STOCHASTIC BLOCK MODEL DYNAMICS

by

Nithish Kumar Kumar

A Dissertation

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the degree of

Master of Science

Department of Computer Science

West Lafayette, Indiana

May 2021

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL

Dr. Simina Brânzei, Chair
School of Computer Science

Dr. Elena Grigorescu
School of Computer Science

Dr. Alex Psomas
School of Computer Science

Approved by:
Dr. Kihong Park

To my family.

ACKNOWLEDGMENTS

I wish to convey my heartfelt gratitude to my advisor, Prof. Simina Brânzei for her guidance and patience over the past two years and especially during my work on this thesis. Her support and advice during my research, especially in circumstances where I was challenged by the depth of my research problem were invaluable. I am also thankful to my committee members Prof. Elena Grigorescu and Prof. Alex Psomas, for their guidance during my work on this thesis.

I thank the Computer Science Department of Purdue University and Prof. Simina Brânzei for funding me the past two years. I am also grateful to Prof. Gireeja Ranade for collaborating with us on this project. Her insights and comments were invaluable for us on several occasions. I thank Prof. Ben Seamone for guiding me through the foundations of research as part of my undergraduate internship. I wouldn't have been able to work on this project without the experience I gained from working with him.

I am incredibly grateful to my parents, Lakshmi, and Kumar for supporting me through thick and thin. I would not be where I am now without their unwavering love, affection, and care. Lastly, I wish to express my indebtedness to my brother Harish, for being a guide, teacher, and most of all a friend throughout my life. His inputs have been crucial for every major step of my life and have helped me see the better path on several crucial junctures of my life.

TABLE OF CONTENTS

ABSTRACT

The past few years have seen an increasing focus on fairness and the long-term impact of algorithmic decision making in the context of Machine learning, Artificial Intelligence and other disciplines. In this thesis, we model hiring processes in enterprises and organizations using dynamic mechanism design. Using a stochastic block model to simulate the workings of a hiring process, we study fairness and long-term evolution in the system.

We first present multiple results on a deterministic variant of our model including convergence and an accurate approximate solution describing the state of the deterministic variant after any time period has elapsed. Using the differential equation method, it can be shown that this deterministic variant is in turn an accurate approximation of the evolution of our stochastic block model with high probability.

Finally, we derive upper and lower bounds on the expected state at each time step, and further show that in the limiting case of the long-term, these upper and lower bounds themselves converge to the state evolution of the deterministic system. These results offer conclusions on the long-term behaviour of our model, thereby allowing reasoning on how fairness in organizations could be achieved. We conclude that without sufficient, systematic incentives, under-represented groups will wane out from organizations over time.

1. INTRODUCTION

With the growth of the importance of Algorithmic decision making in everyday life through Machine learning, Artificial Intelligence and other disciplines, there has been an increasing emphasis on the fairness of the algorithms used for making these decisions. An important area where fairness is critical, is in hiring policies - they deeply affect the dynamics of most organizations and consequently, society as a whole.

The traditional under-representation of various under-privileged communities in business and other organizations results in a pernicious feedback loop when biased hiring practices are included into the situation. Hiring processes are often driven by prior professional or personal contacts, and this results in a significant reduction of upward mobility amongst such communities.

In addition, it is also essential to examine any policy implemented to promote fairness in terms of whether it establishes fairness in the long term. Liu et al. [1] demonstrates that static fairness criteria do not actually produce lasting improvement over the long-term. They show that for classification systems (e.g. for loan eligibility), static constraints introduced to promote fairness can lead to a worsening of the disadvantaged group's situation in the long term. Thus, when examining any dynamic, it is necessary to look at it in the context of its long term behaviour.

The stochastic block model is a generative model for random graphs. The nodes in these graphs are split into communities and occurrences of connections/collaborations (edges) between any two nodes are dependent on the community they belong to. This is relevant for our purposes for it can account for the different likelihoods of two people working together, given which communities they belong to. For instance, considering we have two communities A and B, the existence of a similarity bias can be encoded in such a model using a higher probability for the occurrence of $A - A$ edges than $A - B$ edges. The weighted version of the stochastic block model also allows encoding different values of utility from collaboration between different groups of people. In addition to our work on hiring processes, the model we developed can also capture natural phenomena such as the variation of populations of a pair of symbiotic species in a biome, organizational dynamics between experts and novices who may learn from each other, etc.

The hiring process can be equated to adding new nodes to the graph generated by the stochastic block model. In each step, one node is added, and the identity of the incoming node is based on how established the corresponding community is inside the graph. i.e. if community A has a greater number of edges with higher weights, then the new node joining the graph is likelier to belong to community A.

A new engineer who is hired into a tech company is likelier to belong to the community that is more influential and established inside the company. Further biases cause collaborations (and hence positive results) between engineers with similar communal identities to be likelier. We show in this thesis that in such a system, this behavior would reinforce the existing implicit biases about that community, result in a hiring environment where the new engineers who are hired are more likely to belong to the majority community. Over time, the majority community's representation will increase while that of the minority is diminished. Thus, it is essential that incentives are provided to ensure fairness.

We model these incentives as additional weight (the utility of the collaboration) for edges between two nodes of different communities - this represents additional funding or other institutional support for diverse collaborations. We define a deterministic variant of our model that provides several insights into the behaviour of our model. We analytically derive the system evolution of this determinstic system, and hence the exact circumstances under which this system will converge to a fair proportion of majority and minority and when it would lead to an unfair result in the long term. We then use the Hoeffding inequality to show that this deterministic system is a good representation of our stochastic block model in the long term. Finally, we present an outline to use the differential equation method to prove that with high probability, the deterministic variant is a good approximation for the stochastic system's temporal evolution.

1.1 Organization

Chapter 2 gives an overview of related work in fairness in general. Chapter 3 goes into the required background knowledge in mathematics and probability theory. Chapter 4 introduces and defines our model along with notation, nomenclature and preliminary equations that are frequently used in the rest of this thesis. Chapter 5 introduces the deterministic variant of our stochastic block model and derives numerous results on its temporal evolution. Chapter 6 goes into detail on the stochastic system and proves that it can be approximated by the deterministic system.

Chapter 7 concludes the thesis and offers a discussion of potential directions in which our work may be built upon and improved.

Chapter 8 is a summary of the notation used across this thesis. It is to be used as a reference whenever the reader is in doubt about the meaning and value of a specific variable, function or a constant.

2. RELATED WORK

2.1 Fairness in Algorithmic decision making

The rapid growth of machine learning and algorithmic decision making in the last few years has made analysis of fairness of these techniques crucial. Informally, fairness means to ensure that such techniques do not suffer bias against or towards particular population groups. Gajane and Pechenizkiy [2] present a survey of various formalizations of fairness and a theoretical as well as empirical critique over these notions. Chouldechova and Roth [3], Mehrabi et al. [4] provide a survey of theoretical work aimed at understanding algorithmic bias and fairness in machine learning.

There has been significant research on achieving fairness in the hiring process, mostly focusing on altering the nature of the hiring process to achieve fairness. Hu and Chen [5] propose a dual labor market with a temporary labor market (TLM) - where constraints are placed to ensure statistical parity of workers granted entry, and a permanent labor market (PLM) where the the top performers are hired. The constraints in the TLM achieve an equitable long-term equilibrium in the PLM.

Liu et al. [6] discuss interventions in the broader context of classification algorithms. It stages interventions in two forms: by decoupling the decision rule for classification by group and by subsidizing the cost of investment to acquire a qualification (such as the tuition cost of a university program) in order to motivate more individuals from a disadvantaged groups to invest in obtaining said qualifications. They also lay special focus on situations where investment cost is different for each group (such as higher rate of interest for a low income family). These interventions are aimed at improving the equilibrium qualification rate of disadvantaged groups.

These results in general alter the nature of the hiring process, in addition to targeting benefits to a specific community. In contrast to these results, our work focuses on altering the internal dynamics of the organization. By only incentivizing collaborations between two communities, we demonstrate that it is possible to indirectly influence the hiring process in a way that promotes fair representation in the organization that is preserved over the long-term.

Liu et al. [1] demonstrates that many static fairness criteria do not necessarily promote improvement over the long term. They show that for classification systems (e.g. for loan eligibility), static constraints introduced to promote fairness can lead to a worsening of the disadvantaged group's situation in the long term. In addition, they show measurement errors may mislead us into overestimating the effectiveness of these criteria and depict the importance of measurement in evaluation of fairness criteria. Taking these observations into account, we also examine the long term impact of providing incentives in this thesis.

2.2 The differential equation method

There are a number of random graph processes that are useful in a variety of situations. For instance, one such process is the degree bound graph process where we take an edgeless graph of n vertices and keep adding edges to it until we obtain a graph where the subgraph induced by the vertices of degree less than d is a clique. The differential equation method approximates the trajectories of a random process by the solutions to differential equations. Wormald [7] provides a general framework for applying the differential equation method to discrete-time randomized algorithms and random combinatorial structures. This powerful technique can be applied in a large variety of problems. As a specific example, in the degree bound graph process mentioned above, using the differential equation method and some other arguments, it was shown that asymptotically almost surely the final graph is regular if dn is even, and almost regular, with one vertex of degree $d-1$ and the rest of degree d, otherwise. Warnke [8] provides a simplified and slightly improved version (with better accuracy and less error probability) of the primary result in [7] which we can use for the model described in this thesis.

We now present a number of articles which use the differential equation method for their results. The H-free process is a random graph process defined by starting with an empty graph on n vertices and then adding edges one at a time, chosen uniformly at random subject to the constraint that no H subgraph is formed (where H is a fixed graph). Bohman and Keevash [9] present multiple results for the H-free process including the independence number, Ramsay number and the Turan number for specific classes of graphs by using the differential equation method. Amini and Minca [10] analyse the epidemic process among a population that interacts through a network. They show that in the random regular graphs, voluntary social distancing will always be sub-optimal and they use the differential equation method from [8] and [7] as a key part of their proof. More instances of the application of the differential equation method can be found in Branzei and Peres [11], Manurangsi and Suksompong [12], Yang and Yeung [13].

2.3 The stochastic block model

The stochastic block model is a generative model for random graphs. It is often used to study clustering and community detection. Loosely, community detection is about partitioning the vertices of a graph into clusters that are more densely connected. Abbe [14] provides a detailed survey of recent results in community detection using the stochastic block model. Heimlicher, Lelarge and Massoulié [15] present results for community detection in a labelled version of the stochastic block model. This labelled version is essentially similar to the regular stochastic block model except for an assignment of labels to the edges in the graph and can be useful in a number of specialized applications such as protein-protein interactions which may be exothermic (release energy) or endothermic (require energy).

The community detection (community recovery) in stochastic block model could be classified into multiple categories based on the guaranteed optimality properties: Exact recovery where the communities are recovered with high probability i.e., with probability tending to one as $n-$ the number of vertices tends to infinity and minimum misclustering error rate are some of them. Exact recovery is studied by Abbe, Bandeira and Hall [16], Jog and Loh [17], Abbe and Sandon [18]. Gao, Ma, Zhang and Zhou [19] present a polynomial time algorithm to achieve optimal results in misclustering proportion for stochastic block model under weak regularity conditions. Xu, Jog and Loh [20] study a weighted version of the community detection problem using the stochastic block model where the weight of each edge is generated independently from an unknown probability density determined by the community membership of its endpoints and the observations are collected in a weighted adjacency matrix.

Our usage of the stochastic block model is not for community detection, but rather to model connections and collaborations in an organization.

3. BACKGROUND

3.1 Weighted Stochastic block model

The stochastic block model is a model used to generate random graphs. This model is usually used to produce graphs containing communities, which are subsets characterized by being connected with one another with particular edge densities. For example, edges between two specific communities might be more common than edges between any other pair of communities.

