directly to vertex $v_j$ in the ordering. Note that $X_j$, for $n - s < j \leq n$, is the sum of $(n-s)$ indicator variables, where each variable has value 1 with probability $\psi_j = p(1-p)^{n-j}$ by Lemma 51.
The number of vertices $X(s)$ that attach directly to the last $s$ vertices is therefore

$$EX(s) = \sum_{j=n-s+1}^{n} X_j = \sum_{j=n-s+1}^{n} (n-s)\psi_j = \sum_{j=n-s+1}^{n} (n-s)p(1-p)^{n-j}$$

$$= (n-s)p \cdot \frac{1-(1-p)^s}{1-(1-p)} = (n-s) [1 - (1-p)^s].$$

Let $s = (1 + \alpha)\frac{\log n}{p}$, where $\alpha > 0$. Then, for $\omega \log n/n \leq p \leq 1$, we have $(1 + \alpha) \log n \leq s \leq (1 + \alpha)n/\omega$. Substituting for the above equation we obtain

$$EX(s) \geq (n-s) [1 - e^{-ps}] \geq [n - (1 + \alpha)n/\omega] \left(1 - n^{-(1+\alpha)}\right).$$

Note that if $v_i, v_j < v_k$, the edges from $v_i$ to $v_k$ and $v_j$ to $v_k$ are independent. By applying a Chernoff bound we have, for $\delta = \alpha/(1 + \alpha)$,

$$P(X(s) \leq (1 - \delta)EX(s)) \leq \exp \left\{-\delta^2EX(s)/2\right\} \leq \exp \left\{-cn\right\},$$

where $c = (1 - o(1))[1 - (1 + \alpha)/\omega]^{\alpha^2}/[2(1 + \alpha)^2]$.

Thus, for $\omega \log n/n \leq p \leq 1$, and $(1 + \alpha) \log n \leq s \leq (1 + \alpha)n/\omega$, with probability at least $1 - \exp \left\{-cn\right\}$, where $c = c(\alpha) > 0$, the number of vertices that connect directly to the block of last $s$ vertices in the ordering satisfies $X(s)/(n - s) \to 1$ as $n \to \infty$.

The above theorem gives a relatively simple intuition as to why we should not expect a path of length more than $(1 + \alpha)n/\omega$. Recall that we reveal the edge connections of each vertex from $v_1$ to $v_n$ according to the ordering. Suppose we have exposed the forward edges of the first $n - s$ vertices in the ordering, leaving the last $s$ vertices unrevealed. Note that $(1 + \alpha) \log n \leq s \leq (1 + \alpha)n/\omega$. Since almost all of the first $n - s$ vertices are connected directly to the last $s$ vertices, which gives paths of length 1, then even if the last $s$ vertices form a path (which is very unlikely), the longest path is no more than $1 + s$.

The following lemma gives us a useful tool to bound the sum of some product terms, which is helpful when we compute the probability of existing a path of certain length.
Lemma 56. Consider a product of $k$ variables $x_{a_1} \ldots x_{a_k}$, for any $1 \leq a_1 < \ldots < a_k \leq n$, then the sum of these products satisfy the following bounds.

\[
\left( \sum_{i=1}^{n} x_i \right)^k - \binom{k}{2} \left( \sum_{i=1}^{n} x_i \right)^{k-2} \sum_{i=1}^{n} x_i^2 \leq k! \sum_{1 \leq a_1 < \ldots < a_k \leq n} x_{a_1} \ldots x_{a_k} \leq \left( \sum_{i=1}^{n} x_i \right)^k.
\]

Proof. Consider the Multinomial theorem:

\[
(x_1 + \cdots + x_m)^n = \sum_{k_1 + \cdots + k_m = n} \binom{n}{k_1, \ldots, k_m} \prod_{t=1}^{m} x_t^{k_t}.
\]

Since each of the $k$ variables in the product term $x_{a_1} \ldots x_{a_k}$ is of order 1, then the coefficient is $\binom{k}{1, \ldots, 1} = k!$, yielding the upper bound of the sum of the product terms.

As for the lower bound, we show that

\[
\left( \sum_{i=1}^{n} x_i \right)^k - k! \sum_{1 \leq a_1 < \ldots < a_k \leq n} x_{a_1} \ldots x_{a_k} < \binom{k}{2} \left( \sum_{i=1}^{n} x_i \right)^{k-2} \sum_{i=1}^{n} x_i^2.
\]

Notice that any variable that exists on the left-hand side of the above expression also exists on the right-hand side. Moreover, the coefficient of any term on the left-hand side is less or equal to the coefficient of any term on the right-hand side. To see this, we consider the coefficients from large to small. The largest coefficient $k!$ is subtracted on the left-hand side. The second largest coefficient is $\binom{k}{2}$, which comes from the product terms that contains only one variable of order 2 and all other variables of order 1, such as $x_{a_1}^2 x_{a_2} \cdot x_{a_k}$. Clearly, other coefficients are less than $\binom{k}{2}$, thus the inequality is valid.

Note that the equal case is when $k = 2$, i.e.

\[
\sum_{1 \leq a_1 < a_2 \leq n} x_{a_1} x_{a_2} = \frac{1}{2} \left[ \left( \sum_{i=1}^{n} x_i \right)^2 - \sum_{i=1}^{n} x_i^2 \right].
\]

As we have observed from Theorem 55, it is unlikely to have path of length more than $1 + s$ in the permutation forest $F$ generated by Pick-Last process, where
(1 + \alpha) \log n \leq s \leq (1 + \alpha)n/\omega. We shall provide lower and upper bounds for the longest paths in \( F \).

