UNIVERSIDADE DE SÃO PAULO
Instituto de Ciências Matemáticas e de Computação

Rumour spreading in dynamic random graphs

Vicenzo Bonasorte Reis Pereira

Dissertação de Mestrado do Programa Interinstitucional de Pós-Graduação em Estatística (PIPGEs)

ICMC USP
SÃO CARLOS

SERVIÇO DE PÓS-GRADUAÇÃO DO ICMC-USP

Data de Depósito:

Assinatura: ______________________

Vicenzo Bonasorte Reis Pereira

Rumour spreading in dynamic random graphs

Dissertation submitted to the Institute of Mathematics and Computer Science – ICMC-USP and to the Department of Statistics – DEs-UFSCar – in accordance with the requirements of the Statistics Interagency Graduate Program, for the degree of Master in Statistics. *EXAMINATION BOARD PRESENTATION COPY*

Concentration Area: Statistics

Advisor: Prof. Dr. Alexsandro Giacomo Grimbert Gallo

USP – São Carlos
February 2024

Ficha catalográfica elaborada pela Biblioteca Prof. Achille Bassi
e Seção Técnica de Informática, ICMC/USP,
com os dados inseridos pelo(a) autor(a)

B699r

Bonasorte Reis Pereira, Vicenzo
Rumour spreading in dynamic random graphs / Vicenzo Bonasorte Reis Pereira; orientador Alexsandro Giacomo Grimbert Gallo. -- São Carlos, 2024.
54 p.

Dissertação (Mestrado - Programa Interinstitucional de Pós-graduação em Estatística) -- Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, 2024.

1. Propagação de informação. 2. Tempos estacionários fortes. 3. Grafos aleatórios dinâmicos. 4. Modelo estocástico de blocos. I. Giacomo Grimbert Gallo, Alexsandro, orient. II. Título.

Bibliotecários responsáveis pela estrutura de catalogação da publicação de acordo com a AACR2:
Gláucia Maria Saia Cristianini - CRB - 8/4938
Juliana de Souza Moraes - CRB - 8/6176

Vicenzo Bonasorte Reis Pereira

Propagação de rumor em grafos aleatórios dinâmicos

Dissertação apresentada ao Instituto de Ciências Matemáticas e de Computação – ICMC-USP e ao Departamento de Estatística – DEs-UFSCar, como parte dos requisitos para obtenção do título de Mestre em Estatística – Programa Interinstitucional de Pós-Graduação em Estatística. *EXEMPLAR DE DEFESA*

Área de Concentração: Estatística

Orientador: Prof. Dr. Alexsandro Giacomo Grimbert Gallo

USP – São Carlos
Fevereiro de 2024

UNIVERSIDADE FEDERAL DE SÃO CARLOS

Centro de Ciências Exatas e de Tecnologia
Programa Interinstitucional de Pós-Graduação em Estatística

Folha de Aprovação

Defesa de Dissertação de Mestrado do candidato Vicenzo Bonasorte Reis Pereira, realizada em 19/02/2024.

Comissão Julgadora:

Prof. Dr. Alexsandro Giacomo Grimbert Gallo (UFSCar)

Prof. Dr. Élcio Lebensztayn (UNICAMP)

Prof. Dr. Giulio Iacobelli (UFRJ)

O Relatório de Defesa assinado pelos membros da Comissão Julgadora encontra-se arquivado junto ao Programa Interinstitucional de Pós-Graduação em Estatística.

ACKNOWLEDGEMENTS

I would like to express my deepest appreciation to my advisor Prof. Sandro Gallo, for his patience and teachings.

I also express my gratitude to my beloved family, for their unconditional support.

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001 and by grant 2022/08843-6, São Paulo Research Foundation (FAPESP).

ABSTRACT

PEREIRA, V. B. R. **Rumour spreading in dynamic random graphs**. 2024. 54 p. Dissertação (Mestrado em Estatística – Programa Interinstitucional de Pós-Graduação em Estatística) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2024.

We study rumour spreading in dynamic random graphs. Starting with a single informed vertex, the information flows until it reaches all the vertices of the graph (completion), according to the following process. At each step k, the information is propagated to neighbours, in the k-th generated random graph, of the informed vertices. The way this information propagates from vertex to vertex at each step will depend of the "protocol". First we consider a sequence of graphs in which the presence or absence of an edge follows the dynamic of a Markov chain. We provide a method based on strong stationary times allowing to bound the completion time for the Markov dynamic using bounds on the completion time in the i.i.d. dynamic.

We also consider the rumour spreading according to the Push Protocol (at every round, informed nodes send the rumour to one of their neighbours, chosen uniformly at random) in a sequence of independent stochastic block model random graphs. We are able to bound the completion time in this setting using comparisons with rumour spreading in dynamic random graphs with skeptical nodes (nodes that cannot become informed) and stifler nodes (nodes that, after being informed, do not spread the information further).

Keywords: Dynamic random graphs, Rumour spreading, Stochastic block model, Strong stationary times.

RESUMO

PEREIRA, V. B. R. **Propagação de rumor em grafos aleatórios dinâmicos**. 2024. 54 p. Dissertação (Mestrado em Estatística – Programa Interinstitucional de Pós-Graduação em Estatística) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2024.

Nós estudamos propagação de rumor em grafos aleatórios dinâmicos. Começando com um único vértice informado, a informação se propaga até atingir todos os vértices do grafo (finalização), de acordo com o seguinte processo. A cada passo k, a informação é enviada, no k-ésimo grafo aleatório gerado, para os vizinhos de vértices informados. O modo como essa informação é propagada de vértice para vértice a cada passo depende do "protocolo". Primeiro consideramos uma sequência de grafos em que a presença e ausência de uma aresta segue a dinâmica de uma cadeia de Markov. Propomos um método baseado em tempos estacionários fortes que permite limitar o tempo até a finalização na dinâmica markoviana utilizando limitantes do tempo até a finalização no caso i.i.d..

Também consideramos o rumor se espalhando através do protocolo Push (a cada passo, vértices informados enviam o rumor para um de seus vizinhos, escolhido uniformemente ao acaso) em uma sequência de grafos independentes do modelo estocástico de blocos. Somos capazes de encontrar limitantes para o tempo até a finalização utilizando comparações com propagação de informação em grafos aleatórios dinâmicos com vértices céticos (vértices que não se tornam informados) e vértices contidos (vértices, que após serem informados, não passam a informação adiante).

Palavras-chave: Grafos aleatórios dinâmicos, Propagação de rumor, Modelo estocástico de blocos, Tempos estacionários fortes..

Contents

Chapter 1

Introduction

The study of information spreading in networks has found many applications in recent years. The efforts to contain the COVID-19 pandemic benefited greatly from the existing mathematical theory of epidemiology. Social Networks, such as Facebook and Twitter, had a great impact on recent presidential elections all over the world, making it crucial to gain insight into how fake news spreads on random networks. There are also technological applications to information spreading in random networks, such as the distribution of an update in replicated databases. In computation theory, such problems are studied under the denomination of communication protocols.

Communication protocols on networks. Rumour Spreading is a well-known algorithm for broadcasting information on a large network. Some of the more important versions of this algorithm are the Push Protocol, Flood Protocol, Pull Protocol and Push-Pull Protocol. In all these Protocols we start with a single informed node. In the Push version every round, an informed node chooses one of its neighbours uniformly at random and informs it. In the Flood version, informed nodes inform every neighbour each round. In the Pull version, uninformed nodes choose one of their neighbours, if the chosen node is informed, the initial one also becomes informed. The Push-Pull Protocol combines Push and Pull, informed nodes will send the information to a chosen neighbour and uninformed nodes will try to get the information from one chosen neighbour.

Rapid overview of the literature on random networks. CLEMENTI *et al.* (2016) studied the Push version of this algorithm in an underlying dynamic network. In one setting each graph is a newly sampled (independently of everything else) Erdős-Rényi random graph, $ER_n(p)$. They showed that the completion time (the time until every node is informed) for the Push Protocol is $O(\frac{\log n}{\min\{1,np\}})$ with high probability, that is with probability decreasing polinomially to 0 as n diverges. DOERR e KOSTRYGIN (2017) improved this bound when the edge probability is $\frac{a}{n}$, where n is the number of edges and a is a positive constant. They introduced a method for analyzing rumour spreading that doesn't rely on the specific definition of the protocol, under some sufficient conditions they provide the expectation and a large deviations bound for the completion time for several settings apart from additive constants. The expected completion time (T) is $\log_{2-e^{-a}}(n) + \frac{1}{1-exp(-a)}\log(n) + O(1)$ and $\mathbb{P}(|T - \mathbb{E}(T)| > r) \leq A\exp(-\alpha r)$, where A and α are positive constants.

DAKNAMA (2017) uses the method above to analyze rumour spreading in a sequence of Erdös-Rényi graphs for the Pull and Push-Pull Protocol. He proves that the expected completion time for the Pull Protocol is $\log_{2-e^{-a}}(n) + \frac{1}{a\log(n)} + O(1)$ and for Push-Pull is $\log_{1+\gamma}(n) + \frac{1}{a}\log(n) + O(1)$, for a constant γ.

CLEMENTI *et al.* (2010) and CLEMENTI *et al.* (2016) analyzed rumour spreading in edge-markovian random graphs. In this dynamic graph, there is a 2-state Markov chain attached to each possible edge (independent of one another). If an edge was absent at time t, it will be present at time $t+1$ with probability p, if it was present at time t, it will be absent at time $t+1$ with probability q. In the former they considered the Flood Protocol and prove that for any p and q the completion time is $O(\frac{\log n}{\log(1+np)})$ with high probability. In the latter they considered the rumour spreading according to the Push Protocol and prove that, for $p \geq \frac{1}{n}$ and when q is constant (does not depend on n), the completion time is $O(\log n)$, with high probability.

Main contributions of this text. In this text, we work on this models on two main directions. First, we provide a method based on strong stationary times that, under some sufficient conditions on the edge transition matrix, is able to bound the completion time of several information spreading protocols: Push, Pull, Flooding and Push-Pull. In the Push Protocol setting we prove that the completion time is $O(n^{k-1}\log n)$ when $p = \frac{a}{n^k}$ and $q = 1$, for constants $a > 0$ and $k > 1$, which had not been done previously in the literature.

We also consider two variations of the rumour spreading through the Push Protocol in a sequence of independent random graphs. In the first one we introduce skeptical nodes, that is nodes that cannot be informed. They have probability p_2 of connecting to informed/uninformed nodes. Non-skeptical nodes have probability p_1 of connecting to informed/uninformed nodes. We prove that the completion time is $O(\frac{\log n}{\min\{1,np_1\}})$ with high probability when the number of skeptical nodes is a fixed fraction of the total amount of nodes and $p_2 = O(p_1)$.

In the second variation we introduce bots and stiflers. Bots start informed. Stiflers, when informed, have no interest of spreading the information further. Bots have probability p_1 of connecting amongst themselves and probability p_2 of connecting to stiflers. We prove that the completion time is $O(\frac{\log n}{\min\{1,np_2\}})$, when the number of bots and stiflers are fixed fractions of the total amount of nodes and $p_1 = O(p_2)$.

Lastly, we provide upper bounds for the rumour spreading through the Push Protocol in a sequence of stochastic block model graphs with two communities. Nodes in community $i = 1, 2$ have probability p_i of connecting to another node in community i and probability p_{12} of connecting to a node in the other community. We prove that if $p_1 = \Theta(p_{12})$ and $p_2 = \Omega(p_{12})$ or if $p_1 = \theta(p_{12})$ and $p_2 = o(p_{12})$ the completion time is $O(\frac{\log n}{\min\{1,np_{12}\}})$ with high probability.