In the stochastic block model there are n nodes labelled $\{1, \ldots, n\}$. Each node i has a color $c_{\mathrm{i}} \in \{c_1, \ldots, c_r\}$. A graph is generated as follows: each edge (i, j) is formed with probability $P_{c_{\mathrm{i}}, c_{\mathrm{j}}}$, where P is a symmetric matrix of size $r \times r$.

In the weighted stochastic block the model, the edges have weights. If edge (i, j) exists, then it gets weight $w_{c_{\mathrm{i}}, c_{\mathrm{j}}}$.

In a real world scenario, the color of a node could represent the community or type of that node. In much of the literature, the problem is: given a graph G that was generated according to the stochastic block model, recover the underlying model (i.e. the probability matrix P and weights w).

3.2 Hoeffding inequality

Theorem 3.2.1. *By the Hoeffding inequality([21]), if $X_1, X_2 ..., X_n$ are independent random variables such that $a_{\mathrm{i}} \leq X_{\mathrm{i}} \leq b_{\mathrm{i}}$ for all i and $X = \sum_{\mathrm{i}=1}^{\mathrm{i}=n} X_{\mathrm{i}}, \mu = \mathbb{E}[X]$, $c_{\mathrm{i}} = b_{\mathrm{i}} - a_{\mathrm{i}}$. Then, for all $\Delta > 0$*

$$\mathsf{Pr}[|X - \mu| > \Delta] \leq 2\mathrm{e}^{\frac{-2\Delta^2}{n \sum_{\mathrm{i}=1}^{\mathrm{i}=n} c_{\mathrm{i}}^2}} \tag{3.1}$$

3.3 $\sigma-$field

A $\sigma-$field is a collection of subsets of sample space which satisfy certain properties and are used to establish a formal definition of probability.

Let X be some set and $\mathcal{P}(X)$ be it's power set. Then, if $\Sigma \subseteq \mathcal{P}(X)$, then Σ is a $\sigma-$field if it satisfies the following properties:

- X is in Σ.
- Σ is closed under union operation.
- Σ is closed under the complement operation.

Note that the combination of the first and third properties means that the null set must be in Σ for it to be a $\sigma-$field.

3.4 The differential equation method

Given that we wish to track a derived random sequence associated with an underlying discrete-time stochastic process, Wormald's differential equation method shows that this random sequence can be PAC-approximated (Probably Approximate Correct) by the solution to a differential equation. The following is a refined version of Wormald's theorem by Warnke ([8]).

Theorem 3.4.1. *([8]) Given integers $a, n \geq 1$, a bounded domain $D \subseteq R^{a+1}$, functions $(F_k)_{1\leq k\leq a}$ with $F_k : D \to R$, and $\sigma-$fields $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \cdots$, suppose that the random variables $((Y_k(\mathrm{i}))1 \leq k \leq a$ are $\mathcal{F}_\mathrm{i}$-measurable for $\mathrm{i} \geq 0$. Furthermore, assume that, for all $\mathrm{i} \geq 0$ and $1 \leq k \leq a$, the following conditions hold whenever $(\mathrm{i}/n, Y_1(\mathrm{i})/n, ..., Y_a(\mathrm{i})/n) \in D$:*

- $|\mathbb{E}[Y_k(\mathrm{i}+1) - Y_k(\mathrm{i}) \mid \mathcal{F}_\mathrm{i}] - F_k(\mathrm{i}/n, Y_1(\mathrm{i}/n), \ldots, Y_a(\mathrm{i}/n))| \leq \delta$, *where the function F_k is $L-$Lipschitz continuous on D,*
- $|Y_k(\mathrm{i}+1) - Y_k(\mathrm{i})| \leq \beta$

  and the following condition holds initially:
- $max_{1\leq k\leq a}|Y_k(0) - \hat{y}_k n| \leq \lambda n$ *for some* $(0, \hat{y}_1, \ldots, \hat{y}_a) \in D$

Then there are $R = R(D, (F_k)_{1\le k\le a}, L) \in [1, \infty)$ *and* $T = T(D) \in (0, \infty)$ *such that, whenever* $\lambda \ge \delta min\{T, L^{-1}\} + R/n$*, with probability at least* $1 - 2a\mathrm{e}^{-n\lambda^2/(8T\beta^2)}$ *we have*

$$max_{0\le \mathrm{i}\le \sigma n}\ max_{1\le k\le a}\left|Y_k(\mathrm{i}) - y_k\Big(\frac{\mathrm{i}}{n}\Big)n\right| < 3\mathrm{e}^{LT}\lambda n \tag{3.2}$$

where $(y_k(t))_{1\le k\le a}$ *is the unique solution to the system of differential equations*

$$y_k{}'(t) = F_k(t, y_1(t), \ldots, y_a(t)) \text{ with } y_k(0) = \hat{y}_k \text{ for } 1 \le k \le a \tag{3.3}$$

and $\sigma = \sigma(\hat{y}_1, \ldots, \hat{y}_a(t)) \in (0, T]$ *is any choice of* $\sigma \ge 0$ *with the property that* $(t, y_1(t), \ldots, y_a(t))$ *has* ℓ^∞ *-distance at least* $3\mathrm{e}^{LT}\lambda$ *from the boundary of* D *for all* $t \in [0, \sigma)$*.*

4. THE MODEL

4.1 Weighted Stochastic Block Model

Let $[n] = \{1, \ldots, n\}$ be a set of nodes, such that each node i has color $c_{\mathrm{i}} \in \{red, blue\}$. A random undirected graph is generated by creating each edge (i, j) with probability $P_{c_{\mathrm{i}},c_{\mathrm{j}}}$. The probability matrix P is symmetric, with $P_{\mathrm{i,j}} = P_{\mathrm{j,i}} = a$ if nodes i and j have the same color, and $P_{\mathrm{i,j}} = P_{\mathrm{j,i}} = b$ otherwise.

If edge (i, j) is formed, its weight is $w_{c_{\mathrm{i}},c_{\mathrm{j}}}$, the weight matrix $\mathbf{w}$ is symmetric, with $w_{c_{\mathrm{i}},c_{\mathrm{j}}} = \alpha > 0$ if i and j have the same color and $w_{c_{\mathrm{i}},c_{\mathrm{j}}} = \beta > 0$ if they have different colors.

4.2 Dynamics

We study a dynamic process where at each step, a random graph is generated according to the weighted stochastic block model. Then a new node arrives, taking color red (blue) with probability equal to the edge weight of red (blue).

More formally, we slightly override notation and denote by $w_{\mathrm{i,j}}(t)$ the weight of edge (i, j) at time t. We have:

$$w_{\mathrm{i,j}}(t) = \begin{cases} w_{c_{\mathrm{i}},c_{\mathrm{j}}} & \text{with probability } P_{c_{\mathrm{i}},c_{\mathrm{j}}} \\ 0 & \text{otherwise} \end{cases} \tag{4.1}$$

The sum of weights at time t is $w(t) = \sum_{\mathrm{i,j}=1}^{n} w_{\mathrm{i,j}}(t)$. We refer to $w(t)$ as the total weight at time t. The weight of each color at time t is

$$w_{red}(t) = \sum_{\mathrm{i}:c_{\mathrm{i}}=red} \sum_{\mathrm{j}} w_{\mathrm{i,j}}(t) \quad \text{and} \quad w_{blue}(t) = \sum_{\mathrm{i}:c_{\mathrm{i}}=blue} \sum_{\mathrm{j}} w_{\mathrm{i,j}}(t)\,.$$

The dynamical system we study is formally defined as follows.

Dynamical system definition: At each time step $t = 0, 1, 2, \ldots$,

- a random graph G_t is generated according to probability matrix P and weights $\mathbf{w}$.
- a new node arrives. If $w(t) > 0$, the node takes color red with probability $p_t = \frac{w_r(t)}{w(t)}$ and blue with probability $1 - p_t$. If $w(t) = 0$, then the new node takes each of the colors with probability $p_t = 1/2$.

Notation: We use the following notation through the rest of this thesis. In addition, we also define a few more variables whenever they are required in each section.

Let $\rho = \frac{\nu}{\lambda}$ where $\nu = \frac{a}{b}$ and $\lambda = \frac{\beta}{\alpha}$.

Without loss of generality, we assume that when the dynamic begins, the color red is the minority if one exists. We also assume that we start with more than zero nodes i.e., $n > 0$ for the problem fails to be interesting otherwise. Let $n_r(t)$ and $n_b(t)$ denote the number of red and blue nodes at time t, respectively. Since one node arrives at each point in time, we have $n + t$ nodes in total at time t, with

$$n_r(t) = X(t) \cdot (n + t) \tag{4.2}$$

$$n_b(t) = (1 - X(t)) \cdot (n + t) \tag{4.3}$$

4.3 Questions

For the dynamic we defined earlier, we investigate the following questions :

- A deterministic system that can be useful for understanding the random one is obtained by considering the update rule obtained as $n \to \infty$. At what threshold $\lambda^*(P)$ is there a phase transition? It seems that at $\rho = 0$ the minority will vanish in the long run, while at $\lambda = \infty$ it will achieve fair representation (50%). What is the threshold for this change?
- In the stochastic system, how long does it take for the minority to achieve a certain minimum representation?

- Understand the stochastic system and show what happens with high probability.

4.4 An update rule for the stochastic system

Let $Y(t)$ be an indicator random variable, such that $Y(t) = 1$ if the color of the incoming node at time t is red, and $Y(t) = 0$ otherwise. Thus we have

$$Y(t) = \begin{cases} \begin{cases} 1 \text{ with probability } \dfrac{w_r(t)}{w(t)} \\ 0 \text{ with probability } 1 - \dfrac{w_r(t)}{w(t)} \end{cases} & \text{if } w(t) \neq 0 \\ \begin{cases} 1 \text{ with probability } \dfrac{1}{2} \\ 0 \text{ with probability } \dfrac{1}{2} \end{cases} & \text{if } w(t) = 0 \end{cases} \tag{4.4}$$

Let $\{\mathcal{F}_t\}_{t\geq 0}$ be a set of increasing $\sigma-$fields which denote the history of the stochastic process where $\mathcal{F}_t$ denotes the history after t rounds i.e., after t new nodes are added to the system. It follows that, if $w(t) \neq 0$,

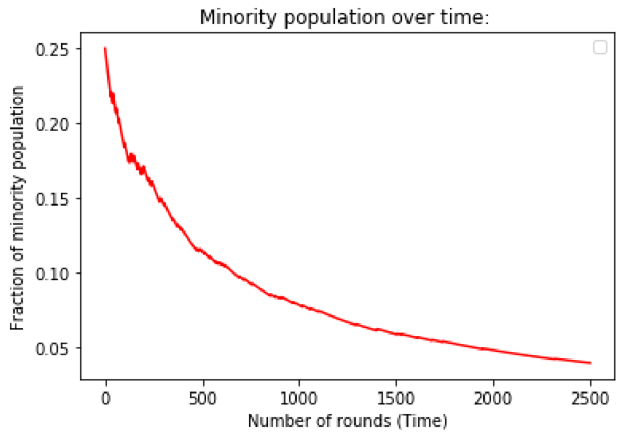
$$\mathbb{E}[Y(t) \mid \mathcal{F}_t] = \mathbb{E}\left[\frac{w_r(t)}{w(t)} \,\middle|\, \mathcal{F}_t\right] \tag{4.5}$$