**Theorem 57.** Let \( F = F_{\pi,G} \) be the permutation forest generated from a random graph \( G(n, p) \) and a permutation \( \pi = (v_1, \cdots, v_n) \) by Pick-Last process. Then, for \( p = o(1/\log n) \) but \( p \geq n^{-\alpha} \) where \( 0 \leq \alpha < 1 \), the expected maximum height \( H \) of trees in \( F \) is bounded by

\[
\log n / \log \log n \leq \mathbb{E}H \leq \omega \log n / \log \log n ,
\]

where \( \omega \to \infty \) arbitrarily slowly.

**Proof.** Determining the maximum height of the trees in the permutation forest is equivalent to finding the longest path to the root in a component. Suppose a path of length \( k \) starting from \( v_i \) is denoted by \( P_i = (v_i, a_1, \cdots, a_k) \) for any vertex \( v_i \) where \( 1 \leq i < n \). The probability that \( P_i \) exists is the product of the probabilities that vertex \( v_i \) is attached to vertex \( a_1 \), and \( a_1 \) is attached to \( a_2 \) and so on until \( a_k \), where the forward connection of \( a_k \) has not been exposed. Recall that Lemma 51 gives the probability that a vertex with index lower than \( j \) attaches directly to \( v_j \), which is \( \psi_j = p(1 - p)^{n-j} \). Therefore,

\[
P(\exists \text{ a } k\text{-path from } v_i) = \sum_{1 \leq i < a_1 < \cdots < a_k \leq n} \psi_{a_1} \psi_{a_2} \cdots \psi_{a_k} . \tag{5.3}
\]

By Lemma 56 we can upper bound the probability as follows.

\[
P(\exists \text{ a } k\text{-path from } v_i) \leq \frac{1}{k!} \left( \sum_{j=2}^{n} \psi_j \right)^k = \frac{1}{k!} \left( \sum_{j=2}^{n} p(1 - p)^{n-j} \right)^k
= \frac{1}{k!} \left[ 1 - (1 - p)^{n-1} \right]^k \leq \left( \frac{e}{k} \right)^k \left[ 1 - (1 - p)^{n-1} \right]^k
\leq \left( \frac{e}{k} \right)^k .
\]

Let \( k = c \log(n p) \) for some variable \( c = c(n) \). Then,

\[
\left( \frac{e}{k} \right)^k = e^{k - k \log k} = \exp\{-c \log(n p) [\log (c \log(n p)) - 1]\}
= (n p)^{-c[\log(c \log(n p)) - 1]} .
\]

161
CHAPTER 5. PERMUTATION FORESTS

We apply a union bound to obtain the expected number of $k$-paths in the permutation forest. Let $p = 1/n^\alpha$, where $0 \leq \alpha < 1$ we have

$$E(\text{number of } k\text{-paths in } F) = n \cdot \mathbb{P}(\exists \text{ a } k\text{-path from vertex } v_i)$$

$$\leq n \cdot (np)^{-c[\log(c \log(np)) - 1]}$$

$$= n^{1-(1-\alpha)f(n,c)} ,$$

where $f(n,c) = c[\log(c(1-\alpha) \log n) - 1]$. Choose $c = \omega / \log \log n$ for $\omega = o(\log \log n)$ but tends to infinity with $n$; for example $\omega = \log \log \log n$. For sufficiently large $n$,

$$f(n,c) = c \left( \log \log n - \log \left[ \frac{e}{c(1-\alpha)} \right] \right)$$

$$= c \left( \log \log n - \log \left[ \frac{e \log \log n}{\omega(1-\alpha)} \right] \right)$$

$$\geq \frac{\omega}{\log \log n} \left( \frac{1}{2} \log \log n \right)$$

$$= \frac{\omega}{2} .$$

Therefore, for $p = 1/n^\alpha$, where $0 \leq \alpha < 1$, there is no path of length $\omega \log n / \log \log n$ with probability $1 - o(1)$.

Now we seek a lower bound on the probability of existing a $k$-path in the permutation forest. Since for the term $\sum \psi_j^2$ we have

$$\sum_{j=2}^{n} \psi_j^2 = \sum_{j=2}^{n} \left[ p(1-p)^{n-j} \right]^2 = \frac{p \left[ 1 - (1-p)^{2(n-1)} \right]}{2 - p} ,$$

162
then by Lemma 56,

$$\mathbb{P}(\exists \text{ a } k\text{-path from } v_i)$$

$$\geq \frac{1}{k!} \left[ \left( \sum_{j=2}^{n} \psi_j \right)^k - \binom{k}{2} \left( \sum_{j=2}^{n} \psi_j \right)^{k-2} \sum_{j=2}^{n} \psi_j^2 \right]$$

$$= \frac{1}{k!} \left\{ \left[ 1 - (1 - p)^{n-1} \right]^k - \frac{k(k-1)}{2} \left[ 1 - (1 - p)^{n-1} \right]^{k-2} p \left[ 1 - (1 - p)^{2(n-1)} \right] \right\}$$

$$\geq \frac{1}{k!} \left[ 1 - (1 - p)^{n-1} \right]^{k-1} \left\{ \left[ 1 - (1 - p)^{n-1} \right] - \frac{pk^2}{2(2 - p)} \cdot \left[ 1 + (1 - p)^{n-1} \right] \right\}$$

$$= \frac{1}{k!} \left[ 1 - (1 - p)^{n-1} \right]^{k-1} \left\{ \left( 1 + \frac{pk^2}{2(2 - p)} \right) \left[ 1 - (1 - p)^{n-1} \right] - \frac{pk^2}{2 - p} \right\}$$

$$= \frac{1}{k!} \left[ 1 - (1 - p)^{n-1} \right]^{k} \left\{ 1 - \left( \frac{2}{1 - (1 - p)^{n-1} - 1} \right) \frac{pk^2}{2(2 - p)} \right\}. \quad (5.4)$$