Organization of the Dissertation. The next Chapter is divided into three Sections. In the first Section, we define rumour spreading in a sequence of independent Erdős-Rényi random graphs and present theorems of the literature that bound the completion time for the Flood, Push, Pull and Push-Pull protocols. In the second section we define rumour spreading in a sequence of random graphs with markovian edges and give a method to bound the completion time using strong stationary times. In the last section we provide theorems that bound the completion time for rumor spreading according to the Push Protocol in a sequence of independent random graphs with communities. Chapter 3 provides proofs for the new Theorems presented in Chapter 2.

Chapter 2

Rumour Spreading in Dynamic Random Graphs: Definitions and New Results

2.1 Rumour Spreading in a Sequence of Independent Erdös-Rényi Graphs

A graph is a pair of sets, $G = (V, E)$, where V is a set of vertices and E a set of pairs of vertices, called edges. Consider a sequence of random graphs $\{G_t\}_{t \geq 0}$, where $G_t = ([n], E_t)$. We have a fixed vertex set $[n] := \{1, 2, \ldots, n\}$ and the edges set E_t evolves in time, at every instant an edge is present with probability $0 < p < 1$ and absent with probability $1 - p$, therefore $\{G_t\}_{t \geq 0}$ is a sequence of independent Erdős-Rényi graphs with edge parameter p (ER$_n(p)$ in what follows). Let $I \subseteq [n]$ be the subset of informed nodes. We start with a single informed node. For $t \in \mathbb{N}$ we generate a new set of edges and the information spreads according to some Protocol. Some of the classic synchronous protocols in the distributed systems literature are Push Protocol, Flood Protocol, Pull Protocol, Push-Pull Protocol, k-Push Protocol, k-Pull Protocol. Our interest is to get information concerning the completion time: the time until every node is informed, that is,$I = [n]$. We aim to find bounds that hold with high probability, that is, with probability at least $1 - \frac{1}{n^\alpha}$, $\alpha > 0$. The definitions of the asymptotic notation used further can be seen in Appendix A

Push Protocol. In the synchronous version of the Push Protocol, at every time step, each of the informed nodes chooses one of its neighbours uniformly at random. If the neighbour was uninformed, it becomes informed.

The number of informed nodes can at most double at every time step, when at each step, each $v \in I$ chooses a distinct uninformed node. In this case the completion time is $\lceil \log_2 n \rceil$

Theorem 2.1, proved by CLEMENTI *et al.* (2016), gives an upper bound on the completion time.

Theorem 2.1 *Let $\mathscr{G} = (G_t)_{t\in\mathbb{N}}$ be a sequence of independent ER$_n(p)$:*
(a) if $p \geq \frac{1}{n}$, the completion time of the push protocol over $\mathscr{G}$ is $O(\log n)$, with high probability.
(b)if $p < \frac{1}{n}$, the completion time of the push protocol over $\mathscr{G}$ is $O(\frac{\log n}{np})$, with high probability.

We note that, since we need at least $\log_2 n$ steps until every vertex is informed, the completion time is $\Theta(\log n)$ when $p = \frac{a}{n}$, $a > 0$.

We can generalize the Push Protocol, allowing the informed nodes to send the rumour to $k-1$, $k > 2$, of its neighbours, chosen uniformly at random. We will call this protocol k-Push Protocol. Upper bounds for the simple Push Protocol are also upper bounds for the k-Push Protocol, since the latter informs at least the same amount of nodes.

Flood Protocol. The Flood Protocol is the fastest information spreading algorithm we will consider. We start with a single informed node. At every time step, each informed node sends the rumour to all of its neighbours. The following Theorem 2.2 of CLEMENTI *et al.* (2010), bounds the maximum completion time over all possible sources, given any sequence of ER$_n(p)$.

Theorem 2.2 *Let $\mathscr{G} = (G_t)_{t\in\mathbb{N}}$ be a sequence of independent ER$_n(p)$. The completion time of the Flood Protocol over $\mathscr{G}$ is $O\left(\frac{\log n}{\log(1+np)}\right)$, with high probability.*

Pull Protocol. In the Pull Protocol we start with a single informed node. At every time step, each uninformed vertex asks one of its neighbours, chosen uniformly at random, for the information. If the chosen neighbour was already informed, the asking node becomes informed. DAKNAMA (2017) gives the expectation and a large deviations bound for the completion time of the Pull Protocol over a sequence of $\text{ER}_n(p)$ with $p = \frac{a}{n}$, $a > 0$.

Theorem 2.3 *Let $\mathscr{G} = (G_t)_{t\in\mathbb{N}}$ be a sequence of independent $ER_n(p)$ with $p = \frac{a}{n}$, $a > 0$. Let T_n be the completion time of the Pull Protocol over $\mathscr{G}$. Then:*

$$\mathbb{E}(T_n) = \log_{2-\exp(-a)}(n) + \frac{1}{a}\log n + O(1),$$

and there are constants $A, \alpha > 0$ such that $\forall\, r, n \in \mathbb{N}$:

$$\mathbb{P}(|T_n - \mathbb{E}(T_n)| \geq r) \leq A\exp(-\alpha r).$$

In particular, Theorem 2.3 implies that the completion time of the Pull Protocol over $\mathscr{G}$ is $O(\log n)$ with high probability.

In similar fashion to the Push Protocol, we can also consider the k-Pull Protocol, where uninformed nodes ask $k-1$ of its neighbours, chosen uniformly at random, for the information. If at least one of them is already informed, the asking vertex becomes informed. Upper bounds for the simple Pull Protocol are also upper bounds for the k-Pull Protocol.

Push-Pull Protocol. The Push-Pull Protocol is a combination of the Push Protocol and the Pull Protocol. We start with a single informed node. At every time step, each informed node chooses one of its neighbours uniformly at random and informs it. And each of the uninformed nodes asks one of its neighbours for the information, if the chosen neighbour is already informed, the asking nodes becomes informed. DAKNAMA (2017) gives the expectation and a large deviations bound for the completion time of the Push-Pull Protocol over a sequence of $\text{ER}_n(p)$ with $p = \frac{a}{n}$, $a > 0$.

Theorem 2.4 *Let $\mathscr{G} = (G_t)_{t\in\mathbb{N}}$ be a sequence of independent $ER_n(p)$ with $p = \frac{a}{n}$, $a > 0$. Let $\lambda = 2(1-\exp(-a)) - \frac{(1-\exp(-a))^2}{a}$ Let T_n be the completion time of the Push-Pull Protocol over $\mathscr{G}$. Then:*

$$\mathbb{E}(T_n) = \log_{1+\lambda}(n) + \frac{1}{a}\log n + O(1),$$

and there are constants $A, \alpha > 0$ such that $\forall\, r, n \in \mathbb{N}$:

$$\mathbb{P}(|T_n - \mathbb{E}(T_n)| \geq r) \leq A\,\text{exp}(-\alpha r).$$

In particular, Theorem 2.4 implies that the completion time of the Push-Pull Protocol over $\mathscr{G}$ is $O(\log n)$ with high probability.

2.2 Rumour Spreading in a Sequence of Random Graphs with Markovian Edges

In this Section we consider a generalization of the rumour spreading model in a sequence of $\text{ER}_n(p)$. Let $\{G_t\}_{t\geq 0}$ be a sequence of random graphs, with $G_t = ([n], E_t)$. In this variation we have a 2-state Markov chain attached to each edge (independently of all the other edges). The Markov chain has state space $S = \{0, 1\}$, 0 means that an edge is absent and 1 means the edge is present. We call this type of graph edge markovian random graphs. The transition matrix of the edge process is:

$$P = \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix}.$$

We will consider the rumour spreading according to the Protocols defined in Section 2.1. The model presented in Section 2.1 is a particular case of this model, taking $q = 1 - p$. We are interested in bounding the completion time.

By construction, the set E_t will only depend on E_{t-1}. $(E_t)_{t\geq 1}$ is also a Markov chain. Our strategy to obtain an upper bound is based on some properties of the chain $(E_t)_{t\geq 0}$.

$(E_t)_{t\geq 0}$ has transition probabilities given by the products of the transition probabilities of each one of the edges. Let $x = (x_1, x_2, \ldots, x_{\binom{n}{2}})$ and $y = (y_1, y_2, \ldots, y_{\binom{n}{2}})$ be two possible states of the chain. We have that the t-step transition probability between x and y is:

$$M^t_{x,y} = P^t_{x_1,y_1} P^t_{x_2,y_2} \ldots P^t_{x_{\binom{n}{2}},y_{\binom{n}{2}}},$$

where $P^t_{i,j}$ is the probability that an edge is in state $i \in \{0, 1\}$ at time s and be in state $j \in \{0, 1\}$ at time $s + t$. $M^t_{x,y}$ is the probability that the edges set is in state $x \in \{0, 1\}^{\binom{n}{2}}$ at time s and be in state $y \in \{0, 1\}^{\binom{n}{2}}$ at time $s + t$.

The same logic applies to the stationary distribution, it's given by the cartesian product of the stationary distributions of each edge. We can compute the stationary distribution of a single edge by solving $\pi P = \pi$, $\sum_{i\in\{0,1\}} \pi_i = 1$, there's an unique solution $\pi = (\frac{q}{p+q}, \frac{p}{p+q})$. Thus the stationary distribution of $(E_t)_{t\geq 0}$, Π, is:

$$\Pi = \pi \times \pi \cdots \times \pi,$$

where we are taking the cartesian product $\binom{n}{2}$ times.

(E_t) has finite state space, given by $\{0,1\}^{\binom{n}{2}}$. It is aperiodic if p or q are greater than 0 (there's a chance that the chain is in the same state two times in a row: $E_t = E_{t+1}$ with positive probability) and irreducible. For Markov chains having these three properties, the t-step transition probability converges to the stationary distribution as $t \to \infty$.

To bound the completion time, we use the strong stationary times theory, the necessary results can be seen in Appendix B. We prove the following.

Theorem 2.5 *Let $G = (G_t)_{t\geq 0} = ([n], E_t)$ be a sequence of edge markovian random graphs with transition matrix:*

$$P = \begin{pmatrix} 1 - f(n) & f(n) \\ 1 - g(n) & g(n) \end{pmatrix},$$

in which $f(n)$, $g(n)$ are decreasing functions with constant limit $\gamma \geq 0$, and $|g(n) - f(n)| \leq \frac{M}{n^\alpha}$, $M, \alpha > 0$. Let $\pi_1 = \frac{f(n)}{1+f(n)-g(n)}$ be the stationary probability of an edge being present. Let T_{Ind} be the completion time of some protocol spreading information in a sequence of independent Erdös-Rényi graphs with parameter π_1. Suppose that T_{Ind} is $O(r(n))$ with high probability. If for sufficiently large s,

$$\exp\left(-(1 - \frac{1}{s})^2 \frac{sr(n)}{2}\right),$$

decreases to 0 polynomially as a function of n, then the completion time of the protocol over G is also $O(r(n))$ with high probability.

The proof of Theorem 2.5 is presented in Section 3.1. Theorem 2.5 shows that under some sufficient conditions on the edge transition matrix, if a protocol spreads fast enough in a sequence of independent random graphs, we can give upper bounds for the completion time on a sequence of edge markovian graphs. As an example, consider the Push Protocol spreading through G such that $\pi(1) > \frac{1}{n}$. Theorem 2.1 tells us that the completion protocol is $O(\log n)$ with high probability. We have that:

$$\exp\left(-(1-\frac{1}{s})^2 \frac{s \log n}{2}\right) = n^{-(1-\frac{1}{s})^2 \frac{s}{2}}. \tag{2.6}$$

For sufficiently large s, Equation 2.6 decreases polynomially to 0 (the exponent is positive), and that is a sufficient condition to state that the Push Protocol takes $O(\log n)$ steps to complete transmission in the Markov dynamic.