Recall the fraction of red nodes at time t is

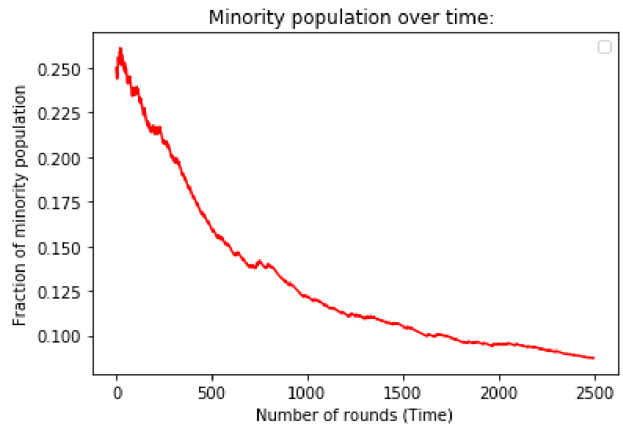
$$X(t) = \frac{n_r(t)}{n + t} \tag{4.6}$$

From the definition of our dynamic, we have the following update rule:

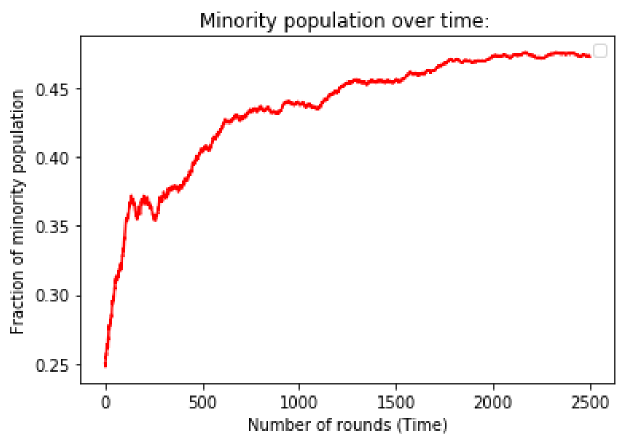
$$X(t+1) = \frac{X(t) \cdot (n + t) + Y(t)}{n + t + 1} \tag{4.7}$$



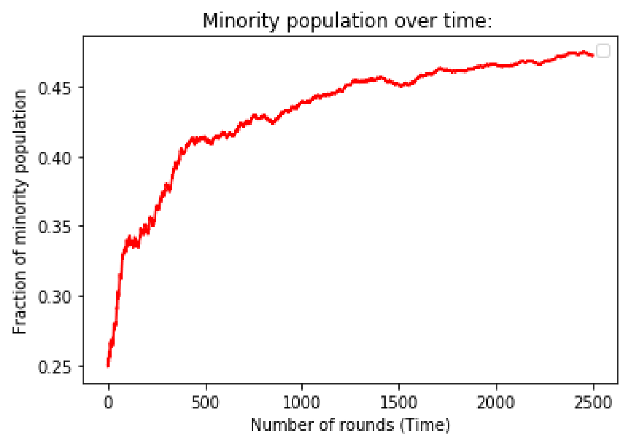
(a)$\rho = 10$



(b)$\rho = 2.5$

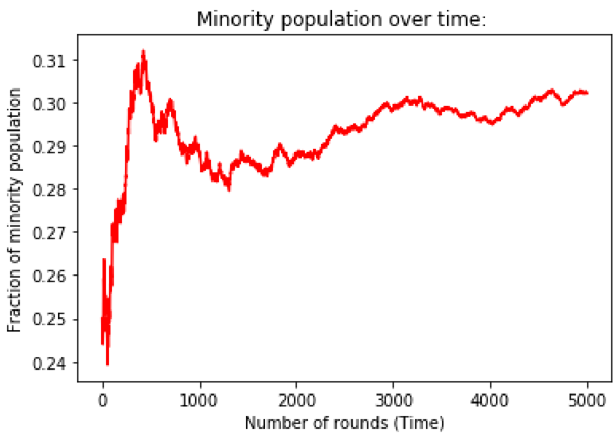


(c)$\rho = 0.1$

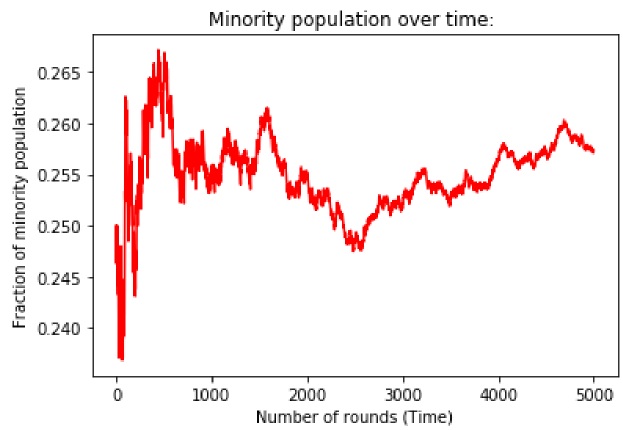


(d)$\rho = 0.01$

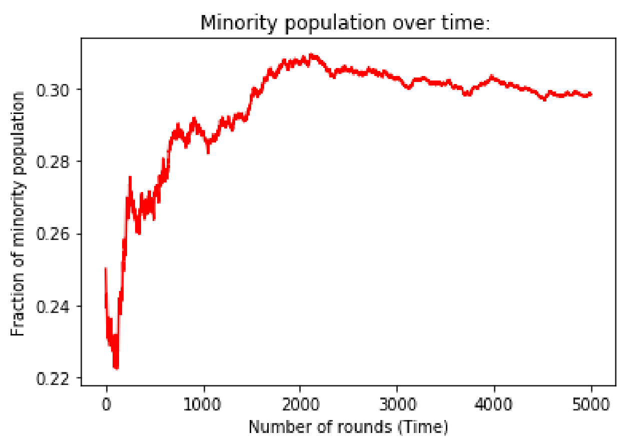
Figure 4.1. Trial runs of the stochastic system for a range of values of ρ. We start with 200 nodes in all cases.



(a)$\rho = 1$



(b)$\rho = 1.0000025$



(c)$\rho = 0.9999$

Figure 4.2. Trial runs of the stochastic system when ρ is close to 1. There is some fluctuation initially as each coin flip greatly affects the fraction. But over time the system stabilizes around a specific point.

4.5 Motivation

We believe that this dynamic will be a good representation of the hiring process in organizations where a newly inducted employee is more likely to belong to the more influential community. Our model aspires to study and also promote fairness in organisations. In addition, we also believe that this model or it's variants can be applied to several interesting problems such as the variation of populations of a pair of symbiotic species in a biome, organizational dynamics between experts and novices who may learn from each other and so forth. Finally, we hope that our work in this model could inspire research in similar dynamics that have a lot of applications in the real world.

5. A DETERMINISTIC SYSTEM

In order to understand the stochastic system, we will first pick and study a deterministic system. Later, we will see that this deterministic system is with high probability a good approximation to the stochastic system.

5.1 The update rule

We consider the deterministic dynamical system given by the following update rule:

$$X_d(t+1) = \frac{X_d(t) \cdot (n+t) + Y_d(t)}{n+t+1} \tag{5.1}$$

where

$$Y_d(t) = \frac{\rho \cdot (X_d(t))^2 + X_d(t) \cdot (1 - X_d(t))}{\rho \cdot \left((X_d(t))^2 + (1 - X_d(t))^2\right) + 2X_d(t) \cdot (1 - X_d(t))}$$

Any variable which has the subscript d denotes that the variable is for the deterministic system.

Let $w_{rr}(t)$ denote the sum of weights of all edges between two red nodes at time t. Then

$$\mathbb{E}[w_{rr}(t) \mid \mathcal{F}_t] = a\alpha(n_r(t))^2 = a\alpha(X(t))^2(n+t)^2 \tag{5.2}$$

The above expression comes because there are $n_r(t) \cdot n_r(t)$ edges that can form between two red nodes. And each of those edges is formed with a probability of a and has a weight of α. Finally, we use (4.2) to rewrite $n_r(t)$ in terms of $X(t)$.

Similarly, if $w_{rb}(t)$ denotes the sum of weights of all edges between a red node and a blue node at time t, then

$$\mathbb{E}[w_{rb}(t) \mid \mathcal{F}_t] = b\beta(n_r(t))(n+t-n_r(t)) = b\beta(X(t))(1-X(t))(n+t)^2 \tag{5.3}$$

Combining (5.2) and (5.3) we get

$$\mathbb{E}[w_r(t) \mid \mathcal{F}_t] = \mathbb{E}[w_{rr}(t) + w_{rb}(t) \mid \mathcal{F}_t] = a\alpha(n+t)^2(X(t))^2 + b\beta(n+t)^2(X(t))(1-X(t)) \tag{5.4}$$

where $w_r(t)$ denotes the sum of weights of all edges which are incident on at least one red node at time t.

Similarly, we get the following sets of expressions for the color blue:

$$\mathbb{E}[w_{bb}(t) \mid \mathcal{F}_t] = a\alpha(n_b(t))^2 = a\alpha(1-X(t))^2(n+t)^2 \tag{5.5}$$

Then, $E[w_b(t)|X(t)]$ will be given by

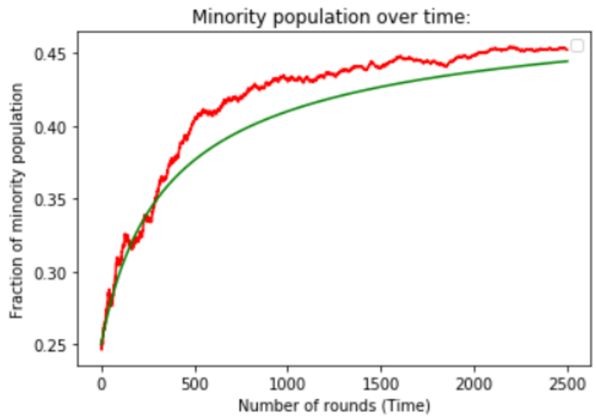
$$\begin{aligned} \mathbb{E}[w_b(t) \mid \mathcal{F}_t] &= \mathbb{E}[w_{bb}(t) + w_{rb}(t) \mid \mathcal{F}_t] \\ &= a\alpha(n+t)^2(1-X(t))^2 + b\beta(n+t)^2(X(t))(1-X(t)) \end{aligned} \tag{5.6}$$

Finally, combining (5.6) and (5.4), we can see that

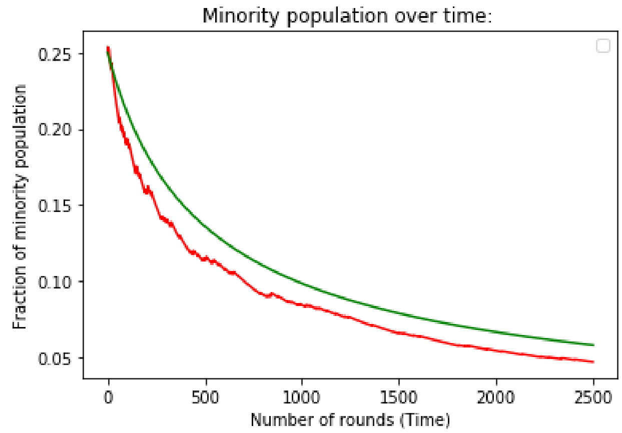
$$\begin{aligned} \mathbb{E}[w(t) \mid \mathcal{F}_t] &= \mathbb{E}[w_r(t) + w_b(t) \mid \mathcal{F}_t] \\ &= a\alpha(n+t)^2 \left((X(t))^2 + (1-X(t))^2\right) + 2b\beta(n+t)^2(X(t))(1-X(t)) \end{aligned} \tag{5.7}$$

We obtain $Y_d(t)$ by taking (4.5) and approximating it as

$$Y_d(t) = \frac{\mathbb{E}[w_r(t) \mid \mathcal{F}_t]}{\mathbb{E}[w(t) \mid \mathcal{F}_t]} \tag{5.8}$$



(a)



(b)

Figure 5.1. Minority Proportion vs Time. In both cases we start with 200 nodes of which 50 are red. In case (b) the value of ρ is $4(>1)$ and the minority diminishes over time, while in (a) we have $\rho = 0.25$.