Note that for any \(|x| < 1\), we have the following Taylor expansion:

$$\frac{x}{1 - x} = \frac{1}{1 - x} - 1 = x + x^2 + \cdots = \sum_{i=1}^{\infty} x^i. \quad (5.5)$$

As \(n^{-\alpha} \leq p \leq 1/(\omega \log n)\), \((1 - p)^{n-1} = O(\epsilon^{n^{1-\alpha}}) = o(1)\). Thus

$$\frac{p}{2 - p} = \frac{p}{2} + \frac{p^2}{4} + \cdots = O \left( \frac{p}{2} \right), \quad \text{and}$$

$$\frac{1}{1 - (1 - p)^{n-1}} = 1 + (1 - p)^{n-1} + (1 - p)^{2(n-1)} + \cdots = 1 + O \left( (1 - p)^{n-1} \right).$$

Substituting for Inequality (5.4) with the above approximation, we obtain the following upper bound on the probability of existing a \(k\)-path.

$$\mathbb{P}(\exists \text{ a } k\text{-path from } v_i)$$

$$\geq \frac{1}{k!} \left[ 1 - (1 - p)^{n-1} \right]^{k} \left\{ 1 - \left[ 1 + O \left( (1 - p)^{n-1} \right) \right] O \left( \frac{p}{2} \right) \frac{k^2}{2} \right\}$$

$$\geq \frac{1}{k!} \left[ 1 - (1 - p)^{n-1} \right]^{k} \left\{ 1 - O \left( \frac{p^2}{4} \right) k^2 \right\}. \quad (5.5)$$

Notice that if \(p = o(1/\log n)\) but \(p \geq n^{-\alpha}\), where \(0 \leq \alpha < 1\), we have \(pk^2 = o(1)\). Since \([1 - (1 - p)^{n-1}]^k \geq [1 - e^{-(n-1)p}]^k \geq 1 - ke^{-np} \geq 1 - ke^{-(1-\alpha)n}\), and \(ke^{-(1-\alpha)n} = \frac{1}{2} \cdot 1 = \frac{1}{2}\).
o(1) for \( k = O(\log(np)) \), then we obtain \([1 - (1 - p)^{n-1}]^k \geq 1 - o(1)\). Therefore, with the above approximation, the expression of the probability that there exists a path of length \( k \) in the Inequality (5.5) becomes

\[
P(\exists \text{ a } k\text{-path from } v_i) \geq \frac{1}{k!} (1 - o(1)).
\]

Let \( k = \frac{\log(np)}{\log \log(np)} \) for \( p = o(1/\log n) \) but \( p \geq n^{-\alpha} \) where \( 0 \leq \alpha < 1 \). Applying a union bound on the above probability and using the Stirling approximation yield

\[
E(\text{number of } k\text{-paths in } F) = n \cdot P(\exists \text{ a } k\text{-path from vertex } i) \\
\geq \frac{n}{k!} (1 - o(1)) = \frac{n}{\sqrt{2\pi k}} \left(\frac{e}{k}\right)^k (1 - o(1)) \\
= \frac{n (1 - o(1))}{\sqrt{2\pi k} \cdot e^{k(\log k - 1)}} \geq \frac{n (1 - o(1))}{e^{k \log k}} \\
= \frac{n (1 - o(1))}{(np)^{1 - \log \log \log(np) / \log \log(np)}} \\
\geq n^{1 - (1 - \alpha)[1 - O(1/\log \log(np))] - \alpha O(1/\log \log(np))} (1 - o(1)) \\
\geq n^\alpha (1 - o(1)) \to \infty.
\]

Hence, for \( p = o(1/\log n) \) but \( p \geq n^{-\alpha} \) where \( 0 \leq \alpha < 1 \), the expected number of paths of length \( \log n / \log \log n \) in the permutation forest is \( \Omega(n^\alpha) \).

5.5 Pick-UAR model

In this section, we contribute to the understanding of Pick-UAR process which is another model for generating permutation forest apart from Pick-First and Pick-Last models. We will also provide analytical results of various properties of the forest generated from an underlying random graph with a given random ordering on the vertices. Given a graph \( G \), we produce our permutation forest under the Pick-UAR rule by first randomly ordering the vertices of \( G \), and then copying all vertices of \( G \) to
the forest. As for the edges to present in the forest, for each vertex according to the
ordering, we examine its neighbours and select an edge uniformly at random which
connects it to another vertex appearing after it in the ordering, if such an edge exists.
After selecting these edges for each vertex, all other edge are discarded. Therefore,
for each vertex at most one ‘forward’ edge form the underlying graph is preserved in
the forest.