We now state several corollaries, bounding the completion time for some information spreading protocols. The proof for these corollaries follow directly from Theorems 2.1 to 2.4 and Theorem 2.5.

Corollary 2.7 *Let $\pi_1 = \frac{f(n)}{1+f(n)-g(n)}$ be the stationary probability of an edge being present. The completion time of the Push Protocol over G is $O(\frac{\log n}{\min\{n\pi_1,1\}})$.*

We note that the Push Protocol takes at least $\lceil \log_2 n \rceil$ steps to complete transmission. So when $\min\{n\pi_1, 1\} = 1$, the completion time is $\Theta(\log n)$. The following Corollary 2.8 highlights a special case of the push protocol over a sequence of edge markovian graphs where $p = \Theta(\frac{1}{n^k})$, $k > 1$. As far as we know, this case was not considered in the literature.

Corollary 2.8 *Let $G = (G_t)_{t\geq 0} = ([n], E_t)$ be a sequence of edge markovian random graphs with $f(n) = \frac{a}{n^k}$ and $g(n) = 0$, in which $a > 0$ and $k > 1$ are constants. The completion time of the Push Protocol over G is $O(n^{k-1} \log n)$*

Corollary 2.9 *Let $\pi_1 = \frac{f(n)}{1+f(n)-g(n)}$ be the stationary probability of an edge being present. The completion time of the Flood Protocol over G is $O\left(\frac{\log n}{\log(1+n\pi_1)}\right)$ with high probability.*

Corollary 2.10 *Let $\pi_1 = \frac{f(n)}{1+f(n)-g(n)}$ be the stationary probability of an edge being present. If $\pi_1 = \frac{a}{n}$, for a positive constant a. Then, the completion time of the Pull Protocol over G is $O(\log n)$ with high probability.*

Corollary 2.11 *Let $\pi_1 = \frac{f(n)}{1+f(n)-g(n)}$ be the stationary probability of an edge being present. If $\pi_1 = \frac{a}{n}$, for a positive constant a. Then, the completion time of the Push-Pull Protocol over G is $O(\log n)$*

2.3 Rumour Spreading in Dynamic Random Graphs with Communities

In this section we are interested in studying rumour spreading in a sequence of independent random graphs with nodes divided into two communities. We start by introducing new types of nodes: skeptical nodes, bots and stiflers. We then explore rumour spreading in a sequence of independent stochastic block model random graphs.

2.3.1 Rumour Spreading in Dynamic Random Graphs with Skeptical Nodes

In this Section we present a variation of the model shown in Section 2.1. We will only consider the rumour spreading through the Push Protocol. Nodes can now be: informed, uninformed or skeptical. There are αn skeptical nodes, $0 < \alpha < 1$. Skeptical nodes cannot become informed. Let S be the set of skeptical nodes. Let $u, v \in [n] \setminus S$ and let $w \in S$. The edge $\{u, v\}$ has probability p_1 of being present at each time step, independently of everything else. The edge $\{u, w\}$ has probability p_2 of being present at each time step, independently of everything else. We start with one informed node. At each time step we generate a new random graph. The informed nodes, then, choose one of their neighbours uniformly at random. If the chosen neighbour was uninformed, it becomes informed. We bound the time until every uninformed node has become informed. Our proofs use the same argument as CLEMENTI *et al.* (2016). The following Theorem 2.12 gives an upper bound for the completion time, it is proved in Section 3.2.

Theorem 2.12 *Let $(G_t)_{t\geq 1}$, $G_t = ([n], E_t)$, be a sequence of random graphs with edge probability between a skeptical and a non-skeptical node $p_2 = O(p_1)$:*
(a) if $p_1 \geq \frac{1}{(1-\alpha)n}$ the completion time of the Push Protocol over (G_t) is $O(\log n)$ with high probability.
(b) if $p_1 < \frac{1}{(1-\alpha)n}$ the completion time of the Push Protocol over (G_t) is $O(\frac{\log n}{np_1})$ with high probability.

In part b) of Theorem 2.12, if $p_1 = \frac{c}{n}$, $0 < c < \frac{1}{1-\alpha}$ the completion time remains $O(\log n)$, however, when $p_1 = \frac{c}{n^k}$, $c > 0$, $k > 1$, the completion time is $O(n^{k-1} \log n)$.

2.3.2 Rumour Spreading in Dynamic Random Graphs with Bots and Stiflers

In this Section we consider a rumour spreading according to the Push Protocol in a sequence of random graphs. The nodes are divided into two groups: Bots and Stiflers. We have αn Bots, $0 < \alpha < 1$, and βn Stiflers, $\beta = 1 - \alpha$. Stiflers can become informed, but do not spread the rumour further. We start with every Bot informed. At each time step an edge between two Bots appears with probability p_1, independently of everything else. An edge between a Bot and a Stifler appears with probability p_2, independently of everything else. At each time step we generate a new graph. Bots choose one of their neighbours uniformly at random. If the chosen neighbour was a Stifler, it becomes informed. We bound the time until every Stifler node becomes informed. Our proofs use the same argument as CLEMENTI *et al.* (2016). The following Theorem whose proof is presented in Section 3.3 provides an upper bound for the completion time.

Theorem 2.13 *Let $(G_t)_{t\geq 1}$, $G_t = ([n], E_t)$, be a sequence of random graphs with edge probability between two bots $p_1 = O(p_2)$:*
(a) if $p_2 \geq \frac{1}{\beta n}$ the completion time of the Push Protocol over (G_t) is $O(\log n)$ with high probability.
(b) if $p_2 < \frac{1}{\beta n}$ the completion time of the Push Protocol over (G_t) is $O(\frac{\log n}{np_2})$ with high probability.

In part b) of Theorem 2.13, if $p_2 = \frac{c}{n}$, $0 < c < \frac{1}{1-\beta}$ the completion time remains $O(\log n)$, however, when $p_2 = \frac{c}{n^k}$, $c > 0$, $k > 1$, the completion time is $O(n^{k-1} \log n)$.

2.3.3 Rumour Spreading in the Dynamic Stochastic Block Model

In this Section we consider a rumour spreading according to the Push Protocol in a sequence of stochastic block model random graphs. The nodes are divided into two communities: 1 and 2. Let C_i be the set of nodes in community $i = 1, 2$. We have that $|C_1| = \alpha n$ and $|C_2| = \beta n = (1 - \alpha)n$. A pair of nodes from community i have probability p_i of being connected, $i = 1, 2$. A node from community i has probability $p_{ij} = p_{ji}$ of being connected to a node from community j, $i \neq j$, $i, j = 1, 2$. We start with a single informed node in community 1. At each time step we sample a new random graph. Informed nodes choose one of their neighbours uniformly at random. If the chosen neighbour wasn't informed, it becomes informed. We bound the time until every node becomes informed. To derive the upper bound we will present variations of this model that are equivalent to the models discussed in Sections 2.3.1 and 2.3.2.

First we will consider the case where p_1 and p_{12} are of the same order of magnitude and p_2 is at least of the same order.

Theorem 2.14 *Let $G = (G_t)_{t\geq 1}$, $G_t = ([n], E_t)$, be a sequence of stochastic block model random graphs with $p_1 = \Theta(p_{12})$ and $p_2 = \Omega(p_{12})$. The completion time of the Push Protocol over G is $O\left(\frac{\log n}{\min\{1, np_{12}\}}\right)$ with high probability.*

We can also compute an upper bound for the special case where nodes in C_2 find it difficult to form connections between themselves and the nodes in C_1 are comparatively highly connected among themselves and among the nodes in C_2.

Theorem 2.15 *Let $G = (G_t)_{t\geq 1}$, $G_t = ([n], E_t)$, be a sequence of stochastic block model random graphs with $p_{12} = \Theta(p_1)$ and $p_2 = o(p_1)$. The completion time of the Push Protocol over G is $O\left(\frac{\log n}{\min\{1, np_1\}}\right)$ with high probability.*

The proofs of both results are presented in Section 3.4.

Chapter 3

Proofs of the New Results

3.1 Proof of the Theorem on Rumour Spreading in Edge Markovian Random Graphs

As seen in section 2.2, under the hypothesis that $(E_t)_{t\geq 0}$ is an ergodic Markov chain, Proposition B.7 establishes that there is an optimal strong stationary time for this chain. When a strong stationary time happens, each edge is present with probability $\pi(1) = \frac{p}{p+q}$. If we let the rumour spread only at the strong stationary times, we go back to the time independent model seen in Section 2.1 and have plenty of results to work with. Since we limit the rumour spreading to strong stationary times, the completion time is at least as large as the completion time on the original model, where new nodes can be informed at every instant. Upper bounds for the completion time on the limited model are also valid for the original model.

Lemmas 3.1 e 3.2, below, show how to bound the separation distance for $(E_t)_{t\geq 0}$ and how to bound the completion time, respectively. In what follows, let $(G_t)_{t\geq 0}$, $G_t = ([n], E_t)$, be a sequence of edge markovian graphs with transition matrix:

$$P = \begin{pmatrix} 1 - f(n) & f(n) \\ 1 - g(n) & g(n) \end{pmatrix}.$$

The next Lemma 3.1 uses the Definition of separation distance that can be seen in B.4.

Lemma 3.1 *Let $S_x(t)$ be the separation distance for $(E_t)_{t\geq 0}$, with initial state $x \in \{0,1\}^{\binom{n}{2}}$. Also define $h(n) := \max\left\{\frac{f(n)}{1-g(n)}, \frac{1-g(n)}{f(n)}\right\}$ Then:*

$$S_x(t) \leq n^2 h(n) \left(|g(n) - f(n)|\right)^t .$$

Proof. We have that for a Markov chain with transition matrix:

$$A = \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix},$$

the t-step transition probabilities are:

$$A^t = \begin{pmatrix} \frac{q+(1-p-q)^t p}{p+q} & \frac{p-(1-p-q)^t p}{p+q} \\ \frac{q-(1-p-q)^t q}{p+q} & \frac{p+(1-p-q)^t q}{p+q} \end{pmatrix}.$$

We have $p = f(n)$ and $q = 1 - g(n)$. So we can compute the separation distance for a single edge:

$$S_0(t) = \max_{y\in\{0,1\}} \left(1 - \frac{P^t_{0,y}}{\pi(y)}\right) = \max\left\{1-1-(1-f(n)-1+g(n))^t \frac{f(n)}{1-g(n)}, 1-1+(1-f(n)+g(n))^t\right\},$$

$$S_1(t) = \max_{y\in\{0,1\}} \left(1 - \frac{P^t_{1,y}}{\pi(y)}\right) = \max\left\{1-1+(1-f(n)-1+g(n))^t, 1-1+(1-f(n)-1+g(n))^t \frac{1-g(n)}{f(n)}\right\}.$$

The separation distance satisfies:

$$S(t) := \max\{S_0(t), S_1(t)\} \leq \left(|g(n) - f(n)|\right)^t h(n).$$

We now compute the separation distance for the edge set $(E_t)_{t\geq 0}$. For every initial state $x \in \{0,1\}^{\binom{n}{2}}$:

$$S_x(t) = 1 - \min_{y\in S} \frac{M^t_{x,y}}{\Pi(y)} = 1 - \min_{y\in S} \frac{P^t_{x_1,y_1}}{\pi(y_1)} \frac{P^t_{x_2,y_2}}{\pi(y_2)} \cdots \frac{P^t_{x_{\binom{n}{2}},y_{\binom{n}{2}}}}{\pi(y_{\binom{n}{2}})} = 1 - \Pi_{i=1}^{\binom{n}{2}}(1 - S_{x_i}(t)) \leq$$

$$1 - \left(1 - \left(|g(n) - f(n)|\right)^t h(n)\right)^{\binom{n}{2}} \leq n^2 h(n) \left(|g(n) - f(n)|\right)^t .$$

■

Lemma 3.2 *Let T_{Mar} be the completion time of an information spreading protocol over $(G_t)_{t\geq 0}$. Suppose that $f(n)$, $g(n)$ are decreasing functions with limit $\gamma \geq 0$ and that $|g(n) - f(n)| \leq \frac{M}{n^\alpha}$, $M, \alpha > 0$. Then for sufficiently large C, D, s:*

$$\mathbb{P}(T_{\mathit{Mar}} > C\ r(n)) \leq \exp\left(-(1-\frac{1}{s})^2 \frac{sr(n)}{2}\right) +$$

$$\mathbb{P}(T_{\mathit{Ind}} > D\ r(n)),$$

in which $r(n)$ is an arbitrary function and T_{Ind} is the completion time of the same protocol on a sequence of independent Erdös-Rényi random graphs with edge parameter $\frac{f(n)}{1+f(n)-g(n)}$.