From (5.4), (5.7) and (5.8), it follows that

$$Y_d(t) = \frac{a\alpha(n+t)^2(X_d(t))^2 + b\beta(n+t)^2(X_d(t))(1-X_d(t))}{a\alpha(n+t)^2\left((X_d(t))^2+(1-X_d(t))^2\right)+2b\beta(n+t)^2(X_d(t))(1-X_d(t))} \tag{5.9}$$

$$= \frac{\rho(X_d(t))^2 + (X_d(t))(1-X_d(t))}{\rho\left((X_d(t))^2+(1-X_d(t))^2\right)+2(X_d(t))(1-X_d(t))} \tag{5.10}$$

Finally, combining (5.1) and (5.8), we obtain,

$$X_d(t+1) = \frac{1}{n+t+1}\Big(X_d(t)\cdot(n+t) + \frac{\rho(X_d(t))^2 + (X_d(t))(1-X_d(t))}{\rho\left((X_d(t))^2+(1-X_d(t))^2\right)+2(X_d(t))(1-X_d(t))}\Big) \tag{5.11}$$

Note that this $Y_d(t)$ can be a fraction. To get an intuitive feel, consider that every node can be a mix of red and blue. So, when a new node is added, it could be partially red and partially blue. In addition, $X_d(t)$ is obtained by adding the fraction of red in every node and normalizing it.

5.2 The fixed points of the deterministic system

Lemma 1. *If, $\rho \neq 1$ then the fixed points of the deterministic system given by (5.11) are at $X_d(t) \in \{0, \frac{1}{2}, 1\}$. If, $\rho = 1$, then every point $X_d(t) \in [0, 1]$ is a fixed point of the deterministic system given by (5.11)*

Proof. In order to find the fixed points of (5.11), we need to find $X_d(t)$ satisfying

$$X_d(t+1) = X_d(t)$$

Using, (5.11), we instead need

$$X_d(t) = \frac{1}{n+t+1} \cdot \Big(X_d(t) \cdot (n+t) + \frac{\rho (X_d(t))^2 + (X_d(t))(1 - X_d(t))}{\rho \left((X_d(t))^2 + (1 - X_d(t))^2 \right) + 2(X_d(t))(1 - X_d(t))} \Big)$$

Simplify the above equation and we get

$$X_d(t) = \frac{\rho (X_d(t))^2 + (X_d(t))(1 - X_d(t))}{\rho \left((X_d(t))^2 + (1 - X_d(t))^2 \right) + 2(X_d(t))(1 - X_d(t))} \tag{5.12}$$

Cross-multiply and simplify (5.12) and we instead need a $X_d(t)$ that satisfies

$$(X_d(t))^3 (2\rho - 2) + (X_d(t))^2 (-2\rho + 2) + \rho X_d(t) = \rho (X_d(t))^2 + X_d(t) (1 - X_d(t)) = 0 \tag{5.13}$$

Simplify and factorise (5.13) and we obtain,

$$(1 - \rho) X_d(t) (X_d(t) - 1) (2X_d(t) - 1) = 0$$

Thus, as long as $\rho \neq 1$, the fixed points are at $X_d(t) = 1$,$X_d(t) = \frac{1}{2}$ and $X_d(t) = 0$.

However, if $\rho = 1$, then every point $X_d(t) \in [0, 1]$ is a fixed point of the deterministic system. □

5.3 Convergence of the deterministic system

Lemma 2. *Consider the function $G : [0,1] \to \mathbb{R}$ given by*

$$G(x) = (1-\rho)(x-1) \cdot x \cdot (2x-1) \tag{5.14}$$

Then $G(X_d(t))$ has the same sign as $X_d(t+1) - X_d(t)$.

Proof. From (5.11) we have

$$X_d(t+1) - X_d(t) = \frac{1}{n+t+1} \cdot \left(-X_d(t) + \frac{\rho(X_d(t))^2 + (X_d(t))(1-X_d(t))}{\rho\left((X_d(t))^2 + (1-X_d(t))^2\right) + 2(X_d(t))(1-X_d(t))} \right)$$

Then we get

$$X_d(t+1) - X_d(t) = \frac{-2(X_d(t))^3\rho + 2(X_d(t))^3 + 3(X_d(t))^2\rho - 3(X_d(t))^2 - (X_d(t))\rho + (X_d(t))}{\Big(\rho\left(2(X_d(t))^2 - 2(X_d(t)) + 1\right) + 2(X_d(t))\left(1 - (X_d(t))\right)\Big)(n+t+1)} \tag{5.15}$$

The denominator in the above expression is always positive and hence it does not affect the sign. Hence, the sign of $X_d(t+1) - X_d(t)$ is the same as the sign of the numerator in (5.15), that is:

$$-2(X_d(t))^3\rho + 2(X_d(t))^3 + 3(X_d(t))^2\rho - 3(X_d(t))^2 - (X_d(t))\rho + (X_d(t)) \tag{5.16}$$

Factorising (5.16), we obtain $(1-\rho)(X_d(t)-1)X_d(t)(2X_d(t)-1) = G(X_d(t))$ as required. $\square$

Lemma 3. *Consider the deterministic system $(X_d)_{n\geq 0}$. Then $X_d(t+1) > X_d(t)$ when one of the following conditions holds*

1. *$\rho > 1$ and $\frac{1}{2} < X_d(t) < 1$*
2. *$\rho < 1$ and $0 < X_d(t) < \frac{1}{2}$*

Proof. We have $X_d(t+1) > X_d(t)$ when $X_d(t+1) - X_d(t) > 0$. From Lemma 2, the inequality is satisfied when $G(X_d(t)) = (1-\rho)(X_d(t)-1)X_d(t)(2X_d(t)-1) > 0$. We have $G(X_d(t)) > 0$ when one of the conditions 1 and 2 holds. $\square$

Lemma 4. *Consider the deterministic system $(X_d)_{n\geq 0}$. Then $X_d(t+1) < X_d(t)$ when one of the following conditions holds*

1. $\rho < 1$ *and* $\frac{1}{2} < X_d(t) < 1$

2. $\rho > 1$ *and* $0 < X_d(t) < \frac{1}{2}$

Proof. We have $X_d(t+1) < X_d(t)$ when $X_d(t+1) - X_d(t) < 0$. From Lemma 2, the inequality is satisfied when $G(X_d(t)) = (1-\rho)(X_d(t)-1)X_d(t)(2X_d(t)-1) < 0$. We have $G(X_d(t)) < 0$ when one of the conditions 1 and 2 holds. □

Lemma 5. *Consider the deterministic system $(X_d)_{n\geq 0}$. Recall that $X_d(0) \leq \frac{1}{2}$. then the minority never becomes an absolute majority i.e., $X_d(t) \leq \frac{1}{2}$ $\forall t \geq 0$*

Proof. In any round, we have from (5.9),

$$Y_d(t) = \frac{\rho(X_d(t))^2 + (X_d(t))(1 - X_d(t))}{\rho\left((X_d(t))^2 + (1 - X_d(t))^2\right) + 2(X_d(t))(1 - X_d(t))} \tag{5.17}$$

$$= \frac{\rho(X_d(t))^2 + (X_d(t))(1 - X_d(t))}{\rho(X_d(t))^2 + (X_d(t))(1 - X_d(t)) + \rho(1 - X_d(t))^2 + (X_d(t))(1 - X_d(t))} \tag{5.18}$$

If $X_d(t) \leq \frac{1}{2}$, then

$$\rho(X_d(t))^2 \leq \rho(1 - X_d(t))^2$$

The above expression is true because $1 - X_d(t) \geq X_d(t)$ and in addition, $X_d(t)$ as well as $1 - X_d(t)$ are positive. Thus, it follows that

$$\rho(X_d(t))^2 + (X_d(t))(1 - X_d(t)) \leq \rho(1 - X_d(t))^2 + (X_d(t))(1 - X_d(t)) \tag{5.19}$$

From (5.19), we can see that

$$2\left(\rho(X_d(t))^2 + (X_d(t))(1 - X_d(t))\right) \leq \rho\left((X_d(t))^2 + (1 - X_d(t))^2\right) + 2(X_d(t))(1 - X_d(t))$$

and thus when we also use (5.17), we have

$$Y_d(t) \leq \frac{1}{2}$$

The above expression means that in any step, if we add a new node, the new node always has a fraction of the majority color that is at least as big as the fraction of the minority color. Extending this over multiple steps, we can see that the fraction of the majority color added over multiple steps as at least as big as the fraction of the minority color added.

Thus, the number of majority nodes is always at least as big as the number of minority nodes and the lemma is proved. □

Theorem 5.3.1. *Assume that the deterministic system does not start at* $X_d(0) \in \{0, \frac{1}{2}, 1\}$*. Then,*

- *If* $\rho > 1$*, the deterministic system converges to 0,1 (i.e the minority moves to 0 and the majority to 1)*
- *if* $\rho < 1$*, the deterministic system converges to* $\frac{1}{2}$
- *if* $\rho = 1$*, the deterministic system stays wherever it is.*

Proof. Recall that $0 \leq X_d(0) \leq \frac{1}{2}$ from our model definition (as red is the minority at the start of the dynamic if one exists). In addition, using our assumptions, we have $0 < X_d(0) < \frac{1}{2}$.

- If $\rho > 1$, by Lemma 4, $X_d(t+1) < X_d(t)$. We have a decreasing sequence which is bounded below (at $X_d(t) = 0$). Thus, it has to converge. Apply Lemma 1 and it follows that $X_d(t)$ has to converge at $X_d(t) = 0$.
- If $\rho < 1$, by Lemma 3, $X_d(t+1) > X_d(t)$. We have a increasing sequence which is bounded above (at $X_d(t) = \frac{1}{2}$ - using Lemma 5). Thus, it has to converge. Apply Lemma 1 and it follows that $X_d(t)$ has to converge at $X_d(t) = \frac{1}{2}$.

□

Theorem 5.3.2. *The deterministic system* $(X_d)_{n\geq 0}$ *is monotonic.*

Proof. We can split this into three cases, depending on whether $\rho > 1$, $\rho = 1$, or $\rho < 1$. We consider them case by case.

- When $\rho > 1$, we can split this further into three sub cases. When $0 < X_d(t) < \frac{1}{2}$, we can apply Lemma 4 to see that $X_d(t+1) < X_d(t)$. Further, when $\frac{1}{2} < X_d(t) < 1$, we can use Lemma 3 to see that $X_d(t+1) > X_d(t)$. Finally, when $X_d(t) \in \{0, \frac{1}{2}, 1\}$, $X_d(t+1) = X_d(t)$ using Lemma 1. Thus, in this case, $X_d(t)$ is monotonic as long as it does not cross a fixed point. By Lemma 5, the system never crosses the fixed point $X_d(t) = \frac{1}{2}$ and in addition it cannot cross $X_d(t) = 0$ or $X_d(t) = 1$ because it's domain is restricted to $[0, 1]$ (since it is a fraction.)

- When $\rho = 1$, by Lemma 1, $X_d(t+1) = X_d(t)$ and thus the system is monotonic.

- When $\rho < 1$, we can split this further into three sub cases. When $0 < X_d(t) < \frac{1}{2}$, we can apply Lemma 3 to see that $X_d(t+1) > X_d(t)$. Further, when $\frac{1}{2} < X_d(t) < 1$, we can use Lemma 4 to see that $X_d(t+1) < X_d(t)$. Finally, when $X_d(t) \in \{0, \frac{1}{2}, 1\}$, $X_d(t+1) = X_d(t)$ using 1. Thus, in this case, $X_d(t)$ is monotonic as long as it does not cross a fixed point. By Lemma 5, the system never crosses the fixed point $X_d(t) = \frac{1}{2}$ and in addition it cannot cross $X_d(t) = 0$ or $X_d(t) = 1$ because it's domain is restricted to $[0, 1]$ (since it is a proper fraction.)