Formally we give the following definition of the process. Let $G(n, p) = G(V, E)$
be a random graph with $|V| = n$. Let $\pi = (v_1, \ldots, v_n)$ be an ordering of $V$, i.e.
a permutation of the vertices of $G$. Define a permutation forest $F = F_{\pi,G} \subseteq G$ as
follows. Let $V(F) = V(G)$. For each $v_i$, $i = \{1, \ldots, n\}$, we add an edge between
$v_i$ and $v_j$, where $j > i$ and $v_j$ is chosen uniformly at random from the set of $v_i$’s
neighbours such that $j \in \{k > i : (i, k) \in E(G)\}$. If there is no edge between $v_i$ and
vertices after $v_i$ in the ordering, we do not add an edge. We say such a vertex ‘fails
to connect’.

We consider the following topics for the Pick-UAR model.
1. Expected size of the largest component
2. Expected maximum height of trees

5.5.1 Expected size of the largest component in Pick-UAR
model

We first compute the probability that vertex $v_i$ connects directly to vertex $v_j$ in the
forest, where $j > i$ in the ordering and $v_j$ is chosen uniformly at random among all
the forward neighbours of $v_i$ in the underlying random graph.

Lemma 58. Let $\pi = (v_1, \ldots, v_n)$ be a random permutation on the vertices of a
random graph $G(n, p)$. Then in the permutation forest $F = F_{\pi,G}$ generated by the
Pick-UAR process, a vertex $v_i$ attaches directly to a uniformly-at-random selected
forward neighbour $v_j$, where $j > i$, with probability $[1 - (1 - p)^{n-i}] / (n-i)$ for $0 \leq p \leq 1$.

Proof. For any $1 \leq i < j \leq n$, let $\psi_i$ be the probability that a vertex $v_i$ attaches directly to vertex $v_j$ in $F$ generated by Pick-UAR process. Note that $v_j$ is chosen uniformly at random among all the neighbours appearing after $v_i$ in the ordering, and the probability that the edge between $v_i$ and $v_j$ exists in the underlying random graph $G$ is $p$.

The probability that $v_j$ is a neighbour of $v_i$, and $v_i$ has exactly $k$ many other neighbours which appear after $v_i$ in the ordering, apart from $v_j$ in the underlying graph $G$, is $p \cdot \sum_{k=0}^{n-i-1} \left( \frac{n-i-1}{k} \right) p^k (1-p)^{n-i-1-k}$. Given this, the probability that $v_i$ chooses $v_j$ among its $k + 1$ forward neighbours is $1/(k+1)$. Thus we have

$$\psi_i = \mathbb{P}(v_i \text{ attaches directly to } v_j, j > i)$$

$$= p \cdot \sum_{k=0}^{n-i-1} \left( \frac{n-i-1}{k} \right) p^k (1-p)^{n-i-1-k} \cdot \frac{1}{k+1}$$

$$= \sum_{k=0}^{n-i-1} \frac{(n-i-1)!}{(k+1)! \cdot (n-i-1-k)!} p^{k+1} (1-p)^{n-i-1-k}$$

$$= \sum_{l=1}^{n-i} \frac{(n-i-1)!}{l! \cdot (n-i-l)!} p^l (1-p)^{n-i-l}$$

$$= \frac{1}{n-i} \sum_{l=1}^{n-i} \left( \frac{n-i}{l} \right) p^l (1-p)^{n-i-l}$$

$$= \frac{1}{n-i} \left[ ((1-p) + p)^{n-i} - (1-p)^{n-i} \right]$$

$$= \frac{1}{n-i} \left[ 1 - (1-p)^{n-i} \right].$$

We next give bounds on the expected size of the largest component in the permutation forest.

Theorem 59. Let $F = F_{\pi,G}$ be the permutation forest generated by Pick-UAR process from a random graph $G(n,p)$ and a permutation $\pi = (v_1, \cdots, v_n)$ on the vertices.
CHAPTER 5. PERMUTATION FORESTS

Provided \( p = o(1) \), the largest component is rooted at \( n \) in expectation. The expected size \( C_n \) is bounded by

\[
1 + e^{-\pi^2/12(n - 2 - \log p)p} \leq E[C_n] \leq 1 + (n - 2)p e^p.
\]

Proof. Using the similar method as in the proof of Lemma 53, we let \( Q_i^{(j)} \) be the probability that vertex \( v_i \) belongs to the component \( C_j \) rooted at \( j \) in \( F \), where \( 1 \leq i \leq j \leq n \). We know that \( Q_j^{(j)} = (1 - p)^{n-j} \) as it equals to the probability that \( j \) is a root. Also, the probability \( Q_{j-1}^{(j)} \) that \( v_{j-1} \) belongs to \( C_j \) is the probability that \( v_{j-1} \) directly connects to \( v_j \) and \( v_j \) is a root. Similarly, the probability \( Q_{j-2}^{(j)} \) equals to the probability that \( v_{j-2} \) connects directly to \( v_j \) and \( v_j \) is a root plus the probability that \( v_{j-2} \) connects to \( v_{j-1} \) and \( v_{j-1} \) belongs to \( C_j \). Recall that the probability that \( v_i \) connects to \( v_j \) directly where \( i < j \) is \( \psi_i = \frac{1}{n-i} [1 - (1-p)^{n-i}] \) by Lemma 58. The event that \( v_j \) is a root is independent of this. Therefore,

\[
Q_j^{(j)} = (1 - p)^{n-j}
\]

\[
Q_{j-1}^{(j)} = \psi_{j-1}Q_j^{(j)}
\]

\[
Q_{j-2}^{(j)} = \psi_{j-2}Q_j^{(j)} + \psi_{j-2}Q_{j-1}^{(j)} = \psi_{j-2}(1 + \psi_{j-1})Q_j^{(j)}.
\]