Proof. Let τ be the optimal strong stationary time for $(E_t)_{t>0}$. We can use Lemma 3.1 and Proposition B.7 to bound the distribution of τ. $\exists\, k > 0$, such that for any initial state and for sufficiently large n:

$$\mathbb{P}(\tau > t) \leq S(t) \leq n^2 h(n)(|g(n) - f(n)|)^t \leq n^2 \left(\frac{M}{n^\alpha}\right)^t \leq \left(\frac{1}{n^k}\right)^{t-l},$$

where $l = \lfloor \frac{2}{k} \rfloor$. Thus, we can establish a stochastic inequality between τ and Y, geometric random variable with parameter $1 - \frac{1}{n^k}$. We have that $Y + l > \tau$. Let $(\tau_i)_{i\geq 1}$ be a sequence of optimal strong stationary times for (E_t) and let $(Y_i)_{i\geq 1}$ be a sequence of iid random variables with the same distribution as Y. If we let the rumour spread only at the sequence $(\tau_i)_{i\geq 1}$, we have that $T_{\text{Mar}} \leq \sum_{i=1}^{T_{\text{Ind}}} \tau_i$. For $C, D > 0$ and an function $r(n)$, we compute:

$$\mathbb{P}\left(T_{\text{Mar}} > C\ r(n)\right) \leq \mathbb{P}\left(\sum_{i=1}^{T_{\text{Ind}}} \tau_i > C\ r(n)\right) \leq \mathbb{P}\left(\sum_{i=1}^{T_{\text{Ind}}} (Y_i + l) > C\ r(n)\right),$$

the last inequality is due to the stochastic domination. Then,

$$\mathbb{P}\left(\sum_{i=1}^{T_{\text{Ind}}} (Y_i + l) > C\ r(n)\right) =$$

$$\mathbb{P}\left(\sum_{i=1}^{T_{\text{Ind}}} (Y_i + l) > C\ r(n), T_{\text{Ind}} > D\ r(n)\right) + \mathbb{P}\left(\sum_{i=1}^{T_{\text{Ind}}} (Y_i + l) > C\ r(n), T_{\text{Ind}} \leq D\ r(n)\right) \leq$$

$$\mathbb{P}\left(T_{\mathsf{Ind}} > D\, r(n)\right) + \mathbb{P}\left(\sum_{i=1}^{D\, r(n)} (Y_i + l) > C\, r(n)\right). \tag{3.3}$$

For the second term of Equation (3.3), we use Proposition A.3. Since $\sum_{i=1}^{D\, r(n)}(Y_i - 1)$ has distribution Neg Bin$(Dr(n), 1 - \frac{1}{n^k})$, then for $s > 1$:

$$\mathbb{P}\left(\sum_{i=1}^{D\, r(n)} (Y_i - 1) > sDr(n)(1 - \frac{1}{n^k}\right) \le \text{exp}\left(-(1 - \frac{1}{s})^2 \frac{sr(n)}{2}\right).$$

Now observe that,

$$\mathbb{P}\left(\sum_{i=1}^{D\, r(n)} (Y_i + l) > C\, r(n)\right) = \mathbb{P}\left(\sum_{i=1}^{D\, r(n)} (Y_i - 1) > (C - D - Dl)r(n)\right) =$$

$$\mathbb{P}\left(\sum_{i=1}^{D\, r(n)} (Y_i - 1) > (\frac{C}{D} - 1 - l)Dr(n)\right) \le \mathbb{P}\left(\sum_{i=1}^{D\, r(n)} (Y_i - 1) > (\frac{C}{D} - 1 - l)Dr(n)(1 - \frac{1}{n^k})\right).$$

So if $\frac{C}{D} - 1 - l =: s > 1$, we get:

$$\mathbb{P}\left(\sum_{i=1}^{D\, r(n)} (Y_i + l) > C\, r(n)\right) \le \text{exp}\left(-(1 - \frac{1}{s})^2 \frac{sr(n)}{2}\right). \tag{3.4}$$

From (3.3) and (3.4), we conclude that:

$$\mathbb{P}(T_{\mathsf{Mar}} > C\, r(n)) \le \text{exp}\left(-(1 - \frac{1}{s})^2 \frac{sr(n)}{2}\right) +$$

$$\mathbb{P}(T_{\mathsf{Ind}} > D\, r(n))$$

■

We can now conclude the proof of Theorem 2.5.

Proof of Theorem 2.5. When T_{ind} is $O(r(n))$ with high probability and

$$\text{exp}\left(-(1 - \frac{1}{s})^2 \frac{sr(n)}{2}\right),$$

vanishes polynomially as a function of n, a direct application of Lemma 3.2 shows that $\mathbb{P}(T_{\text{Mar}} > Cr(n)) \leq \frac{a}{n^\beta}$, $\beta \geq 1$. This implies that the completion time of the same protocol over a sequence of edge markovian random graphs is $O(r(n))$ with high probability. Concluding the proof. ■

3.2 Proof of the Theorem on Rumour Spreading with Skeptical Nodes

To prove the first part of Theorem 2.12 we need a technical Definition and Lemma.

Definition 3.5 ((I, b)-modified graph) *Let G be a graph with vertex set $V = [n] = \{1, 2, \ldots, n\}$ and edges set E. Let $I \subseteq [n]$ be the set of informed nodes. Let $1 \leq b \leq n$ be an arbitrary number and $\{v_1, \ldots v_b\}$ be a set of "virtual" nodes. Then the (I, b)-modified version of G is the graph H with vertex set $V(H) = [n] \cup \{v_1, \ldots v_b\}$ and edges set $E(H)$ obtained from E by the following operations:*

1. *remove all edges between any pair of nodes both in I,*
2. *for every $u \in I$ add all edges $\{u, v_1\}, \{u, v_2\}, \ldots, \{u, v_b\}$.*

This definition is useful to deal with the dependence that arises from two informed vertices being connected.

We can now compute the expected amount of new informed nodes in a single step of the Push Protocol in H. A coupling shows that the Push Protocol in G informs at least as many nodes as in H.

Lemma 3.6 (Increasing rate of informed nodes) *Let $I \subseteq [n]$ be the set of informed nodes and $S \subset [n]$ be the set of skeptical nodes. Let $G = ([n], E)$ be a new graph with edge probability between two non-skeptical nodes $p_1 \geq \frac{1}{(1-\alpha)n}$ and edge probability between a skeptical and a non-skeptical node $p_2 = O(p_1)$. Let X be a random variable counting the number of new informed nodes after a single step of the Push Protocol in G. Then, there exists $0 < \gamma < 1$:*

$$\mathbb{P}(X > \gamma \min\{|I|, n - |I| - |S|\}) \geq \gamma.$$

Proof. Let H be the $(I, 2\alpha np_2 + 2(1-\alpha)np_1)$-modified version of G. $2\alpha np_2 + 2(1-\alpha)np_1$ is two times the expected degree of a non-skeptical node. Let $\text{deg}_G(u, I) = |\{u' \in I : \{u, u'\} \in E\}|$ be the random variable counting the degree of u among informed nodes in G. Let $J = \{u \in I : \text{deg}_G(u, I) \leq 2\alpha np_2 + 2(1-\alpha)np_1\}$ be a random variable indicating the set of informed nodes with less than $2\alpha np_2 + 2(1-\alpha)np_1$ informed neighbours in G. Let $v \in [n] \setminus (S \cup I)$ be an uninformed node. Let $u \in J$. We consider a single step of the push protocol in H and compute the probability of u sending the rumour to v.

Let $\delta_H(u)$ be a random variable indicating the node chosen by u in the single step of the Push Protocol in H. Then:

$$\mathbb{P}(\delta_H(u) = v | J) = p_1 \mathbb{P}(\delta_H(u) = v | J, \{u, v\} \in E).$$

Conditioning on u having at most $3\alpha np_2$ skeptical neighbours and at most $3(1-\alpha)p_1$ uninformed nodes we get that:

$$\mathbb{P}(\delta_H(u) = v | J, \{u, v\} \in E) \geq$$

$$\mathbb{P}(\delta_H(u) = v | J, \{u, v\} \in E, \text{deg}_G(u, [n] \setminus (I \cup \{v\} \cup S)) \leq 3(1-\alpha)np_1 - 1, \text{deg}_G(u, S) \leq 3\alpha p_2)$$

$$\mathbb{P}(\text{deg}_G(u, [n] \setminus (I \cup \{v\} \cup S)) \leq 3(1-\alpha)np_1 - 1) \mathbb{P}(\text{deg}_G(u, S) \leq 3\alpha p_2).$$

Now u chooses one of its neighbours uniformly at random, so the first term satisfies:

$$\mathbb{P}(\delta_H(u) = v | J, \{u, v\} \in E, \text{deg}_G(u, [n] \setminus (I \cup \{v\} \cup S)) \leq 3(1-\alpha)np_1 - 1, \text{deg}_G(u, S) \leq 3\alpha p_2) \geq$$

$$\frac{1}{5\left((1-\alpha)np_1 + \alpha np_2\right)}.$$

We can bound the second term using the fact that $p_1 \geq \frac{1}{(1-\alpha)n}$ and the Markov inequality:

$$\mathbb{P}(\text{deg}_G(u, [n] \setminus (I \cup \{v\} \cup S)) \leq 3(1-\alpha)np_1 - 1) = 1 - \mathbb{P}(\text{deg}_G(u, [n] \setminus (I \cup \{v\} \cup S)) > 3(1-\alpha)np_1 - 1) \geq$$

$$1 - \mathbb{P}(\text{deg}_G(u, [n] \setminus (I \cup \{v\} \cup S)) > 2(1-\alpha)np_1) \geq 1 - \frac{\left((1-\alpha)n - |I| - 1\right)p_1}{2(1-\alpha)p_1} \geq 1 - \frac{(1-\alpha)np_1}{2(1-\alpha)np_1} = \frac{1}{2}.$$

We can also bound the third term using the Markov inequality:

$$\mathbb{P}(\text{deg}_G(u, S) \leq 3\alpha p_2) = 1 - \mathbb{P}(\text{deg}_G(u, S) > 3\alpha p_2) \geq 1 - \frac{\alpha np_2}{3\alpha np_2} = \frac{2}{3}.$$

As a result:

$$\mathbb{P}(\delta_H(u) = v \mid J) \geq \frac{1}{5\left((1-\alpha)np_1 + \alpha np_2\right)} \frac{1}{2}\frac{2}{3} p_1 \geq \frac{1}{3}\frac{1}{5\left((1-\alpha)n + \beta n\right)} = \frac{\lambda}{n},$$

where constant β is an upper bound on the ratio between p_1 and p_2, and $\lambda = \frac{1}{15[(1-\alpha)n+\beta n]}$.