Thus, the system is monotonic in all three cases and is monotonic overall. $\square$

5.4 A continuous approximation of the deterministic system

We now look at a continuous time approximation of (5.11) specified by the following set of equations.

$$\begin{cases} \xi(t+h) = \dfrac{1}{n+t+h} \cdot \left(\xi(t) \cdot (n+t) + \dfrac{\rho(\xi(t))^2 + (\xi(t))(1-\xi(t))}{\rho\left((\xi(t))^2 + (1-\xi(t))^2\right) + 2(\xi(t))(1-\xi(t))} \cdot h\right) \\ \xi(0) = X_d(0) \end{cases} \tag{5.20}$$

Theorem 5.4.1. *The solution to the continuous system specified at (5.20) is given by the solution of the differential equation:*

$$\begin{cases} \dfrac{d\xi(t)}{dt} = -\dfrac{2\left(\rho-1\right)\left(\xi(t)-1\right)\xi(t)+\rho}{\left(\rho-1\right)\left(\xi(t)-1\right)\xi(t)\left(2\xi(t)-1\right)} \cdot \dfrac{1}{n+t} \\ \xi(0) = X_d(0) \end{cases} \tag{5.21}$$

The solution to the differential equation described in (5.21) satisfies the following expression.

$$\left.\frac{(\rho+1)\ln\Big(|2\xi(t)-1|\Big) - \rho\Big(\ln\Big(|\xi(t)|\Big)+\ln\Big(|\xi(t)-1|\Big)\Big)}{\rho-1}\right|_{\xi_0}^{\xi_T} = ln|t+n|\Big|_{t_0}^{t_T} \tag{5.22}$$

Proof. First, notice that

$$\frac{d\xi(t)}{h} = \frac{\xi(t+h)-\xi(t)}{h} \tag{5.23}$$

Then, using (5.23) and (5.20), we can see that

$$\frac{d\xi(t)}{h} = \frac{1}{n+t+h} \cdot \Big(-\xi(t)\cdot(h) + \frac{\rho(\xi(t))^2+(\xi(t))(1-\xi(t))}{\rho\left((\xi(t))^2+(1-\xi(t))^2\right)+2(\xi(t))(1-\xi(t))} \cdot h\Big) \cdot \frac{1}{h}$$

Simplifying the above expression, we can see that

$$\frac{d\xi(t)}{h} = \frac{1}{n+t+h} \cdot \frac{-\left(\rho-1\right)\left(\xi(t)-1\right)\xi(t)\left(2\xi(t)-1\right)}{2\left(\rho-1\right)\left(\xi(t)-1\right)\xi(t)+\rho}$$

Take the limit of the above equation as h approaches 0 and we have

$$\lim_{h\to 0}\frac{d\xi(t)}{h} = \frac{d\xi(t)}{dt} = \lim_{h\to 0}\Big(\frac{1}{n+t+h} \cdot \frac{-\left(\rho-1\right)\left(\xi(t)-1\right)\xi(t)\left(2\xi(t)-1\right)}{2\left(\rho-1\right)\left(\xi(t)-1\right)\xi(t)+\rho}\Big)$$

Applying the limit, we get

$$\frac{d\xi(t)}{dt} = \frac{1}{n+t} \cdot \frac{-(\rho-1)(\xi(t)-1)\,\xi(t)\,(2\xi(t)-1)}{2(\rho-1)(\xi(t)-1)\,\xi(t)+\rho}$$

which is the first result of the theorem.

Next, we will obtain the solution to this differential equation. First, cross multiply and we get a new equation

$$-\frac{2(\rho-1)(\xi(t)-1)\,\xi(t)+\rho}{(\rho-1)(\xi(t)-1)\,\xi(t)\,(2\xi(t)-1)} \cdot d\xi(t) = \frac{dt}{n+t} \tag{5.24}$$

Using partial fractions, the left-hand-side of (5.24) can be rewritten as

$$\frac{1}{\rho-1} \cdot \left((-2\rho-2) \cdot \frac{1}{2\xi(t)-1} d\xi(t) + \rho \cdot \frac{1}{\xi(t)} d\xi(t) + \rho \cdot \frac{1}{\xi(t)-1} d\xi(t)\right)$$

Integrate each of these terms and we obtain the following

$$\frac{(2\rho+2)\cdot ln|2\xi(t)-1|}{2(\rho-1)} - \frac{\rho ln|\xi(t)|}{(\rho-1)} - \frac{\rho ln|\xi(t)-1|}{(\rho-1)} + C_1$$

where C_1 is some constant.

Simplify it and then apply limits and we obtain the left-hand-side in (5.22).

Integrating the right hand side of (5.21), we get

$$ln|t+n|\Big|_{t_0}^{t_T} + C_2$$

where C_2 is some constant. Applying limits and we obtain the right hand side of (5.22).

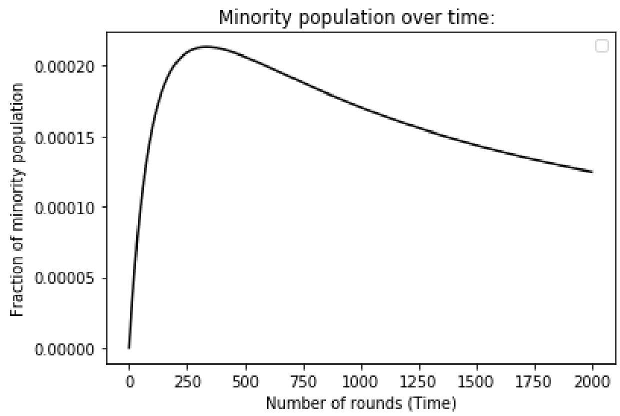
Thus, we obtain the solution of the differential equation described at (5.21). □

The solution of this differential equation is a good approximation to our deterministic system. We do not present a formal result on the accuracy of the system described by this continuous equation. However, our experiments suggest that this approximation is very close to the original deterministic system.

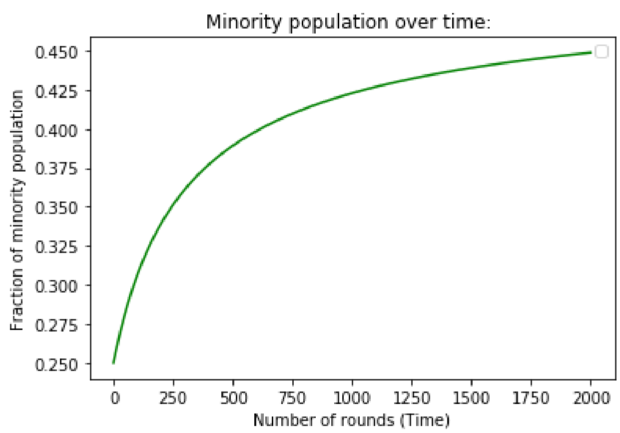
The expression given in Theorem 5.4.1 gives us an approximation of the deterministic system with very good accuracy. Using the differential equation method [8], one could also see that it gives a good approximation of the stochastic system with high probability.

Although Theorem 5.4.1 does not let us find $\xi(t)$ directly, one could use a root-finding algorithm like the Newton-Raphson method to get the value of $\xi(t)$ for any time interval from t_0 to t_T. However, proper care needs to be taken to ensure that in the case of multiple roots, we take the correct root i.e., the root which satisfies our results in the earlier sections. For instance, in the event that $\xi(0) < \frac{1}{2}$, if we have two roots for $\xi(t)$ with one root more than $\frac{1}{2}$ and another less than $\frac{1}{2}$, then the correct root which fits the original deterministic system would be then one less than $\frac{1}{2}$.

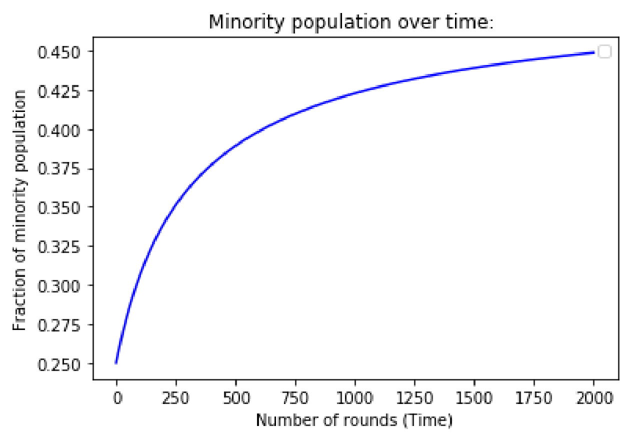
When we do that and plot the difference between the discrete deterministic system and the continuous approximation, we notice that the continuous system is an excellent approximation of the deterministic system. Their difference appears to be very less across multiple experiments.



(a)



(b)



(c)

Figure 5.2. A plot of the continuous approximation against the deterministic system. In (a), we show the difference between the value as per our continuous approximation and the deterministic system. (b) shows the values taken by the actual deterministic system, while (c) shows the values that we predict using the approximation at integer time points. We start with 200 nodes of which 50 are red. We also have $\rho = 0.19$.

6. THE STOCHASTIC SYSTEM

We will now look at the stochastic system that we talked about earlier in the model section. We will see some bounds for the update rule which give us an idea of how our system behaves in the long run.

6.1 Upper and lower bounds on the update rule

We use the following notation for this section in addition to more variables to be defined later:

$$\begin{cases} M_{ab} = max(a, b) \\ m_{a,b} = min(a, b) \\ M_{\alpha\beta} = max(\alpha, \beta) \\ m_{\alpha,\beta} = min(\alpha, \beta) \end{cases} \tag{6.1}$$

Lemma 6. *For the stochastic system, we have the following set of inequalities,*

$$\begin{cases} \mathbb{E}[w(t) \mid \mathcal{F}_t] \leq M_{ab} \cdot M_{\alpha\beta} \cdot (n+t)^2 \\ \mathbb{E}[w(t) \mid \mathcal{F}_t] \geq m_{a,b} \cdot m_{\alpha,\beta} \cdot (n+t)^2 \end{cases}$$

Proof. First, from (5.7), we have

$$\mathbb{E}[w(t) \mid \mathcal{F}_t] = a\alpha\Big((n_r(t))^2 + (n_b(t))^2\Big) + 2b\beta \cdot n_r(t) \cdot n_b(t)$$

Using the definition of M_{ab} and $M_{\alpha\beta}$, we get

$$\mathbb{E}[w(t) \mid \mathcal{F}_t] \leq M_{ab} \cdot M_{\alpha\beta} \cdot \Big((n_r(t))^2 + (n_b(t))^2\Big) + 2M_{ab}M_{\alpha\beta} \cdot n_r(t) \cdot n_b(t)$$

This can be rewritten as

$$\mathbb{E}[w(t) \mid \mathcal{F}_t] \leq M_{ab} \cdot M_{\alpha\beta} \cdot \Big((n_r(t))^2 + n_r(t) \cdot n_b(t) + (n_b(t))^2 + n_r(t) \cdot n_b(t)\Big)$$

which then becomes

$$\mathbb{E}[w(t) \mid \mathcal{F}_t] \leq M_{ab} \cdot M_{\alpha\beta}\Big(n_r(t)(n_r(t) + n_b(t)) + n_b(t)(n_r(t) + n_b(t))\Big)$$

followed by

$$\mathbb{E}[w(t) \mid \mathcal{F}_t] \leq M_{ab} \cdot M_{\alpha\beta}\Big(n_r(t)(n+t) + n_b(t)(n+t)\Big)$$

which finally becomes

$$\mathbb{E}[w(t) \mid \mathcal{F}_t] \leq M_{ab} \cdot M_{\alpha\beta} \cdot (n+t)^2$$

which is the required statement.