Inductively, the probability \( Q_i^{(j)} \) equals to the probability that \( v_i \) connects directly to \( v_j \) and \( v_j \) is a root, plus the probability that \( v_i \) connects to \( v_{j-1} \) and \( v_{j-1} \) belongs to \( C_j \), plus the probability that \( v_i \) connects to \( v_{j-2} \) and \( v_{j-2} \) belongs to \( C_j \) and so on, till \( v_i \) connects to \( v_{i+1} \) and \( v_{i+1} \) belongs to \( C_j \). Therefore, for \( 1 \leq i \leq j-2 \),

\[
Q_i^{(j)} = \psi_iQ_j^{(j)} + \psi_iQ_{j-1}^{(j)} + \cdots + \psi_iQ_{i+1}^{(j)} = \psi_i(1 + \psi_{i+1}) \cdots (1 + \psi_{j-1})Q_j^{(j)}
\]

\[
= \psi_iQ_j^{(j)} \prod_{k=i+1}^{j-1} (1 + \psi_k).
\]

(5.6)

Notice that the probability \( \psi_i \) that \( v_i \) directly connects to \( v_j \) where \( j > i \) is independent of \( j \), and \( Q_n^{(n)} \) maximises \( Q_j^{(j)} \) and the product in Equation (5.6). Then \( Q_i^{(j)} \) is maximised when \( j = n \). Therefore, the largest component is in \( F \) in expectation.
CHAPTER 5. PERMUTATION FORESTS

is $C_n$. The expected size of $C_n$ can be computed as follows.

$$
E[C_n] = 1 + p + \sum_{i=1}^{n-2} Q_i^{(j)} = 1 + p + \sum_{i=1}^{n-2} \psi_i \prod_{k=i+1}^{n-1} (1 + \psi_k).
$$

We first seek an upper bound on the expected size of $C_n$. Note that for any $1 \leq k < n$, $\psi_k \leq 1/(n-k)[1 - (1 - (n-k)p)] = p < 1$; and that

$$
\sum_{k=i+1}^{n-1} \psi_k = \sum_{k=i+1}^{n-1} \frac{1}{n-k} - \sum_{k=i+1}^{n-1} \frac{(1-p)^{n-k}}{n-k} = H_{n-i-1} + \left[ -\log p - O\left(\frac{(1-p)^{n-i}}{n-i}\right) \right] = \log(1 - (1-p)^{n-i}/(n-i)).
$$

Hence,

$$
E[C_n] \leq 1 + p + \sum_{i=1}^{n-2} \psi_i \prod_{k=i+1}^{n-1} e^{\psi_k} = 1 + p + \sum_{i=1}^{n-2} \psi_i e^{\sum_{k=i+1}^{n-1} \psi_k}
$$

$$
= 1 + p + \sum_{i=1}^{n-2} \psi_i [(n-i-1)p \cdot e^{1-p}]
$$

(since $O\left(\frac{(1-p)^{n-i}/(n-i)}{(n-i)}\right) \leq O\left(\frac{(1-p)^2}{2} \leq 1 - p\right)$)

$$
\leq 1 + p + pe^{1-p} \sum_{i=1}^{n-2} [1 - (1-p)^{n-i}]
$$

$$
= 1 + p + pe^{1-p} \left( (n-2) - \frac{1-p}{p} [1 - (1-p)^{n-2}] \right)
$$

$$
= 1 + p + e^{1-p}(n-2)p - e^{1-p}(1-p)(1-o(1))
$$

$$
\leq 1 + (n-2)pe, \text{ for } p = o(1).
$$

As for the lower bound of $E[C_n]$, we use the following inequality:

$$
1 + x = \exp\{\log(1 + x)\}
$$

$$
= \exp\{x - x^2/2 + x^3/3 - x^4/4 + \cdots\}
$$

$$
= \exp\{x - x^2/2 + x^3(1 - 3x/4) + \cdots\}
$$

$$
\geq \exp\{x - x^2/2\}.
$$
Let $x = \psi_k = 1/(n-k)[1 - (1-p)^{n-k}]$, then $x^2 \leq 1/(n-k)^2$. Therefore,

$$\prod_{k=i+1}^{n-1} (1 + \psi_k) \geq \prod_{k=i+1}^{n-1} \exp \left\{ \psi_k - \frac{1}{2(n-k)^2} \right\}$$

$$= \exp \left\{ \sum_{k=i+1}^{n-1} \psi_k \right\} \cdot \exp \left\{ - \sum_{k=i+1}^{n-1} \frac{1}{2(n-k)^2} \right\}$$

$$\geq \exp \left\{ \sum_{k=i+1}^{n-1} \psi_k \right\} \cdot \exp \left\{ - \pi^2/12 \right\} \quad \text{(since } \sum_{k=1}^{\infty} 1/k^2 = \pi^2/6 \text{)} ,$$