We can now compute the probability of v not receiving the information from any of the nodes in J.

$$\mathbb{P}(\cap_{u\in J}\{\delta_H(u) \neq v\}|J) \leq \left(1 - \frac{\lambda}{n}\right)^{|J|} \leq \exp\left(-\frac{\lambda|J|}{n}\right).$$

Let Y be a random variable that counts the number of nodes in $[n] \setminus (I \cup S)$ that receive the rumour in H from a vertex in J. Then:

$$\mathbb{E}(Y|J) \geq (n - |I| - |S|)\left(1 - \exp\left(-\frac{\lambda|J|}{n}\right)\right) \geq ((1-\alpha)n - |I|)\,\frac{\lambda}{2}\frac{|J|}{n}.$$

To compute the expectation of Y we first analyze the probability of a node being a member of J and the expectation of J.

$$\mathbb{P}(u \in J) = \mathbb{P}(\text{deg}_G(u, I) \leq 2\alpha np_2 + 2(1-\alpha)np_1) =$$

$$1 - \mathbb{P}(\text{deg}_G(u, I) > 2\alpha np_2 + 2(1-\alpha)np_1) \geq \frac{1}{2}.$$

Since the probability of an informed node being part of J is at least $\frac{1}{2}$, we have that $\mathbb{E}(J) \geq \frac{|I|}{2}$. Using the Law of Total Expectation, we obtain that the unconditional expectation of Y satisfies:

$$\mathbb{E}(Y) \geq \frac{\lambda}{2n}\left((1-\alpha)n - |I|\right)\frac{|I|}{2}.$$

We have 2 cases. If $|I| \geq \frac{(1-\alpha)n}{2}$:

$$\mathbb{E}(Y) \leq \frac{\lambda}{8}(1-\alpha)|I|.$$

If $|I| > \frac{(1-\alpha)n}{2}$:

$$\frac{\lambda}{8}(1-\alpha)\left((1-\alpha)n - |I|\right).$$

We get that:

$$\mathbb{E}(Y) \geq \frac{\lambda}{8}(1-\alpha)\,\text{min}\{|I|, (1-\alpha)n - |I|\}.$$

Since $Y \leq$ min$\{|I|, (1-\alpha)n - |I|\}$, Proposition A.2 implies that:

$$\mathbb{P}(Y \geq \frac{\lambda(1-\alpha)}{16} \text{min}\{|I|, (1-\alpha)n - |I|\}) \geq \frac{\lambda(1-\alpha)}{16}.$$

Let X be a random variable that counts the number of new nodes that receive the rumour in G after a single step of the Push Protocol. We will show through a coupling that X is stochastically at least as large as Y. Let $u \in J$, consider the following coupling:

- If $\delta_H(u) \in [n] \setminus |I|$ then $\delta_G(u) = \delta_H(u)$.
- Let k, l, h be the number of uninformed, skeptical and informed neighbours of u in G respectively. If $\delta_H(u)$ is one of the virtual nodes, then let $\delta_G(u)$ be uniform among the uninformed neighbours with probability $\frac{k}{h+k+l} \frac{2\alpha np_2 + 2(1-\alpha)np_1 - h}{2\alpha np_2 + 2(1-\alpha)np_1}$. Let $\delta_G(u)$ be uniform among the skeptical neighbours with probability $\frac{l}{h+k+l} \frac{2\alpha np_2 + 2(1-\alpha)np_1 - h}{2\alpha np_2 + 2(1-\alpha)np_1}$. And let $\delta_G(u)$ be uniform among the informed neighbours with probability

$$\frac{h}{h+k+l} \frac{2\alpha np_2 + 2(1-\alpha)np_1 + k + l}{2\alpha np_2 + 2(1-\alpha)np_1}$$

.

Informed nodes in G that are not in J, meaning they have degree between informed nodes greater than $2\alpha np_2 + 2(1-\alpha)np_1$, perform the Push Protocol in G independently of the Push Protocol in H.

Every time a node in J informs a new node in H, it also informs a new node in G. We conclude that:

$$\mathbb{P}(X \geq \frac{\lambda(1-\alpha)}{16} \text{min}\{|I|, (1-\alpha)n - |I|\}) \geq \frac{\lambda(1-\alpha)}{16}.$$

■

We can now prove the first part of the Theorem, the case where $p_1 \geq \frac{1}{(1-\alpha)n}$

Proof of part (a) from Theorem 2.12. Consider a step t of the Push Protocol, let I_t be the set of informed nodes at time t and let $m_t = |I_t|$ be its size.If $m_t < \frac{(1-\alpha)n}{2}$, Lemma 3.6 implies that $\mathbb{P}(m_{t+1} \geq (1+\gamma)m_t) \geq \gamma$. Let $\varepsilon_t = \{m_t \geq (1+\gamma)m_{t-1}\} \cup \{m_{t-1} \geq \frac{n}{2}\}$ and let $Y_t = Y_t((E_1, I_1), \ldots, (E_t, I_t))$ be its indicator variable. For $t = \frac{\log((1-\alpha)n)}{\log(1+\gamma)}$, we have that: $(1+\gamma)^t \geq \frac{(1-\alpha)n}{2}$. Let $T_1 = \frac{2}{\gamma} \frac{\log((1-\alpha)n)}{\log(1+\gamma)}$, we compute the probability of taking longer than T_1 steps to reach at least $\frac{(1-\alpha)n}{2}$ informed nodes.

$$\mathbb{P}(m_{T_1} \leq \frac{(1-\alpha)n}{2}) \leq \mathbb{P}(\sum_{t=1}^{T_1} Y_t \leq \frac{\gamma}{2} T_1).$$

This probability is not larger than the probability of a binomial random variable with parameters T_1 and γ being less than $\frac{\gamma}{2}T_1$. A direct application of the Chernoff Bound (Proposition A.1) shows that:

$$\mathbb{P}(\sum_{t=1}^{T_1} Y_t \leq \frac{\gamma}{2} T_1) \leq \text{exp}(-\frac{\gamma}{4} T_1) = ((1-\alpha)n)^{\eta_1},$$

$\eta_1 > 0$. We have shown that after T_1 time steps, we reach at least $\frac{(1-\alpha)n}{2}$ informed nodes with high probability.

If $m_{T_1} \geq \frac{(1-\alpha)n}{2}$, Lemma 3.6 implies that, for $t > T_1$, $\mathbb{P}((1-\alpha)n - m_{t+1} \leq (1-\gamma)((1-\alpha)n - m_t)) \geq \gamma$. Let $t = \frac{\log((1-\alpha)n}{\gamma}$, then $(1-\gamma)^t \leq 1$. For $T_2 = \frac{2}{\lambda}\frac{\log((1-\alpha)n}{\gamma} + T_1$, the probability that the Push Protocol has not completed transmission by T_2 is:

$$\mathbb{P}(m_{T_2} < (1-\alpha)n) \leq \mathbb{P}(m_{T_2} < (1-\alpha)n | m_{T_1} \geq \frac{(1-\alpha)n}{2}) + \mathbb{P}(m_{T_1} < \frac{(1-\alpha)n}{2}).$$

Applying the same logic used in the first regime (until $\frac{(1-\alpha)n}{2}$ informed nodes), the first probability is not larger than the probability of a binomial random variable with parameters T_2 and γ being less than $\frac{\gamma}{2}T_2$. A direct application of the Chernoff Bound shows that:

$$\mathbb{P}(m_{T_2} < (1-\alpha)n | m_{T_1} \geq \frac{(1-\alpha)n}{2}) \leq \text{exp}(-\frac{1}{4}\gamma T_2) = ((1-\alpha)n)^{-\eta_2}.$$

After T_2 time steps, the protocol has completed transmission with high probability. This concludes the proof. ■

We state the following Lemma 3.7 without proof. It shows that a single Push on the union of sequence of graphs informs stochastically less nodes then Push operations performed on every single graph.

Lemma 3.7 *Let $G = \{G_t = ([n], E_t)\}_{t=1,\ldots,T}$ be a finite sequence of graphs, let $I \subseteq [n]$ be the subset of informed nodes, and let X be a random variable counting the number of informed nodes at time T. Let $H = ([n], F)$, $F = \cup_{i=1}^{T} E_t$ and let Y be a random variable counting the number of informed nodes after a single Push operation over H.*

Then, X is stochastically larger than Y, that is:

$$\mathbb{P}(X \leq l) \leq \mathbb{P}(Y \leq l).$$

We can now offer a proof of the upper bound for the completion time when the edge probability between two non-skeptical nodes is less than $\frac{1}{(1-\alpha)n}$.

Proof of part (b) of Theorem 2.12. Consider the sequence of random graphs $H = \{H_s = ([n], F_s) : s \in \mathbb{N}\}$, $F_s = E_{sT} \cup E_{sT+1} \cup \cdots \cup E_{sT+T-1}$, where $T = \frac{2}{(1-\alpha)np_1}$. The probability of an edge between two non-skeptical nodes being absent in H is $1 - p_1^H = (1-p_1)^T \leq \exp(-p_1 T) = \exp(-\frac{2}{(1-\alpha)n})$. Therefore, the probability that an edge between two non-skeptical nodes exists is $p_1^H = 1 - \exp(-\frac{2}{(1-\alpha)n}) \geq \frac{1}{(1-\alpha)n}$. p_1^H satisfies the conditions of part (a) of Theorem 2.12, we now have to check p_2^H.The probability of an edge between a skeptical node and a non-skeptical node being present in H is $p_2^H = 1-(1-p_2)^T \leq 1-(1-p_2)^{\lceil T} \leq 1-(1-\lceil T \rceil p_2) = \lceil T \rceil p_2 = \frac{p_2}{p_1}\frac{2}{(1-\alpha)n}$. So we have that:

$$\frac{p_2^H}{p_1^H} \leq \frac{\frac{p_2}{p_1}\frac{2}{(1-\alpha)n}}{\frac{1}{(1-\alpha)n}} \leq \frac{2p_2}{p_1}. \tag{3.8}$$

The last Equation (3.8) tells us that part (a) of Theorem 2.12 holds for H. Let τ_G and τ_H be the completion time of the Push Protocol over G and over H respectively. From part (a) of Theorem 2.12, we know that τ_H is $O(\log n)$ with high probability. From Lemma 3.7 it holds that:

$$\mathbb{P}(\tau_G \geq Tt) \leq \mathbb{P}(\tau_H \geq t).$$

Taking $t = D \log n$, for some sufficiently large constant D, we have that:

$$\mathbb{P}(\tau_G \geq TD \log n) \leq \mathbb{P}(\tau_H \geq D \log n) \leq \frac{1}{n^\eta},$$

where $\eta > 0$. Since $T = \frac{2}{(1-\alpha)np_1}$, this implies that τ_G is $O(\frac{\log n}{np_1})$ with high probability. ■

3.3 Proof of the Theorem on Rumour Spreading with Bots and Stiflers

To prove part (a) of Theorem 2.13 we start by computing the increasing rate of informed Stifler nodes. The proof of the following Lemma 3.9 will use the modified graph defined in 3.5.