We will now see the proof of the other inequality. First, from (5.7), we have

$$\mathbb{E}[w(t) \mid \mathcal{F}_t] = a\alpha\Big((n_r(t))^2 + (n_b(t))^2\Big) + 2b\beta \cdot n_r(t) \cdot n_b(t)$$

Using the definition of m_{ab} and $m_{\alpha\beta}$, we get

$$\mathbb{E}[w(t) \mid \mathcal{F}_t] \geq m_{ab} \cdot m_{\alpha\beta} \cdot \Big((n_r(t))^2 + (n_b(t))^2\Big) + 2m_{ab}m_{\alpha\beta} \cdot n_r(t) \cdot n_b(t)$$

This can be rewritten as

$$\mathbb{E}[w(t) \mid \mathcal{F}_t] \geq m_{ab} \cdot m_{\alpha\beta}\Big(n_r(t)(n_r(t) + n_b(t)) + n_b(t)(n_r(t) + n_b(t))\Big)$$

followed by

$$\mathbb{E}[w(t) \mid \mathcal{F}_t] \geq m_{ab} \cdot m_{\alpha\beta}\Big(n_r(t)(n+t) + n_b(t)(n+t)\Big)$$

which finally becomes

$$\mathbb{E}[w(t) \mid \mathcal{F}_t] \geq m_{ab} \cdot m_{\alpha\beta} \cdot (n+t)^2$$

which is the required statement.

$\square$

Lemma 7. *For any $\delta > 0$ we have*

$$\begin{cases} \mathsf{Pr}\Big[|w_r(t) - \mathbb{E}[w_r(t) \mid \mathcal{F}_t]| > \delta\Big] & < 2\mathrm{e}^{\frac{-2\delta^2}{n_r(t)\cdot n(t)\cdot M_{\alpha\beta}^2}} \\ \mathsf{Pr}\Big[|w(t) - \mathbb{E}[w(t) \mid \mathcal{F}_t]| > \delta\Big] & < 2\mathrm{e}^{\frac{-2\delta^2}{(n+t)^2\cdot M_{\alpha\beta}^2}} \end{cases} \tag{6.2}$$

Proof. Recall that $w_{\mathrm{i,j}}(t)$ is a random variable that denotes the weight of the edge between node i and node j at the time t (4.1). First, see that

$$w_r(t) = \sum_{\mathrm{i}\in R(t),\mathrm{j}\in R(t)} w_{\mathrm{i,j}}(t) + \sum_{\mathrm{i}\in R(t),\mathrm{j}\in B(t)} w_{\mathrm{i,j}}(t) \tag{6.3}$$

where $R(t), B(t)$ is the set of red nodes, blue nodes at time t respectively.

We can interpret $w_r(t)$ as the sum of $n_r(t)\cdot n_r(t) + n_r(t)\cdot n_b(t) = n_r(t)\cdot(n+t)$ variables. Of these variables, $n_r(t)\cdot n_r(t)$ variables correspond to edges between two red nodes and hence take a value between 0 and $\alpha \leq M_{\alpha\beta}$ while $n_r(t)\cdot n_b(t)$ variables correspond to edges between a red node and a blue node and hence takes a value between 0 and $\beta \leq M_{\alpha\beta}$. Using these observations and the hoeffding inequality (3.1) and (6.3), we can see that

$$Pr(|w_r(t) - \mathbb{E}[w_r(t) \mid \mathcal{F}_t]| > \delta) < 2\mathrm{e}^{\frac{-2\delta^2}{n_r(t)\cdot(n+t)\cdot M_{\alpha\beta}^2}}$$

which is the first result required in the lemma.

Similarly, we can also see that

$$w(t) = \sum_{\mathrm{i}\in R(t),\mathrm{j}\in R(t)} w_{\mathrm{i,j}}(t) + 2\sum_{\mathrm{i}\in R(t),\mathrm{j}\in B(t)} w_{\mathrm{i,j}}(t) + \sum_{\mathrm{i}\in B(t),\mathrm{j}\in B(t)} w_{\mathrm{i,j}}(t) \tag{6.4}$$

We can interpret $w(t)$ as the sum of $n_r(t)\cdot n_r(t)+2n_r(t)\cdot n_b(t)+n_b(t)\cdot n_b(t) = (n_r(t))^2 + (n_b(t))^2+2n_b(t)\cdot n_r(t) = \Big(n_r(t)+n_b(t)\Big)^2 = (n+t)^2$ variables. Of these variables, $n_r(t)\cdot n_r(t)$ variables correspond to edges between two red nodes and hence take a value between 0 and $\alpha \leq M_{\alpha\beta}$ while $2n_r(t)\cdot n_b(t)$ variables correspond to edges between a red node and a blue node and hence takes a value between 0 and $\beta \leq M_{\alpha\beta}$ and finally $n_b(t)\cdot n_b(t)$ variables correspond to edges between two blue nodes and hence take a value between 0 and $\alpha \leq M_{\alpha\beta}$. Using these observations and (3.1) and (6.4), we can see that

$$Pr(|w(t) - \mathbb{E}[w(t) \mid \mathcal{F}_t]| > \delta) < 2\mathrm{e}^{\frac{-2\delta^2}{(n+t)^2\cdot M_{\alpha\beta}^2}}$$

which is the second result. □

Lemma 8. *Let $D > 0$ be some constant to be fixed later and, let $G_t(D)$ - be the event that both of the following statements*

$$\begin{cases} w_r(t) \geq \mathbb{E}[w_r(t) \mid \mathcal{F}_t] - M_{\alpha\beta}\cdot\sqrt{D}\cdot\sqrt{(n+t)\cdot n_r(t)\cdot log(n+t)} \\ w_r(t) \leq \mathbb{E}[w_r(t) \mid \mathcal{F}_t] + M_{\alpha\beta}\cdot\sqrt{D}\cdot\sqrt{(n+t)\cdot n_r(t)\cdot log(n+t)} \end{cases} \tag{6.5}$$

and

$$\begin{cases} w(t) \geq \mathbb{E}[w(t) \mid \mathcal{F}_t] - M_{\alpha\beta}\cdot\sqrt{D}\cdot(n+t)\cdot\sqrt{log(n+t)} \\ w(t) \leq \mathbb{E}[w(t) \mid \mathcal{F}_t] + M_{\alpha\beta}\cdot\sqrt{D}\cdot(n+t)\cdot\sqrt{log(n+t)} \end{cases} \tag{6.6}$$

hold true. Then, $\mathsf{Pr}[G_t(D)] \geq 1 - \frac{4}{(n+t)^{2D}}$

Proof. From (6.2), substitute $\delta = M_{\alpha\beta}\cdot\sqrt{D}\cdot\sqrt{(n+t)\cdot n_r(t)\cdot log(n+t)}$ and we have

$$\mathsf{Pr}\Big[|w_r(t) - \mathbb{E}[w_r(t) \mid \mathcal{F}_t]| > M_{\alpha\beta}\cdot\sqrt{D}\cdot\sqrt{(n+t)\cdot n_r(t)\cdot log(n+t)}\Big] < 2\mathrm{e}^{-2Dlog(n+t)}$$

Using the properties of logarithms, this changes to

$$\mathsf{Pr}\Big[|w_r(t) - \mathbb{E}[w_r(t) \mid \mathcal{F}_t]| > M_{\alpha\beta}\cdot\sqrt{D}\cdot\sqrt{n(t)\cdot n_r(t)\cdot log(n+t)}\Big] < \frac{2}{(n+t)^{2D}}$$

Thus, we can say that the following expression

$$\begin{cases} w_r(t) \geq \mathbb{E}[w_r(t) \mid \mathcal{F}_t] - M_{\alpha\beta} \cdot \sqrt{D} \cdot \sqrt{(n+t) \cdot n_r(t) \cdot log(n+t)} \\ w_r(t) \leq \mathbb{E}[w_r(t) \mid \mathcal{F}_t] + M_{\alpha\beta} \cdot \sqrt{D} \cdot \sqrt{(n+t) \cdot n_r(t) \cdot log(n+t)} \end{cases} \tag{6.7}$$

holds with probability $\frac{2}{(n+t)^{2D}}$.

When we substitute $\delta = M_{\alpha\beta} \cdot \sqrt{D} \cdot (n+t) \cdot \sqrt{log(n+t)}$ in (6.2), we will get

$$\mathsf{Pr}\Big[|w(t) - \mathbb{E}[w(t) \mid \mathcal{F}_t]| > M_{\alpha\beta} \cdot \sqrt{D} \cdot (n+t) \cdot \sqrt{log(n+t)}\Big] < 2\mathrm{e}^{-2Dlog(n+t)}$$

Using the properties of logarithms, this changes to

$$\mathsf{Pr}\Big[|w(t) - \mathbb{E}[w(t) \mid \mathcal{F}_t]| > M_{\alpha\beta} \cdot \sqrt{D} \cdot (n+t) \cdot \sqrt{log(n+t)}\Big] < \frac{2}{(n+t)^{2D}}$$

$$\begin{cases} w(t) \geq \mathbb{E}[w(t) \mid \mathcal{F}_t] - M_{\alpha\beta} \cdot \sqrt{D} \cdot (n+t) \cdot \sqrt{log(n+t)} \\ w(t) \leq \mathbb{E}[w(t) \mid \mathcal{F}_t] + M_{\alpha\beta} \cdot \sqrt{D} \cdot (n+t) \cdot \sqrt{log(n+t)} \end{cases} \tag{6.8}$$

holds with probability $\frac{2}{(n+t)^{2D}}$.

Combining (6.7) and (6.8) and applying the union bound, we obtain that $G_t(D)$ holds with probability $\frac{4}{(n+t)^{2D}}$ □

Lemma 9. *On the event $G_t(D)$, if $0 < D < \left(m_{\alpha,\beta} \cdot \frac{m_{a,b}}{M_{\alpha\beta}}\right)^2 \cdot (\frac{1}{h(t)})^2$, then $w(t) > 0$.*

Proof. All of the following expressions hold on the event $G_t(D)$ and $D > 0$. First, if $G_t(D)$ holds true, and $D > 0$, then we can apply Lemma 8 and thus we have

$$\mathbb{E}[w(t) \mid \mathcal{F}_t] - M_{\alpha\beta} \cdot \sqrt{D} \cdot (n+t) \cdot \sqrt{log(n+t)} \leq w(t)$$

Thus, if we can show

$$\mathbb{E}[w(t) \mid \mathcal{F}_t] - M_{\alpha\beta} \cdot \sqrt{D} \cdot (n+t) \cdot \sqrt{log(n+t)} > 0 \tag{6.9}$$

then we can show that $w(t) > 0$.

(6.9) is true when

$$\mathbb{E}[w(t) \mid \mathcal{F}_t] > M_{\alpha\beta} \cdot \sqrt{D} \cdot (n+t) \cdot \sqrt{log(n+t)}$$

Apply Lemma 6 and we just need to show

$$m_{a,b} \cdot m_{\alpha,\beta} \cdot (n+t)^2 > M_{\alpha\beta} \cdot \sqrt{D} \cdot (n+t) \cdot \sqrt{log(n+t)}$$

Cancelling $n+t$ on both sides and cross multiplying, we instead have

$$\frac{m_{a,b} \cdot m_{\alpha,\beta} \cdot (n+t)}{M_{\alpha\beta} \cdot \sqrt{log(n+t)}} > \sqrt{D}$$

Squaring on both sides, we finally have

$$D < \left(m_{a,b} \cdot \frac{m_{\alpha,\beta}}{M_{\alpha\beta}}\right)^2 \cdot \left(\frac{n+t}{\sqrt{log(n+t)}}\right)^2$$

and thus the lemma holds true under the assumption $0 < D < \left(m_{\alpha,\beta} \cdot \frac{m_{a,b}}{M_{\alpha\beta}}\right)^2 \cdot \left(\frac{1}{h(t)}\right)^2$. $\square$

We define some new notation.

$$\begin{cases} u = \dfrac{M_{\alpha\beta} \cdot \sqrt{D}}{m_{a,b} \cdot m_{\alpha,\beta}} \\ v = \dfrac{\sqrt{D}}{M_{ab}} \end{cases}$$

Lemma 10. *On the event $G_t(D)$, the following inequalities hold:*

$$\begin{cases} \frac{w_r(t)}{w(t)} \geq \Big(1 + u \cdot h(t)\Big)^{-1} \cdot \Big(Y_d(t) - v \cdot h(t)\Big) \\ \frac{w_r(t)}{w(t)} \leq \Big(1 - v \cdot h(t)\Big)^{-1} \cdot \Big(Y_d(t) + u \cdot h(t)\Big) \end{cases} \tag{6.10}$$

Proof. All of the following expressions assume that the event $G_t(D)$ happens.