Recall that $\sum_{k=i+1}^{n-1} \psi_k = \log[(n-i-1)p] + o(1) \geq \log[(n-i-1)p] \approx \log[(n-i)p]$ for $n \to \infty$. Then, $\prod_{k=i+1}^{n-1} (1 + \psi_k) \geq e^{-\pi^2/12(n-i)p}$, and thus

$$E[C_n] \geq 1 + p + e^{-\pi^2/12} p \sum_{i=1}^{n-2} \psi_i(n-i)$$

$$= 1 + p + e^{-\pi^2/12} p \sum_{i=1}^{n-2} [1 - (1-p)^{n-i}]$$

$$= 1 + p + e^{-\pi^2/12} p \{(n-2) - [- \log p - O(1-p)]\}$$

$$(\text{note that for } |x| < 1, \log(1-x) = x + x^2 + x^3 + \cdots)$$

$$\geq 1 + e^{-\pi^2/12} (n - 2 + \log p)p .$$

$\square$

### 5.5.2 Maximum height of trees in Pick-UAR model

The following theorem gives bounds on the expected length of the longest paths in the permutation forest.

**Theorem 60.** Let $F = F_{\pi,G}$ be the permutation forest generated by Pick-UAR process from a random graph $G(n, p)$ and a permutation $\pi = (v_1, \cdots, v_n)$ on the vertices. Then, for $p = o(1/\log n)$ but $p \geq n^{-\alpha}$ where $0 \leq \alpha < 1$, the expected maximum height $H$ of trees in $F$ is bounded by

$$\log n/\omega \leq EH \leq 2e \log n ,$$

where $\omega = o(\log n) \to \infty$ arbitrarily slowly.
Proof. We will find the longest path in a component and argue that the probability of the existence of a path longer than $O(\log np)$ tends to 0 as $n$ increases. For any vertex $1 \leq u < n$, suppose a path of length $k$ from $u$ is denoted by $P_u = (u, a_1, \ldots, a_k)$. The probability that $P_u$ exists equals to the product of the probabilities that $u$ is attached to $a_1$, and $a_1$ is attached to $a_2$, and $a_2$ is attached to $a_3$, and so on until $a_k$. Note that we do not expose the forward connection of $a_k$. According to Lemma 58, the probability that $v_i$ connects directly to a vertex $v_j$, for $1 \leq i < j \leq n$, is
\[ \psi_i = \frac{[1 - (1 - p)^{n-i}]}{(n-i)} . \]

Therefore,
\[ P(\exists \text{ a } k\text{-path from } u) = \sum_{1 \leq u < a_1 < \ldots < a_k < n} \psi_u \psi_{a_1} \psi_{a_2} \cdots \psi_{a_{k-1}} . \] (5.7)

In order to evaluate the above probability, we will give upper and lower bounds using Lemma 56. We compute the term $\sum \psi_i$ first as it appears on both sides of the inequality in Lemma 56.
\[ \sum_{i=1}^{n-1} \psi_i = \sum_{i=1}^{n-1} \frac{1}{n-i} [1 - (1 - p)^{n-i}] = \sum_{j=1}^{n-1} \frac{1}{j} - \sum_{j=1}^{n-1} \frac{(1-p)^j}{j} . \]
Notice that the term $\sum_{j=1}^{n-1} 1/j$ is the $(n-1)$-th Harmonic number $H_{n-1} \approx \log n$. As for the second term, we know that $\log(1-x) = - \sum_{i=1}^{\infty} x^i/i$, therefore
\[ \sum_{j=1}^{n-1} \frac{(1-p)^j}{j} = - \log p + O\left(\frac{(1-p)^n}{n}\right) \approx - \log p . \]
Substitute with the approximations and we get
\[ \sum_{i=1}^{n-1} \psi_i = H_{n-1} + \log p - O\left(\frac{(1-p)^n}{n}\right) \approx \log(np) . \] (5.8)

We begin with an upper bound using Lemma 56.
\[ P(\exists \text{ a } k\text{-path from } u) \leq \frac{1}{k!} \left( \sum_{i=1}^{n-1} \psi_i \right)^k \text{ (by Lemma 56) } \]
\[ \approx \frac{1}{k!} [\log(np)]^k \text{ (by Equation (5.8)) } \]
\[ \leq \left( \frac{e}{k} \right)^k [\log(np)]^k = \left( \frac{e \log(np)}{k} \right)^k . \] (5.9)
Let $k = c \log(np)$ for some variable $c = c(n)$. Then by the linearity of expectation and Inequality (5.9) we know that the expected number of $k$-paths in $F$ is

$$E[\text{number of } k\text{-paths}] = n \cdot P(\exists \text{ a } k\text{-path from } u) \leq n \left( \frac{e \log(np)}{k} \right)^k = \frac{n}{(np)^c \log c}. $$

It is required that $p$ is not too small to ensure that this probability tends to 0 as $n$ increases to infinity. Therefore, if $p = o(1/\log n)$ but $p \geq n^{-\alpha}$ where $0 \leq \alpha < 1$,

$$E[\text{number of } k\text{-paths}] = \frac{n}{(np)^c \log c} = n^{-(1-\alpha)c \log c} = o(1).$$

Hence, for any $p = o(1/\log n)$ but $p \geq n^{-\alpha}$ where $0 \leq \alpha < 1$, there is no path longer than $\Omega(\log(np))$ with probability at least $1 - o(1)$.