Lemma 3.9 *Let $B \subseteq [n]$ be the set of Bots and $I \subset [n]$ be the set of Informed Stiflers. Let $G = ([n], E)$ be a new graph with edge probability between a Bot and a Stifler $p_2 \geq \frac{1}{\beta n}$ and edge probability between two Bots $p_1 = O(p_2)$. Let X be a random variable counting the number of new Stiflers informed after a single step of the Push Protocol in G. Then, there exists a positive constant γ:*

$$\mathbb{P}(X \geq \gamma \min\{\alpha n, \beta n - |I|\}) \geq \gamma.$$

Proof. Let H be the $(B, 2\alpha np_1 + 2\beta np_2)$-modified version of G. $2\alpha np_1 + 2\beta np_2$ is two times the expected degree of a bot. Let $\text{deg}_G(u, B) = |\{u' \in B : \{u, u'\} \in E\}|$ be the random variable counting the degree of u among bots in G. Let $J = \{u \in B : \text{deg}_G(u, B) \leq 2\alpha np_1 + 2\beta np_2\}$ be a random variable indicating the set of bots with less than $2\alpha np_1 + 2\beta np_2$ bot neighbours in G. Let $v \in [n] \setminus B$ be an uninformed stifler. Let $u \in J$. We consider a single step of the push protocol in H and compute the probability of u sending the rumour to v.

Let $\delta_H(u)$ be a random variable indicating the node chosen by u in the single step of the Push Protocol in H. Then:

$$\mathbb{P}(\delta_H(u) = v|J) = p_2 \mathbb{P}(\delta_H(u) = v|J, \{u, v\} \in E).$$

Conditioning on u having at most $3\beta np_2$ stifler neighbours we get that:

$$\mathbb{P}(\delta_H(u) = v|J, \{u, v\} \in E) \geq$$

$$\mathbb{P}(\delta_H(u) = v|J, \{u, v\} \in E, \text{deg}_G(u, [n] \setminus (B \cup \{v\} \cup I)) \leq 3\beta np_2 - 1)$$

$$\mathbb{P}(\text{deg}_G(u, [n] \setminus (B \cup \{v\} \cup I)) \leq 3\beta np_2 - 1).$$

Since u chooses one of its neighbours uniformly at random, for the first term we have that:

$$\mathbb{P}(\delta_H(u) = v|J, \{u, v\} \in E, \text{deg}_G(u, [n] \setminus (B \cup \{v\} \cup I)) \leq 3\beta np_2 - 1) \geq$$

$$\frac{1}{(5\beta np_2 + 2\alpha np_1)}.$$

We can bound the second term using the fact that $p_2 \geq \frac{1}{\beta n}$ and the Markov inequality:

$$\mathbb{P}(\text{deg}_G(u,[n]\setminus(B\cup\{v\}\cup I)) \leq 3\beta np_2-1) = 1-\mathbb{P}(\text{deg}_G(u,[n]\setminus(B\cup\{v\}\cup I)) > 3\beta np_2-1) \geq$$

$$1-\mathbb{P}(\text{deg}_G(u,[n]\setminus(B\cup\{v\}\cup I)) > 2\beta np_2) \geq 1-\frac{(n-|B|-1-|I|)\,p_2}{2\beta p_2} \geq 1-\frac{\beta np_2}{2\beta np_2} = \frac{1}{2}.$$

As a result:

$$\mathbb{P}(\delta_H(u)=v|J) \geq \frac{1}{(5\beta np_2+2\alpha np_1)}\frac{1}{2}p_2 \geq \frac{\lambda}{n}.$$

where λ is a suitable constant.

We can now compute the probability of v not receiving the information from any of the nodes in J.

$$\mathbb{P}(\cap_{u\in J}\{\delta_H(u)\neq v\}|J) \leq \left(1-\frac{\lambda}{n}\right)^{|J|} \leq \exp\left(-\frac{\lambda|J|}{n}\right).$$

Let Y be a random variable that counts the number of nodes in $[n]\setminus(I\cup B)$ that receive the rumour in H from a vertex in J. Then:

$$\mathbb{E}(Y|J) \geq (n-|I|-|B|)\left(1-\exp\left(-\frac{\lambda|J|}{n}\right)\right) \geq (\beta n-|I|)\,\frac{\lambda}{2}\frac{|J|}{n}.$$

To compute the expectation of Y we first analyze the probability of a node being a member of J and the expectation of J.

$$\mathbb{P}(u\in J) = \mathbb{P}(\text{deg}_G(u,B) \leq 2\alpha np_1+2\beta np_2) =$$

$$1-\mathbb{P}(\text{deg}_G(u,B) > 2\alpha np_1+2\beta np_2) \geq \frac{1}{2}.$$

Since the probability of and informed node being part of J is at least $\frac{1}{2}$, we have that $\mathbb{E}(J) \geq \frac{|B|}{2} = \frac{\alpha n}{2}$. Using the Law of Total Expectation, we obtain that the unconditional expectation of Y satisfies:

$$\mathbb{E}(Y) \geq \frac{\lambda}{2n}(\beta n-|I|)\frac{|B|}{2} \geq \frac{\eta}{4}\min\{\beta n-|I|,\alpha n\}.$$

Since $Y \leq \min\{\beta n-|I|,\alpha n\}$, Proposition A.2 implies that:

$$\mathbb{P}(Y \geq \frac{\eta}{8}\min\{\beta n-|I|,\alpha n\}) \geq \frac{\eta}{8}.$$

Let X be a random variable that counts the number of new nodes that receive the rumour in G after a single step of the Push Protocol. We will show through a coupling that X is stochastically at least as large as Y. Let $u \in J$, consider the following coupling:

- If $\delta_H(u) \in [n] \setminus B$ then $\delta_G(u) = \delta_H(u)$.

- Let k, h, l be the number of uninformed stifler, bot and informed stifler neighbours of u in G respectively. If $\delta_H(u)$ is one of the virtual nodes, then let $\delta_G(u)$ be uniform among the uninformed stifler neighbours with probability $\frac{k}{h+k+l}\frac{2\alpha np_1+2\beta np_2-h}{2\alpha np_1+2\beta np_2}$. Let $\delta_G(u)$ be uniform among the informed stifler neighbours with probability $\frac{l}{h+k+l}\frac{2\alpha np_1+2\beta np_2-h}{2\alpha np_1+2\beta np_2}$. And let $\delta_G(u)$ be uniform among the bot neighbours with probability $\frac{h}{h+k+l}\frac{2\alpha np_1+2\beta np_2+k+l}{2\alpha np_1+2\beta np_2}$.

Bots in G that are not in J, meaning they have degree among bots greater than $2\alpha np_1 + 2\beta np_2$, perform the Push Protocol in G independently of the Push Protocol in H.

Every time a node in J informs a new node in H, it also informs a new node in G. We conclude that:

$$\mathbb{P}(X \geq \frac{\eta}{8}\min\{\beta n - |I|, \alpha n\}) \geq \frac{\eta}{8}.$$

■

Lemma 3.9 tells us that when $\alpha < 0.5$ there are two growth regimes of informed stiflers. We start with fewer bots than uninformed stiflers, so there is a cap on the amount of stiflers we can inform in one step, as the push protocol progresses, the number of bots becomes greater than the number of uninformed stiflers, we could possibly inform every single remaining uninformed node in a single step. However, when $\alpha \geq 0.5$, the number of bots is greater than the number of uninformed stiflers from the beginning, meaning we start in the second regime.

Proof of Part (a) of Theorem 2.13. Consider a step t of the Push Protocol, let I_t be the set of informed stifler nodes at time t and let $m_t = |I_t|$ be its size. We have two cases.

Suppose that $\alpha < 0.5$. For any t such that $m_t \leq (1-2\alpha)n$ Lemma 3.9 implies that $\mathbb{P}(m_{t+1} \geq m_t + \gamma\alpha n) \geq \gamma$. Let $\varepsilon_t = \{m_t \geq (m_{t-1} + \gamma\alpha n\} \cup \{m_{t-1} \geq (1-2\alpha)n\}$ and let$Y_t = Y_t((E_1, I_1), \dots, (E_t, I_t))$ be its indicator variable. Starting at 0 informed Stiflers if we inform $\gamma\alpha n$ nodes each step of the push protocol, we reach $(1-2\alpha)n$ informed ones after $\frac{(1-2\alpha)}{\gamma\alpha}$ time steps. Let $T_1 = \frac{2}{\lambda}\frac{(1-2\alpha)}{\gamma\alpha}$, we compute the probability of taking longer than T_1 steps to reach at least $(1-2\alpha)n$ informed nodes.

$$\mathbb{P}(m_{T_1} \leq (1-2\alpha)n) \leq \mathbb{P}(\sum_{t=1}^{T_1} Y_t \leq \frac{(1-2\alpha)}{\alpha\lambda}) \leq \mathbb{P}(\sum_{t=1}^{T_1} Y_t \leq \frac{\gamma}{2}T_1).$$

This probability is not larger than the probability of a binomial random variable with parameters T_1 and γ being less than $\frac{\gamma}{2}T_1$. A direct application of the Chernoff Bound shows that:

$$\mathbb{P}(\sum_{t=1}^{T_1} Y_t \leq \frac{\gamma}{2}T_1) \leq \exp(-\frac{\gamma}{8}T_1) = (n)^{\eta_1},$$

$\eta_1 > 0$. We have shown that after T_1 time steps, we reach at least $(1-2\alpha)n$ informed nodes with high probability.

If $m_{T_1} \geq (1-2\alpha)n$, Lemma 3.6 implies that, for $t > T_1$, $\mathbb{P}(\beta n - m_{t+1} \leq (1-\gamma)(\beta n - m_t)) \geq \gamma$. Let $t = \frac{\log(2\alpha n)}{\gamma}$, then $(1-\gamma)^t \leq \frac{1}{2\alpha n}$. For $T_2 = \frac{2}{\lambda}\frac{\log(2\alpha n}{\gamma} + T_1$, the probability that the Push Protocol has not completed transmission by T_2 is:

$$\mathbb{P}(m_{T_2} < \beta n) \leq \mathbb{P}(m_{T_2} < \alpha n | m_{T_1} \geq (1-2\alpha)n) + \mathbb{P}\left(m_{T_1} < \frac{(1-\alpha)n}{2}\right).$$

Applying the same logic used in the first regime (until $(1-2\alpha)n$ informed stifler nodes), the first probability is not larger than the probability of a binomial random variable with parameters T_2 and γ being less than $\frac{\gamma}{2}T_2$. A direct application of the Chernoff Bound shows that:

$$\mathbb{P}(m_{T_2} < (1-\alpha)n | m_{T_1} \geq (1-2\alpha)n) \leq \exp(-\frac{1}{4}\gamma T_2) = ((2\alpha)n)^{-\eta_2}.$$

After T_2 time steps, the protocol has completed transmission with high probability. This concludes the proof of the first case.