From (6.5) and Lemma 6

$$\frac{w_r(t)}{\mathbb{E}[w(t) \mid \mathcal{F}_t]} \leq \frac{\mathbb{E}[w_r(t) \mid \mathcal{F}_t]}{\mathbb{E}[w(t) \mid \mathcal{F}_t]} + \frac{M_{\alpha\beta} \cdot \sqrt{D} \cdot \sqrt{(n+t) \cdot n_r(t) \cdot log(n+t)}}{d \cdot \epsilon(n+t)^2}$$

Thus, fixing $u = \frac{M_{\alpha\beta} \cdot \sqrt{D}}{m_{a,b} \cdot m_{\alpha,\beta}}$ and using the above expression, we get a new inequality

$$\frac{w_r(t)}{\mathbb{E}[w(t) \mid \mathcal{F}_t]} \leq \frac{\mathbb{E}[w_r(t) \mid \mathcal{F}_t]}{\mathbb{E}[w(t) \mid \mathcal{F}_t]} + \frac{u \cdot \sqrt{log(n+t)}}{(n+t)} \tag{6.11}$$

Again using (6.5) and Lemma 6, we have

$$\frac{w_r(t)}{\mathbb{E}[w(t) \mid \mathcal{F}_t]} \geq \frac{\mathbb{E}[w_r(t) \mid \mathcal{F}_t]}{\mathbb{E}[w(t) \mid \mathcal{F}_t]} - \frac{M_{\alpha\beta} \cdot \sqrt{D} \cdot \sqrt{(n+t) \cdot n_r(t) \cdot log(n+t)}}{M_{ab} \cdot M_{\alpha\beta}(n+t)^2}$$

Fixing $v = \frac{\sqrt{D}}{M_{ab}}$ and using the above expression, we get a new inequality

$$\frac{w_r(t)}{\mathbb{E}[w(t) \mid \mathcal{F}_t]} \geq \frac{\mathbb{E}[w_r(t) \mid \mathcal{F}_t]}{\mathbb{E}[w(t) \mid \mathcal{F}_t]} - \frac{v \cdot \sqrt{log(n+t)}}{(n+t)} \tag{6.12}$$

We will obtain similar inequalities for $w(t)$.

From (6.6) and Lemma 6, we get

$$\frac{w(t)}{\mathbb{E}[w(t) \mid \mathcal{F}_t]} \leq \frac{\mathbb{E}[w(t) \mid \mathcal{F}_t]}{\mathbb{E}[w(t) \mid \mathcal{F}_t]} + \frac{M_{\alpha\beta} \cdot \sqrt{D} \cdot \sqrt{log(n+t)}}{m_{a,b} \cdot m_{\alpha,\beta} \cdot (n+t)}$$

Thus, fixing $u = \frac{M_{\alpha\beta} \cdot \sqrt{D}}{m_{a,b} \cdot m_{\alpha,\beta}}$ and using the above expression, we get a new inequality

$$\frac{w(t)}{\mathbb{E}[w(t) \mid \mathcal{F}_t]} \leq 1 + \frac{u \cdot \sqrt{log(n+t)}}{(n+t)} \tag{6.13}$$

Finally, following similar steps as earlier, we also get the following inequality

$$\frac{w(t)}{\mathbb{E}[w(t) \mid \mathcal{F}_t]} \geq 1 - \frac{v \cdot \sqrt{log(n+t)}}{(n+t)} \tag{6.14}$$

Combine (6.13) and (6.12), we have

$$\frac{w_r(t)}{w(t)} \geq \big(1 + \frac{u \cdot \sqrt{log(n+t)}}{(n+t)}\big)^{-1} \cdot \big(\frac{\mathbb{E}[w_r(t) \mid \mathcal{F}_t]}{\mathbb{E}[w(t) \mid \mathcal{F}_t]} - \frac{v \cdot \sqrt{log(n+t)}}{(n+t)}\big)$$

Similarly, combining (6.11) and (6.14), we have

$$\frac{w_r(t)}{w(t)} \leq \big(1 - \frac{v \cdot \sqrt{log(n+t)}}{(n+t)}\big)^{-1} \cdot \big(\frac{\mathbb{E}[w_r(t) \mid \mathcal{F}_t]}{\mathbb{E}[w(t) \mid \mathcal{F}_t]} + \frac{u \cdot \sqrt{log(n+t)}}{(n+t)}\big)$$

Thus, we have proved the inequalities hold true on $G_t(D)$. In addition, because we only made an assumption that $G_t(D)$ happens and $\mathsf{Pr}[G_t(D)] \geq 1 - \frac{4}{(n+t)^{2D}}$, we have proved that the inequalities hold with probability at least $1 - \frac{4}{(n+t)^{2D}}$. $\square$

Lemma 11. *Let $0 < D < \left(m_{\alpha,\beta} \cdot \frac{m_{a,b}}{M_{\alpha\beta}}\right)^2 \cdot \left(\frac{1}{h(t)}\right)^2$, then on the event $G_t(D)$ the following inequalities are true:*

$$\begin{cases} \mathbb{E}[X(t+1) \mid \mathcal{F}_t] \geq \dfrac{1}{n+t+1} \cdot \left(X(t) \cdot (n+t) + \dfrac{Y_d(t) - v \cdot h(t)}{1 + u \cdot h(t)}\right) \\ \mathbb{E}[X(t+1) \mid \mathcal{F}_t] \leq \dfrac{1}{n+t+1} \cdot \left(X(t) \cdot (n+t) + \dfrac{Y_d(t) + u \cdot h(t)}{1 - v \cdot h(t)}\right) \end{cases}$$

Proof. From, (4.7) we have

$$\mathbb{E}[X(t+1) \mid \mathcal{F}_t] = \mathbb{E}\left[\frac{X(t) \cdot (n+t) + Y(t)}{n+t+1} \,\middle|\, \mathcal{F}_t\right] \tag{6.15}$$

By linearity of expectation, we obtain

$$\mathbb{E}[X(t+1) \mid \mathcal{F}_t] = \frac{X(t) \cdot (n+t) + \mathbb{E}[Y(t) \mid \mathcal{F}_t]}{n+t+1} \tag{6.16}$$

By Lemma 9, on the event $G_t(D)$ if $D < \left(m_{\alpha,\beta} \cdot \frac{m_{a,b}}{M_{\alpha\beta}}\right)^2 \cdot \left(\frac{n+t}{\sqrt{log(n+t)}}\right)^2$, then $w(t) > 0$. Thus as, $w(t) > 0$, (6.16) can be simplified to

$$\mathbb{E}[X(t+1) \mid \mathcal{F}_t] = \frac{X(t) \cdot (n+t) + \mathbb{E}\left[\frac{w_r(t)}{w(t)} \,\middle|\, \mathcal{F}_t\right]}{n+t+1} \tag{6.17}$$

On the event $G_t(D)$, (6.10) will hold true and thus

$$\mathbb{E}[X(t+1) \mid \mathcal{F}_t] \geq \left(\frac{1}{n+t+1} \cdot \left(X(t) \cdot (n+t) + \frac{Y_d(t) - v \cdot h(t)}{1 + u \cdot h(t)}\right)\right)$$

and

$$\mathbb{E}[X(t+1) \mid \mathcal{F}_t] \leq \frac{1}{n+t+1} \cdot \left(X(t) \cdot (n+t) + \frac{Y_d(t) + u \cdot h(t)}{1 - v \cdot h(t)}\right)$$

will hold true. Hence the required statement is true. □

We now define two new sets of $\sigma-$fields $\{\mathcal{H}_t\}_{t\geq 0}$ and $\{\tilde{\mathcal{H}}_t\}_{t\geq 0}$ where

- $\mathcal{H}_t$: is a 'good' history, i.e. a $\sigma-$field which denotes the history of the stochastic process after t rounds where the event $G_t(D)$ happens.
- $\tilde{\mathcal{H}}_t$: is a 'bad' history, i.e. a $\sigma-$field which denotes the history of the stochastic process after t rounds where the event $G_t(D)$ does not happen.

Theorem 6.1.1. *For the stochastic system $(X(t))_{t\geq 0}$, if $0 < D < \left(m_{\alpha,\beta} \cdot \frac{m_{a,b}}{M_{\alpha\beta}}\right)^2 \cdot \left(\frac{1}{h(t)}\right)^2$, the following expressions are satisfied.*

$$\begin{cases}\mathbb{E}[X(t+1) \mid \mathcal{F}_t] \leq \dfrac{1}{n+t+1}\left(X(t)\cdot(n+t) + \dfrac{Y_d(t)+u\cdot h(t)}{1-v\cdot h(t)}\cdot\big(1-\dfrac{4}{(n+t)^{2D}}\big) + \big(\dfrac{4}{(n+t)^{2D}}\big)\right)\\ \mathbb{E}[X(t+1) \mid \mathcal{F}_t] \geq \dfrac{1}{n+t+1}\left(X(t)\cdot(n+t) + \dfrac{Y_d(t)-v\cdot h(t)}{1+u\cdot h(t)}\cdot\big(1-\dfrac{4}{(n+t)^{2D}}\big)\right)\end{cases} \tag{6.18}$$

Proof. First, observe that

$$\mathbb{E}[X(t+1) \mid \tilde{\mathcal{H}}_t] \leq \frac{X(t)\cdot(n+t)+1}{n+t+1} \tag{6.19}$$

Inequality (6.19) is true because the indicator variable $Y(t)$ always takes a value between 0 and 1. This also means that

$$\mathbb{E}[X(t+1) \mid \tilde{\mathcal{H}}_t] \geq \frac{X(t)\cdot(n+t)}{n+t+1} \tag{6.20}$$

Since $\mathsf{Pr}\left[G_t(D)\right] + \mathsf{Pr}\left[G_t^c(D)\right] = 1$, we can apply the law of total expectation and we have

$$\mathbb{E}[X(t+1) \mid \mathcal{F}_t] = \mathbb{E}[X(t+1) \mid \mathcal{H}_t]\cdot\mathsf{Pr}[G_t(D)] + \mathbb{E}[X(t+1) \mid \tilde{\mathcal{H}}_t]\cdot\mathsf{Pr}[G_t^c(D)] \tag{6.21}$$

Combining, Lemma 8 and Lemma 11 and (6.19) and (6.21), we finally obtain

$$\mathbb{E}[X(t+1) \mid \mathcal{F}_t] \leq \frac{1}{n+t+1}\left(X(t)\cdot(n+t) + \frac{Y_d(t)+u\cdot h(t)}{1-v\cdot h(t)}\cdot\left(1-\frac{4}{(n+t)^{2D}}\right) + \left(\frac{4}{(n+t)^{2D}}\right)\right)$$

which is the required result.