Now, we seek a lower bound on the probability (5.7) that there exists a $k$-path. By Lemma 56 , we obtain

$$P(\exists \text{ a } k\text{-path from } u) \geq \frac{1}{k!} \left[ \left( \sum_{i=1}^{n-1} \psi_i \right)^k - \binom{k}{2} \left( \sum_{i=1}^{n-1} \psi_i \right)^{k-2} \left( \frac{1}{n} \right) \sum_{i=1}^{n-1} \psi_i^2 \right]. $$

(5.10)

Recall that $\sum_{i=1}^{n-1} \psi_i \approx \log(np)$, and for $\sum_{i=1}^{n-1} \psi_i^2$ we have

$$\sum_{i=1}^{n-1} \psi_i^2 = \sum_{i=1}^{n-1} \left( \frac{1}{n-i} \left[ 1 - (1-p)^{n-i} \right] \right)^2 = \sum_{j=1}^{\infty} \frac{1}{j^2} - 2 \sum_{j=1}^{n-1} \frac{(1-p)^j}{j} + \sum_{j=1}^{n-1} \frac{(1-p)^{2j}}{j}$$

$$\approx \frac{\pi^2}{6} + 2 \log[1 - (1-p)] - \log[1 - (1-p)^2]$$

$$= \frac{\pi^2}{6} + 2 \log p - \log (2p - p^2)$$

$$= \frac{\pi^2}{6} - \log \left( \frac{2-p}{p} \right) \leq \frac{\pi^2}{6}. $$
Therefore, by substituting for Inequality (5.10) we obtain the following bound.

\[
\Pr(\exists \text{ a } k\text{-path from } u) \geq \frac{1}{k!} \left[ \log^k(np) - \frac{k(k-1)}{2} \log^{k-2}(np) \frac{\pi^2}{6} \right]
= \frac{1}{k!} \log^k(np) \left[ 1 - \frac{\pi^2 k^2}{12 \log^2(np)} \right]
\geq \frac{1}{e \sqrt{k}} \left( \frac{e \log(np)}{k} \right)^k \left[ 1 - \frac{\pi^2 k^2}{12 \log^2(np)} \right].
\]

(5.11)

Let \( k = \log(np)/\omega \), where \( \omega = o(\log n) \) but tends to infinity with \( n \); for instance \( \omega = \log \log n \). For sufficiently large \( n \), the probability (5.11) that there exists a \( k \)-path from a vertex becomes

\[
\Pr(\exists \text{ a } k\text{-path from } u) \geq \frac{(\log(np))^k}{e^{k \log k}} \left( 1 - \frac{\pi^2}{12 \omega^2} \right)
\geq (np)^{\log \omega/\omega} (1 - o(1)).
\]

By the linearity of expectation we obtain for \( p = o(1/\log n) \) but \( p \geq n^{-\alpha} \) where \( 0 \leq \alpha < 1 \), the expected number of \( k \)-paths is

\[
\mathbb{E}(\text{number of } k\text{-paths in } F) = n \cdot \Pr(\exists \text{ a } k\text{-path from } u)
\geq n^{1-(1-\alpha) \log \omega/\omega} (1 - o(1)) \to \infty.
\]

\[\square\]

### 5.6 Experimental results for Permutation Forest models

For given \( p \), the three models have the same expected number of components which is asymptotic to \( 1/p \). The size and structure of the trees in the forest differs between the models; see Tables 5.4 and 5.5. We discuss the experimental outputs on the following topics.

1. Permutation forests around \( p = 1/\sqrt{n} \)
2. Largest component as a function of \( p \)
3. Maximum tree height as a function of \( p \)
5.6.1 Permutation forests around $p = 1/\sqrt{n}$

It was proved in [36] that the Pick-First model has a threshold for the emergence of a giant component, a component of size $\Theta(n)$, at $p = \sqrt{c/n}$. We compare tree structure for the three models at $p = 1/\sqrt{n}$, to provide a snapshot of the three processes at the critical point. The tree sizes are only plotted from the earliest vertex (in the ordering) to join the tree. Thus in the middle plot of Figure 5.6, at $p = 1/\sqrt{n}$ only three components begin to grow near the beginning (of the ordering), and two other large components start later (at $v_i \sim 2000$ and $v_i \sim 3000$ in the ordering). Note that $x$-axis shows the indices of the vertices in the ordering, and $y$-axis shows the component sizes. The Pick-UAR and Pick-Last processes have many components which grow from near the beginning, and thus differ substantially from Pick-First at $p = 1/\sqrt{n}$. In Theorem 55, we proved that in the Pick-Last process, almost all the first $n - s$ vertices in the ordering directly connect to the last $s$ vertices with high probability, where $s = O(\log n/p)$. The experiment (see the right-hand side picture of Figure 5.7) is consistent with this result, but does not directly confirm this.

![Figure 5.6: Pick-First process at $p = \sqrt{c/n}$ for $c = 1/5$, 1, 5 and $n = 10,000$.](image-url)
CHAPTER 5. PERMUTATION FORESTS

Figure 5.7: Pick-UAR and Pick-Last processes at \( p = 1/\sqrt{n} \) and \( n = 10,000 \).

5.6.2 Largest component as a function of \( p \)

For convenience of the reader we reproduce the table of theoretical results on maximum component size (Table 5.4) below.