Suppose that $\alpha > 0.5$. Lemma 3.9 implies that $\mathbb{P}(\beta n - m_{t+1} \leq (1-\gamma)(\beta n - m_t)) \geq \gamma$. Let $t = \frac{\log(2\beta n)}{\gamma}$, then $(1-\gamma)^t \leq \frac{1}{2\beta n}$. For $T = \frac{2}{\lambda}\frac{\log(2\beta n}{\lambda}$, the probability that the push protocol has not completed transmission by T is not larger than the probability of a binomial random variable with parameters T and γ being less than $\frac{\gamma}{2}T$, which we can bound through the Chernoff bound.

$$\mathbb{P}(m_T < \beta n) \leq \exp(-\frac{1}{4}\frac{\log(2\beta n)}{\gamma}) = \frac{1}{(2\beta n)^{\eta}}.$$

After T time steps we have completed transmission with high probability. This concludes the proof of the second case. ■

Proof of Part (b) of Theorem 2.13. Consider the sequence of random graphs $H = \{H_s = ([n], F_s) : s \in \mathbb{N}\}$, $F_s = E_{sT} \cup E_{sT+1} \cup \cdots \cup E_{sT+T-1}$, where $T = \frac{2}{\beta n p_2}$. The probability of an edge between a stifler and a bot being absent in H p_2^H is $(1-p_2)^T \leq \exp(-p_2 T) = \exp(-\frac{2}{\beta n})$. Therefore, the probability that an edge exists is $1 - \exp(-\frac{2}{\beta n}) \geq \frac{1}{\beta n}$. p_2^H satisfies the conditions of part (a) of Theorem 2.13, we now have to check p_1^H. The probability of an edge between two bots being present in H is $p_1^H = 1-(1-p_1)^T \leq 1-(1-p_1)^{\lceil T} \leq 1-(1-\lceil T \rceil p_1) = \lceil T \rceil p_1 = \frac{p_1}{p_2}\frac{2}{(1-\beta)n}$. So we have that:

$$\frac{p_1^H}{p_2^H} \leq \frac{\frac{p_1}{p_2}\frac{2}{(1-\beta)n}}{\frac{1}{(1-\beta)n}} \leq \frac{2p_1}{p_2}. \tag{3.10}$$

The last Equation (3.10) tells us that part (a) of Theorem 2.13 holds for H. Let τ_G and τ_H be the completion time of the Push Protocol over G and over H respectively. From Theorem 2.13 part (a) , we know that τ_H is $O(\log n)$ with high probability. From Lemma 3.7 it holds that:

$$\mathbb{P}(\tau_G \geq Tt) \leq \mathbb{P}(\tau_H \geq t).$$

Taking $t = D\log n$, for some sufficiently large constant D, we have that:

$$\mathbb{P}(\tau_G \geq TD\log n) \leq \mathbb{P}(\tau_H \geq D\log n) \leq \frac{1}{n^\eta},$$

where $\eta > 0$. Since $T = \frac{2}{\beta n p_2}$, this implies that τ_G is $O(\frac{\log n}{np_2})$ with high probability. ■

3.4 Proof of the Theorems on Rumour Spreading in the Stochastic Block Model

Proof of Theorem 2.14. We introduce a slower version of the model and find an upper bound for the completion time. First, suppose that every time a node from community 1 tries to inform a node from community 2, we cancel this step of the process and move to the next one. That is, we are considering community 2 to be skeptical. Since $p_1 = \Theta(p_{12}) \implies p_{12} = O(p_1)$, Theorem 2.12 tells us that after $O(\frac{\log n}{\min\{1, np_1\}})$ time steps, every node in community 1 is informed. We move into the second step of the proof. We no longer consider community 2 to be skeptical. We bound the geometric time until one node in community 2 has been informed by a node in community 1. Let X be a geometric random variable with parameter $q = \mathbb{P}(\cup_{u \in C_1} \delta(u) \in C_2)$. We have that:

$$\begin{aligned} q \geq \mathbb{P}(\delta(u) \in C_2) \geq \\ \mathbb{P}(\delta(u) \in C_2 | \text{deg}(u, C_1) \leq \rho_1 \alpha n p_1, \text{deg}(u, C_2) \leq (1-\rho_2)\beta n p_{12}) \\ \mathbb{P}(\text{deg}(u, C_1) \leq \rho_1 \alpha n p_1)\mathbb{P}(\text{deg}(u, C_2) \leq (1-\rho_2)\beta n p_{12}), \end{aligned} \tag{3.11}$$

where $\rho_1 > 0$ and $0 < \rho_2 < 1$. Since u sends the rumour uniformly at random and $p_1 = \Theta(p_{12})$, the first term of (3.11) is at least:

$$\frac{(1-\rho_2)\beta n p_{12}}{(1-\rho_2)\beta n p_{12} + \rho_1 \alpha n p_1} \geq \eta_1, \tag{3.12}$$

for a sufficiently large constant η_1. We can bound the second term of (3.11) using the Markov inequality.

$$\mathbb{P}(\text{deg}(u, C_1) \leq \rho_1 \alpha n p_1) \geq 1 - \frac{\alpha n p_1}{\rho_1 \alpha n p_1} \geq \eta_2. \tag{3.13}$$

for a sufficiently large constant η_2. We can bound the third term of 3.11 using the Chernoff Bound.

$$\mathbb{P}(\text{deg}(u, C_2) \leq (1-\rho_2)\beta n p_{12}) \geq 1 - \text{exp}\left(-\frac{\rho_2^2}{2}\beta n p_{12}\right). \tag{3.14}$$

From (3.12),(3.13), (3.14) and (3.11), the parameter $q \geq \eta \left(1 - \text{exp}\left(-\frac{\rho_2^2}{2}\beta n p_{12}\right)\right)$, $\eta = \eta_1 \eta_2$. The probability that X takes longer than $D\frac{\log n}{\min\{1, n p_{12}\}}$, for a sufficiently large constant D, is:

$$\mathbb{P}(X \geq D\frac{\log n}{\min\{1, n p_{12}\}}) \leq (1-q)^{D\frac{\log n}{\min\{1, n p_{12}\}}}.$$

If $p_{12} = \Omega(\frac{1}{n})$, then $1 - \text{exp}\left(-\frac{\rho_2^2}{2}\beta n p_{12}\right) \geq 1 - \text{exp}\left(-\frac{\rho_2^2}{2}\beta\right) \geq \lambda$, $\lambda > 0$, meaning that: $(1-q)^{D\log n} \leq \text{exp}(\log(1-\eta\lambda) D \log n) \leq \frac{1}{n^{\log(1-\eta\lambda)D}}$. Since $\log(1-q) < 0$, we take $O(\log n)$ time steps until at least one node of community 2 is informed. If $p_{12} = \theta(\frac{1}{n^k})$, $k > 1$, $q = \eta - \eta\,\text{exp}(\frac{\lambda}{n^{k-1}})$, therefore:

$$(1-q)^{Dn^{k-1}\log n} \leq (1-\eta+\eta\,\text{exp}(-\frac{\lambda}{n^{k-1}}))^{Dn^{k-1}\log n} \leq \text{exp}\left((\eta - \eta\,\text{exp}(-\frac{\lambda}{n^{k-1}}))Dn^{k-1}\log n\right) \leq$$

$$\exp\left(-\eta(\frac{\lambda}{n^{k-1}} - \frac{1}{2}\frac{\lambda^2}{n^{2k-2}})Dn^{k-1}\log n\right) = \exp\left(-\eta D\lambda\log n\right)\exp\left(\eta D\lambda^2\frac{\log n}{n^{k-1}}\right) \leq$$

$$\exp\left(-\eta D\lambda\log n\right).$$

We take $O(n^{k-1}\log n)$ time steps until at least one node of community 2 is informed, with high probability.

So far, we have proved that after $O(\frac{\log n}{\min\{1,np_{12}\}})$, every node in community 1 and at least a node in community 2 are informed. Since nodes in community 1 are all informed and are not allowed to inform nodes in community 2, we can consider community one to be skeptical. Since $p_2 = \Omega(p_{12}) \implies p_{12} = O(p_2)$, Theorem 2.12 tells us that after $O(\frac{\log n}{\min\{1,np_1\}})$ time steps, every node in community 2 is informed.

After $O(\frac{\log n}{\min\{1,np_1\}})$ time steps every node is informed with high probability in the slower version of the model. This bound is also valid for the original version. ■

Proof of Theorem 2.15. We begin by cancelling out the rumour spreading of nodes in C_1 that try to inform a node in C_2. This is equivalent to considering C_2 to be skeptical. Since $p_{12} = \Theta(p_1) \implies p_{12} = O(p_1)$, by Theorem 2.12, after $O\left(\frac{\log n}{\min\{1,np_1\}}\right)$ every node in C_1 is informed with high probability. We now allow nodes in C_1 to inform nodes in C_2 but cancel the rumour spreading of nodes in C_2. This is equivalent to considering C_1 to be bots and C_2 to be stiflers. Since $p_{12} = \Theta(p_1) \implies p_1 = O(p_{12})$, by Theorem 2.13, after $O\left(\frac{\log n}{\min\{1,np_{12}\}}\right)$ time steps, every node in C_2 is informed with high probability. Since $p_{12} = \Theta(p_1)$, the completion time of the Push Protocol over G is $O\left(\frac{\log n}{\min\{1,np_1\}}\right)$ with high probability. ■

Chapter 4

Conclusion

We analyzed rumour spreading in dynamic random graphs in different settings: several algorithms on edge-markovian random graphs, Push Protocol with skeptical nodes, Push Protocol with bots and stiflers and Push Protocol in a dynamic stochastic block model random graph with 2 communities.

In the first setting we proposed a proof method based on strong stationary times that allows us to bound the completion time of the Push, Pull, Flood and Push-Pull protocol in the markovian dynamic using bounds of the completion time for the i.i.d. case. We were able to get results that, as far as we know, have not been considered in the literature. Namely, we proved that the completion time of the Pull Protocol and the completion time of the Push-Pull Protocol are both $O(\log n)$ when the stationary probability of an edge being present is $\pi_1 = \frac{a}{n}$, $a > 0$. We were also able to prove that the completion time of the Push Protocol is $O(n^{k-1}\log n)$, $k > 1$, when the edge transition matrix is of the form:

$$P = \begin{pmatrix} 1-\frac{a}{n^k} & \frac{a}{n^k} \\ 1 & 0 \end{pmatrix}.$$

In the second and third settings we introduced skeptical, stifler and bot nodes. We proved that the completion time of the Push Protocol is $O(\frac{\log n}{np_1})$ in the former and $O(\frac{\log n}{np_2})$ in the latter. Using comparisons with these two variations, we were also able to show that similar bounds hold for the completion time of the Push Protocol in a sequence of stochastic block model random graphs with two communities. We proved that the push protocol completes transmission after $O(\frac{\log n}{\min\{1,np_1\}})$ in two different settings. When the two communities have similar connectivity, that is when p_1, p_2 and p_{12} are of the same order ($p_2, p_{12} = \theta(p_1)$). And when nodes in community one are highly connected amongst themselves and with nodes in community two, but community two nodes are mostly connected to nodes in community one. That is when $p_{12} = \theta(p_1)$ and $p_2 = o(p_1)$.

Bibliography

BROWN, D. G. (2011). How i wasted too long finding a concentration inequality for sums of geometric variables. *Available at https://cs.uwaterloo.ca/ browndg/negbin*.

CLEMENTI, A. *et al.* (2010). Flooding time of edge-markovian evolving graphs. *SIAM journal on discrete mathematics*, **24 - n. 4**, pp. 1694–1712.

CLEMENTI, A. *et al.* (2016). Rumor spreading in random evolving graphs. *Random Structures Algorithms*, **48**, pp. 290–312.

DAKNAMA, R. (2017). Pull and pushpull in random evolving graphs. *arXiv:1801.00316*.

DOERR, B. e KOSTRYGIN, A. (2017). Randomized rumor spreading revisited. Em *Proceedings of the 44th International Colloquium on Automata, Languages, and Programming (ICALP)*, volume 80, página pp. 138:1–138:14.

HOFSTAD, R. V. D. (2017). *Random Graphs and Complex Networks Volume 1*. Cambridge University Press, Cambridge.

LEVIN, D. A. e PERES, Y. (2017). *Markov Chains and Mixing Times*. American Mathematical Society.

Appendix A

Useful Definitions and Inequalities

In this section we present some inequalities and define asymptotic notation that will be useful later.