For the other inequality, combine Lemma 8 and Lemma 11 and (6.20) and (6.21) and obtain

$$\mathbb{E}[X(t+1) \mid \mathcal{F}_t] \geq \frac{1}{n+t+1}\left(X(t)\cdot(n+t) + \frac{Y_d(t) - v\cdot h(t)}{1+u\cdot h(t)}\cdot\left(1 - \frac{4}{(n+t)^{2D}}\right)\right)$$

□

Theorem 6.1.1 has great implications. From the following argument, one could see that it implies that the behaviour of the stochastic system is almost identical to that of the deterministic system when we have a lot of nodes and thus our earlier results regarding convergence of the deterministic system apply for the stochastic system as well.

First, from (5.1), we have

$$X_d(t+1) = \frac{X_d(t)\cdot(n+t) + Y_d(t)}{n+t+1}$$

and from (4.7) we have

$$X(t+1) = \frac{X(t)\cdot(n+t) + Y(t)}{n+t+1}$$

From the above equations, we can see that if $X(t) = X_d(t)$ then

$$Y(t) = Y_d(t) \implies X(t+1) = X_d(t+1)$$

Thus, if $X(t) = X_d(t)$

$$\lim_{t\to\infty} Y(t) = Y_d(t) \implies \lim_{t\to\infty} X(t+1) = X_d(t+1) \tag{6.22}$$

Comparing (6.18) and (4.7), we can see that

$$\begin{cases} \mathbb{E}[Y(t) \mid \mathcal{F}_t] \leq \dfrac{Y_d(t) + u \cdot h(t)}{1 - v \cdot h(t)} \cdot \Big(1 - \dfrac{4}{(n+t)^{2D}}\Big) + \Big(\dfrac{4}{(n+t)^{2D}}\Big) \\ \mathbb{E}[Y(t) \mid \mathcal{F}_t] \geq \dfrac{Y_d(t) - v \cdot h(t)}{1 + u \cdot h(t)} \cdot \Big(1 - \dfrac{4}{(n+t)^{2D}}\Big) \end{cases} \tag{6.23}$$

From (6.1), we have $h(t) = \frac{\sqrt{log(n+t)}}{n+t}$ and thus

$$\lim_{t\to\infty} h(t) = \lim_{t\to\infty} \frac{\sqrt{log(n+t)}}{n+t}$$

Applying L'Hospital's rule, we have

$$\lim_{t\to\infty} h(t) = \lim_{t\to\infty} \frac{1}{2\,(n+t)\,\sqrt{\ln{(n+t)}}} = 0 \tag{6.24}$$

Thus, from (6.23) and (6.24) we have

$$\lim_{t\to\infty} \mathbb{E}[Y(t) \mid \mathcal{F}_t] \leq \lim_{t\to\infty} \frac{Y_d(t) + u \cdot h(t)}{1 - v \cdot h(t)} \cdot \Big(1 - \frac{4}{(n+t)^{2D}}\Big) + \Big(\frac{4}{(n+t)^{2D}}\Big) = \frac{Y_d(t) + u \cdot 0}{1 - v \cdot 0} \cdot \Big(1 - 0\Big) + \Big(0\Big)$$

which becomes

$$\lim_{t\to\infty} \mathbb{E}[Y(t) \mid \mathcal{F}_t] \leq Y_d(t) \tag{6.25}$$

Taking the upper bound from (6.23), and (6.24) we also have

$$\lim_{t\to\infty} \mathbb{E}[Y(t) \mid \mathcal{F}_t] \geq \lim_{t\to\infty} \frac{Y_d(t) - v \cdot h(t)}{1 + u \cdot h(t)} \cdot \Big(1 - \frac{4}{(n+t)^{2D}}\Big) = \frac{Y_d(t) - v \cdot 0}{1 + u \cdot 0} \cdot \Big(1 - 0\Big)$$

which becomes

$$\lim_{t\to\infty} \mathbb{E}[Y(t) \mid \mathcal{F}_t] \geq Y_d(t) \tag{6.26}$$

Combine (6.25) and (6.26) we can finally see that

$$\lim_{t\to\infty} \mathbb{E}[Y(t) \mid \mathcal{F}_t] = Y_d(t)$$

This shows that in the limit, the stochastic system behaves in a way that is similar to the deterministic system.

7. CONCLUSION AND FUTURE WORK

7.1 How accurate an approximation is the deterministic system?

Theorem 3.4.1 by Warnke [8] provides a framework for proving that the deterministic system is a good approximation of the stochastic block model with high probability. Application of this theorem is quite involved however and requires a number of pre-conditions to be satisfied. The most challenging step would be to get a useful Lipschitz-constant L for the function F which is used to predict the status of the stochastic model after a time period. We do not show our entire work, but it is possible to show that $L \leq \frac{1}{n+1}$ when $0.1 \leq X(t) \leq 0.9\ \forall t \in [0, T]$. This value of the Lipschitz constant will enable us to get a very good approximation with high probability as long as the population does not increase significantly i.e., as long as the population does not increase several fold of it's initial value. Unfortunately, when we try to use Wormald's theorem for the long term, the approximation is valid only with negligible probability.

7.2 Interpretation of our results

Our concluding discussion in the stochastic system shows that its behaviour would be very similar to the deterministic system in the limit. Combined with 5.3.1, this means that when $\rho = \frac{a\alpha}{b\beta} > 1$ the minority would essentially vanish from the organization in the long run. Thus, to ensure that the minority could thrive in the long run, we need to make $\beta > \frac{a\alpha}{b}$. Recall that β is the weight of an edge between a red node and blue node. Thus, in this model, one would have to provide incentives (which could be additional funding or institutional support) for any collaborations to ensure that the minority thrives in the long run.

7.3 Future Work

There are a number of questions related to this model that we think would be interesting to look at:

- Can we formally show that the stochastic system converges in the same way as the deterministic system?
- What happens when we allow more than two communities?
- What happens if we allow different edge weights for red-red edges and blue-blue edges. We hypothesize that the system behaviour would be similar, but convergence would happen to a point other than $X(t) = \frac{1}{2}$.
- Can we obtain theoretical bound on the approximation of the stochastic system in the short term? Our experiments suggest that a bound much better than what is acheived by Wormald's theorem is possible.

8. NOTATION

8.1 Variables, constants and functions introduced for the model

- $c(\mathrm{i})-$ denotes the color of the node i
- $\nu = \frac{a}{b}-$ where a, b are model parameters
- $\lambda = \frac{\beta}{\alpha}$ where α, β are model parameters
- $\rho = \frac{\nu}{\lambda}$
- $w(t) = \sum_{\mathrm{i,j}=1}^{n} w_{\mathrm{i,j}}(t)$
- $w_r(t) = \sum_{\mathrm{i}:c_{\mathrm{i}}=red} \sum_{\mathrm{j}} w_{\mathrm{i,j}}(t)$
- $w_b(t) = \sum_{\mathrm{i}:c_{\mathrm{i}}=blue} \sum_{\mathrm{j}} w_{\mathrm{i,j}}(t)$
-

P	Red	Blue
Red	a	b
Blue	b	a

w	Red	Blue
Red	α	β
Blue	β	α

- $n + t-$ number of nodes at time t
- $X(t)-$ fraction of red nodes at time t
- $n_r(t) = X(t) \cdot (n+t)-$ number of red nodes at time t
- $n_b(t) = (1 - X(t)) \cdot (n+t)-$ number of blue nodes at time t

8.2 Variables, constants and functions introduced for the deterministic version

- $Y_d(t) = \frac{\mathbb{E}[w_r(t)]}{\mathbb{E}[w(t)]}$

8.3 Variables, constants and functions introduced for Stochastic version

- $M_{ab} = max(a, b)$
- $m_{ab} = min(a, b)$

- $M_{\alpha\beta} = max(\alpha, \beta)$

- $m_{\alpha\beta} = min(\alpha, \beta)$

- $u = \frac{M_{\alpha,\beta} \cdot \sqrt{D}}{m_{a,b} \cdot m_{\alpha,\beta}}$

- $v = \frac{\sqrt{D}}{M_{a,b}}$

- $h(t) = \frac{\sqrt{log(n+t)}}{n+t}$

REFERENCES

[1] L. T. Liu, S. Dean, E. Rolf, M. Simchowitz, and M. Hardt, "Delayed impact of fair machine learning," in *International Conference on Machine Learning*, PMLR, 2018, pp. 3150–3158.

[2] P. Gajane and M. Pechenizkiy, "On formalizing fairness in prediction with machine learning," *arXiv preprint arXiv:1710.03184*, 2017.

[3] A. Chouldechova and A. Roth, "The frontiers of fairness in machine learning," *arXiv preprint arXiv:1810.08810*, 2018.

[4] N. Mehrabi, F. Morstatter, N. Saxena, K. Lerman, and A. Galstyan, "A survey on bias and fairness in machine learning," *arXiv preprint arXiv:1908.09635*, 2019.

[5] L. Hu and Y. Chen, "A short-term intervention for long-term fairness in the labor market," in *Proceedings of the 2018 World Wide Web Conference*, 2018, pp. 1389–1398.

[6] L. T. Liu, A. Wilson, N. Haghtalab, A. T. Kalai, C. Borgs, and J. Chayes, "The disparate equilibria of algorithmic decision making when individuals invest rationally," in *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 2020, pp. 381–391.

[7] N. C. Wormald *et al.*, "The differential equation method for random graph processes and greedy algorithms," *Lectures on approximation and randomized algorithms*, vol. 73, p. 155, 1999.

[8] L. Warnke, "On wormald's differential equation method," *arXiv preprint arXiv:1905.08928*, 2019.

[9] T. Bohman and P. Keevash, "The early evolution of the h-free process," *Inventiones mathematicae*, vol. 181, no. 2, pp. 291–336, 2010.

[10] H. Amini and A. Minca, "Pandemic risks and equilibrium social distancing in heterogeneous networks," *Available at SSRN 3645889*, 2020.

[11] S. Branzei and Y. Peres, "Consensus with bounded space and minimal communication," *arXiv preprint arXiv:2101.00025*, 2020.

[12] P. Manurangsi and W. Suksompong, "Closing gaps in asymptotic fair division," *SIAM Journal on Discrete Mathematics*, vol. 35, no. 2, pp. 668–706, 2021.

[13] S. Yang and R. W. Yeung, "Batched sparse codes," *IEEE Transactions on Information Theory*, vol. 60, no. 9, pp. 5322–5346, 2014.

[14] E. Abbe, "Community detection and stochastic block models: Recent developments," *The Journal of Machine Learning Research*, vol. 18, no. 1, pp. 6446–6531, 2017.

[15] S. Heimlicher, M. Lelarge, and L. Massoulié, "Community detection in the labelled stochastic block model," *arXiv preprint arXiv:1209.2910*, 2012.

[16] E. Abbe, A. S. Bandeira, and G. Hall, "Exact recovery in the stochastic block model," *IEEE Transactions on Information Theory*, vol. 62, no. 1, pp. 471–487, 2015.

[17] V. Jog and P.-L. Loh, "Information-theoretic bounds for exact recovery in weighted stochastic block models using the renyi divergence," *arXiv preprint arXiv:1509.06418*, 2015.

[18] E. Abbe and C. Sandon, "Recovering communities in the general stochastic block model without knowing the parameters," *arXiv preprint arXiv:1506.03729*, 2015.

[19] C. Gao, Z. Ma, A. Y. Zhang, and H. H. Zhou, "Achieving optimal misclassification proportion in stochastic block models," *The Journal of Machine Learning Research*, vol. 18, no. 1, pp. 1980–2024, 2017.

[20] M. Xu, V. Jog, P.-L. Loh, *et al.*, "Optimal rates for community estimation in the weighted stochastic block model," *The Annals of Statistics*, vol. 48, no. 1, pp. 183–204, 2020.

[21] W. Hoeffding, "Probability inequalities for sums of bounded random variables," in *The Collected Works of Wassily Hoeffding*, Springer, 1994, pp. 409–426.