<table>
<thead>
<tr>
<th>Model</th>
<th>Max. component size</th>
<th>Range of validity</th>
<th>Where proved:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pick-First</td>
<td>( \Theta(n) )</td>
<td>( \omega/\sqrt{n} \leq p \leq 1 )</td>
<td>In [36] and Theorem 49</td>
</tr>
<tr>
<td>Pick-Last</td>
<td>( 1 + \Theta(np) )</td>
<td>( 0 \leq p &lt; 1 )</td>
<td>Theorem 54</td>
</tr>
<tr>
<td>Pick-UAR</td>
<td>( 1 + \Theta(np) )</td>
<td>( 0 \leq p &lt; 1 )</td>
<td>Theorem 59</td>
</tr>
</tbody>
</table>

Figure 5.8: Order of magnitude results for maximum component size.

**Pick-First Model**  The emergence of a giant component in the Pick-First model was studied by Cooper, Kang, McDowell, Radzik and Vu [36]. The emergence of a unique giant component in this model seems well-behaved even on graphs as small as \( n = 1,000 \).
Figure 5.9: Emergence of component of size about $n$ in Pick-First model for $n = 1000$. The vertical line is at $p = 1/\sqrt{n}$.

The following figures give a closer look at the threshold around $p = \sqrt{c/n}$. The value $p = \sqrt{1/n}$ seems reasonably central to this.

Figure 5.10: Details of emergence of component of size about $n$ in Pick-First model for $n = 1000$ and $n = 100,000$. The vertical line is at $p = 1/\sqrt{n}$.

Figure 5.11 (taken from [36]) plots the results of simulations run for 5000 distinct choices of $p$ on $n = 10^5$ vertices, modelling $F$ obtained by Pick-First protocol and showing the sizes of the two largest components, in blue and red respectively. We
can see that there is a clear steep increase in the largest component size, plateauing after passing the threshold for linearity.

Figure 5.11: Sizes of largest (blue) and the second largest (red) component as a function of $p$, with $n = 10^5$. Note: $p = 1/\sqrt{n}$ corresponds to $p = 0.0032$.

Pick-Last and Pick-UAR Models Figure 5.12 illustrates the simulations of the largest component size as a function of $p$ in Pick-UAR and Pick-Last processes, on $n = 10^4$ and $n = 10^5$ respectively. There appears to be a large interval $[0, p_0]$ where the size does not appear to be concentrated, with a wide variance in the values occurring. Currently we do not have any theoretical way to determine $p_0$ which seems to be about 0.8. Visually, the plots do not contradict a largest component of expected size $\Theta(np)$ for $p < p_0$. 

176
Figure 5.12: Size of largest component as a function of $p$. For UAR, $n = 10,000$, for Last, $n = 100,000$.

### 5.6.3 Maximum tree height as a function of $p$

For convenience of the reader we reproduce the table of theoretical results on maximum path lengths (Table 5.5) below.

<table>
<thead>
<tr>
<th>Model</th>
<th>Max. tree height $H$</th>
<th>Range of validity</th>
<th>Where proved:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pick-First</td>
<td>$H = \Omega(np)$</td>
<td>$\omega/n \leq p \leq 1$</td>
<td>Corollary 47</td>
</tr>
<tr>
<td>Pick-Last</td>
<td>$\log n/\log \log n \leq EH$, but $EH \leq \omega \log n/\log \log n$</td>
<td>$n^{-\alpha} \leq p \leq o(1/\log n)$, $\alpha &lt; 1$</td>
<td>Theorem 57</td>
</tr>
<tr>
<td>Pick-UAR</td>
<td>$\log n/\omega \leq EH$, but $EH \leq 2e \log n$</td>
<td>$n^{-\alpha} \leq p \leq o(1/\log n)$, $\alpha &lt; 1$</td>
<td>Theorem 60</td>
</tr>
</tbody>
</table>

Figure 5.13: Order of magnitude results for maximum tree height $H$ in the three models.

**Pick-First Model**  The longest path length in this model seems to be a linear function of $path = n$ of the number of vertices. Currently we do not have an explanation of this, except in the case $p = 1$ where the forest is a Hamilton path from vertex 1 to vertex $n$.  

177
CHAPTER 5. PERMUTATION FORESTS

Figure 5.14: Size of longest path in Pick-First as a function of $p$, for $n = 100,000$.

**Pick-Last and Pick-UAR Models** As the results here are only order of magnitude ones for the expected maximum path length, it is difficult to draw meaningful conclusions. The expected lengths are respectively, $\log n / \log \log n$ (Last) and $\log n$ (UAR) which further increases the difficulty. The results presented below do tend to confirm the theoretical findings, but only in outline. The simulations are time consuming taking from 6 to 12 hours for the largest case, and are not particularly consistent.

Figure 5.15: Size of longest path in Pick-Last model as a function of $p$, for $n = 10,000$, $n = 100,000$ and $n = 1,000,000$. The horizontal line is at $\log n / \log \log n$. 

178
Figure 5.16: Size of longest path in Pick-UAR model as a function of $p$, for $n = 1,000$, $n = 10,000$ and $n = 100,000$. The horizontal line is at $2 \log n$. 

Bibliography


[37] Colin Cooper, Tomasz Radzik, and Nicolás Rivera. Improved cover time bounds for the coalescing-branching random walk on graphs. In *Proceedings of the 29th


[105] Hiroto Yasumi, Fukuhito Ooshita, and Michiko Inoue. Uniform partition in population protocol model under weak fairness. In Pascal Felber, Roy Friedman, Seth Gilbert, and Avery Miller, editors, *23rd International Conference*