Proposition A.1 (Chernoff Bound for Binomial Random Variable) *Let X be a binomial random variable with parameters n and p. Let $\mu = \mathbb{E}(X)$. Then for every $0 < \delta < 1$ we have that:*

$$\mathbb{P}(X \leq (1-\delta)\mu) \leq \text{exp}(-\frac{\delta^2}{2}\mu).$$

The proof for Proposition A.1 can be found in HOFSTAD (2017).

Proposition A.2 *Let X be a random variable taking values between 0 and m, $m > 0$. If $\mathbb{E}(X) \geq \lambda m$, $0 \leq \lambda \leq 1$, then:*

$$\mathbb{P}(X \geq \frac{\lambda}{2}m) \geq \frac{\lambda}{2}.$$

Proof. Using the Markov inequality and the hypothesis that $\mathbb{E}(X) \geq \lambda m$:

$$\mathbb{P}(X < \frac{\lambda}{2}m) = \mathbb{P}(-X > \frac{\lambda}{2}m) = \mathbb{P}(m - X > \left(1 - \frac{\lambda}{2}\right) m) \leq \frac{\mathbb{E}(m-X)}{\left(1-\frac{\lambda}{2}\right)m} \leq \frac{1-\lambda}{1-\frac{\lambda}{2}}.$$

We have that:

$$\mathbb{P}(X \geq \frac{\lambda}{2}m) \geq 1 - \frac{1-\lambda}{1-\frac{\lambda}{2}} = \frac{\frac{\lambda}{2}}{1-\frac{\lambda}{2}} \geq \frac{\lambda}{2}.$$

■

Proposition A.3 (Chernoff Bound for Negative Binomial Random Variable) *Let X be a negative binomial random variable with parameters n and $\frac{1}{p}$. Let $k > 1$ be a* *constant. Then we have that:*

$$\mathbb{P}(X > knp) \leq \text{exp}(\frac{-(1-\frac{1}{k})^2}{2}kn).$$

The proof for Proposition A.3 can be seen in BROWN (2011).

Asymptotic notation is useful when dealing with the completion time of a rumour spreading protocol over a dynamic graph.

Definition A.4 (Big-O notation) *Let $f(n), g(n)$ be arbitrary non-negative functions. If* $\lim_{n\to\infty} \frac{f(n)}{g(n)} = l$*,* $0 \leq l < \infty$*, we say that* $f(n) = O(g(n))$*.*

Definition A.5 (Small-o notation) *Let $f(n), g(n)$ be arbitrary non-negative functions. If* $\lim_{n\to\infty} \frac{f(n)}{g(n)} = o$*, we say that* $f(n) = o(g(n))$*.*

Note that if $f(n) = o(g(n))$ then $f(n) = O(g(n))$.

Definition A.6 (Big-Ω notation) *Let $f(n), g(n)$ be arbitrary non-negative functions. If* $\lim_{n\to\infty} \frac{f(n)}{g(n)} = l$*,* $0 < l \leq \infty$*, we say that* $f(n) = \Omega(g(n))$*.*

Definition A.7 (Big-Θ notation) *Let $f(n), g(n)$ be arbitrary non-negative functions. If* $\lim_{n\to\infty} \frac{f(n)}{g(n)} = l$*,* $0 < l < \infty$*, we say that* $f(n) = \Theta(g(n))$*.*

Appendix B

Strong Stationary Times

This section is based on Chapter 6 of LEVIN e PERES (2017), we will present the concept of Strong Stationary Times and a Proposition that guarantees their existence in ergodic Markov chains. We start with some definitions.

Definition B.1 (Random Function Representation) *The random function representation of a Markov chain with states space S and transition matrix P is a functions $f : S \times \Omega \to S$, where Ω is the support of a random variable U, such that $\mathbb{P}(f(x, U) = y) = P_{x,y}$.*

Definition B.2 (Randomized Stopping Time) *τ is a randomized stopping time for $(X_t)_{t\geq 0}$ if it is a stopping time for the sequence of iid random variables $(U_t)_{t\geq 1}$ used in the random function representation of (X_t).*

Definition B.3 (Total Variation Distance) *The total variation distance between two distributions with sample space Ω is:*

$$|\mu - \nu|_{TV} = \sum_{x\in\Omega:\mu(x)>\nu(x)} \mu(x) - \nu(x).$$

In Markov chains the total variation distance is useful to measure the distance between the t-step transition probability and the stationary distribution. We define:

$$d(t) = \max_{x\in S} |P^t_{x,\cdot} - \pi|_{TV},$$

where $P^t_{x,\cdot}$ is containing the t-step transition probability when the initial state is $X_0 = x$ and π is the stationary distribution of the chain.

An alternative way to measure the distance between the t-step transition probability and the stationary distribution is the separation distance.

Definition B.4 (Separation Distance) *The separation distance for a Markov chain with initial state $x \in S$ is:*

$$S_x(t) = \max_{y \in S}(1 - \frac{P^t_{x,y}}{\pi(y)}).$$

It is also useful to define:

$$S(t) = \max_{x \in S}(\max_{y \in S} 1 - \frac{P^t_{x,y}}{\pi(y)}).$$

Definition B.5 (Stationary Time) *Let $(X_t)_{t \geq 0}$ be an irreducible Markov chain with stationary distribution π. A stationary time τ for (X_t) is a randomized stopping time such that X_τ is chosen according to π. That is:*

$$\mathbb{P}(X_\tau = y) = \pi(y).$$

Definition B.6 (Strong Stationary Time) *A strong stationary time for a Markov chain $(X_t)_{t \geq 0}$ with stationary distribution π is a randomized stopping time τ, such that:*

$$\mathbb{P}_x(\tau = t, X_t = y) = \mathbb{P}_x(\tau = t)\pi(y),$$

where $\mathbb{P}_x(.)$ denotes a probability conditioned on the initial state of the chain being x.

The following Proposition B.7 proves the existence of a strong stationary time for ergodic Markov chains, we will call it optimal strong stationary time.

Proposition B.7 *Let $(X_t)_{t \geq 0}$ be an irreducible and aperiodic Markov chain with states space S, for every initial state $x \in S$ there is an strong stationary time τ such that for $t \geq 0$:*

$$S_x(t) = \mathbb{P}_x(\tau > t).$$

We prove some auxiliary results before proving Proposition B.7

Lemma B.8 *Let $(X_t)_{t \geq 0}$ be an irreducible Markov chain with stationary distribution π. If τ is a strong stationary time for (X_t), then for every $t > 0$:*

$$\mathbb{P}_x(\tau \leq t, X_t = y) = \mathbb{P}_x(\tau \leq t)\pi(y).$$

Lemma B.8 tells us that the chain follows the stationary law in every time step after τ.

Proof. Let $Z_1, Z_2, \ldots$ be the sequence of iid random variables used in the random function representation of (X_t), then for any $s \leq t$:

$$\mathbb{P}_x(\tau = s, X_t = y) = \sum_{i \in S} \mathbb{P}_x(X_t = y \mid \tau = s, X_s = i)\mathbb{P}_x(\tau = s, X_s = i).$$

We can take the composition of r random function representations, getting $f_r : S \times \Omega^r \to S$ such that:

$$X_{s+r} = f_r(X_s, Z_{s+1}, Z_{s+2}, \ldots, Z_{s+r})$$

Since $\tau = s$ is a randomized stopping time, it depends only on $(Z_1, Z_2, \ldots, Z_s)$, and is independent of $(Z_{s+1}, Z_{s+2}, \ldots, Z_{s+r})$. We have that:

$$\mathbb{P}_x(X_t = y \mid \tau = s, X_s = i) = \mathbb{P}_x(f_{t-s}(i, Z_{s+1}, Z_{s+2}, \ldots, Z_{s+t}) = y \mid \tau = s, X_s = i) = P_{i,y}^{t-s},$$

Therefore:

$$\mathbb{P}_x(\tau = s, X_t = y) = \sum_{i \in S} P_{i,y}^{t-s}\pi(i)\mathbb{P}_x(\tau = s) = \pi(y)\mathbb{P}_x(\tau = s) \implies$$

$$\sum_{s \leq t} \mathbb{P}_x(\tau = s, X_t = y) = \sum_{s \leq t} \pi(y)\mathbb{P}_x(\tau = s) \implies$$

$$\mathbb{P}_x(\tau \leq t, X_t = y) = \pi(y)\mathbb{P}_x(\tau \leq t).$$

■

Proof of Proposition B.7. Fix a state $x \in S$ and let $a_t = \min_y \frac{P_{x,y}^t}{\pi(y)} = 1 - S_x(t)$. If there is a strong stationary time that satisfies $S_x(t) = \mathbb{P}_x(\tau > t)$, it also holds that:

$$\mathbb{P}(\tau = t) = \mathbb{P}(\tau > t-1) - \mathbb{P}(\tau > t)$$

$$= S_x(t-1) - S_x(t) = (1 - S_x(t)) - (1 - S_x(t-1)) = a_t - a_{t-1},$$

we conclude that:

$$\mathbb{P}_x(X_t = y, \tau = t) = \pi(y)\big(a_t - a_{t-1}\big). \tag{B.9}$$

By Lemma B.8:

$$\mathbb{P}(X_t = y, \tau \leq t) = \mathbb{P}_x(\tau \leq t)\pi(y) = \pi(y)a_t. \tag{B.10}$$

Let $(U_i)_{i\geq 1}$ be a sequence of iid random variables taking values in $(0,1)$ and independently of (X_t). We define:

$$\tau = \min \left\{ t \geq 1 : U_t \leq \frac{a_t - a_{t-1}}{\frac{P^t_{x,X_t}}{\pi(X_t)} - a_{t-1}} \right\}.$$

By construction, we have that:

$$\mathbb{P}(\tau = t \mid X_t = y, \tau > t-1) = \frac{a_t - a_{t-1}}{\frac{P^t_{x,y}}{\pi(y)} - a_{t-1}}.$$

We need to show that τ satisfies (B.9) for every $t \geq 1$. We prove by induction.

$$\mathbb{P}_x(X_t = y, \tau = t) = \mathbb{P}(\tau = t \mid X_t = y, \tau > t-1)[\mathbb{P}(X_t = y) - \mathbb{P}(X_t = y, \tau \leq t-1)].$$

The base case $t = 1$:

$$\mathbb{P}_x(X_t = y, \tau = 1) = \mathbb{P}(\tau = 1 \mid X_1 = y, \tau > 0)[\mathbb{P}(X_1 = y) - \mathbb{P}(X_1 = y, \tau \leq 0)] =$$

$$\frac{a_1 - a_0}{\frac{P_{x,y}}{\pi(y)} - a_0}[P_{x,y} - \mathbb{P}(X_1 = y, \tau \leq 0)],$$

we have $a_0 = \min_y \frac{P^0_{x,y}}{\pi(y)} = 0$ e $\mathbb{P}(X_1 = y, \tau \leq 0) = 0$, thus:

$$\mathbb{P}_x(X_t = y, \tau = 1) = \pi(y)(a_1 - a_0),$$

concluding the proof of the base case. Assume that the statement holds for every $s < t$. Then:

$$\mathbb{P}_x(X_t = y, \tau = t) = \frac{a_t - a_{t-1}}{\frac{P^t_{x,y}}{\pi(y)} - a_{t-1}}[P^t_{x,y} - \mathbb{P}(X_t = y, \tau \leq t-1)],$$

By the induction hypothesis, we know that (B.10) holds, meaning that:

$$\mathbb{P}_x(X_t = y, \tau = t) = \frac{a_t - a_{t-1}}{\frac{P^t_{x,y}}{\pi(y)} - a_{t-1}}[P^t_{x,y} - \pi(y)a_{t-1}] = \pi(y)[a_t - a_{t-1}],$$

concluding the proof. ■