ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Bootstrap Percolation on Complex Networks

Master Thesis

C. Swaneet K. Sahoo

Friday 1st March, 2019

Advisors: Prof. Dr. Bernd Gärtner, Ahad N. Zehmakan
Department of Computer Science, ETH Zürich

Abstract

α-bootstrap percolation is a dynamic process on a graph in which every node starts with an initial binary state - either black or white. Every round, a node becomes black if at least α-fraction of its neighbors are black. Otherwise it is unchanged. A minimum dynamic monopoly D is a smallest set of nodes in a graph that have the property that the entire graph becomes black eventually if these nodes D nodes are black initially. We investigate the size of D under α-bootstrap percolation on the preferential attachment graph model (also known as the Barabási-Albert model) as an example of a complex network. While we did not achieve a final result, we summarize our intermediate results, share insights and give ideas on how one might continue this investigation or use them in another context.

Contents

Chapter 1

Introduction

1.1 Complex Networks and Bootstrap Percolation

Complex networks are ubiquitous in nature and in our human-created world. A complex network is a network of entities (such as humans, genes, species, websites etc.) where the relationships are created from interdependent and non-trivial interactions between the entities. These complex network ignite a sense of fascination and are of high interest in diverse disciplines in sciences and industrial application. Popular examples of include the social networks (e.g. Twitter/Facebook), networks of scientific/artistic collaboration, the internet-topology graph and many more. From the scientific point of view it is insightful to gain a better understanding of the different complex networks that arise in nature and society. Such an understanding is on one hand tremendously useful in practise - such as effective epidemic prevention from knowledge on the network of human interactions. On the other hand, it can help us get an intuition of the complexity of nature and society in general as one might find re-occurring general featured and properties in widely diverse contexts. In the past five decades much progress in understanding complex networks was possible due to the increasing availability of large networks in form of data-sets. Insights from such data-sets are a difficult task - thus, there have been many attempts to get a gain an understanding of these networks by searching for concise mathematical models that reproduce key features.

Preferential Attachment Among these models is the preferential attachment graph model first introduced by Barabási and Albert in [6]. Their key observation was that networks grow into their structure when new nodes (entities) connect to existing nodes with a preference towards those who are already well connected. This is consistent with everyday experiences - for example whenever one joins a new social group (e.g. by joining a writing club) one is more likely to connect to someone who is already popular and well known than to someone who is present rarely. The model concisely expresses this preference mathematically: Let $n, m \geq 1$ be parameters denoting the final number of nodes n and m the number of out-going edges per node. The model defines a preferential attachment graph (shortly PA graph) G_m^n. Starting with the graph G_m^1 containing a single node with m loops - G_m^{t+1} is then recursively built from G_m^t by adding a new node $t+1$ and connecting each of its m outgoing edges to nodes $i \in V(G_m^{t+1})$ with probability proportional to the degree $\deg(i)$ of node i. The degree proportional connection probability encapsulates this preference. This model enjoys many properties found in real-world complex network.

Dynamics and Bootstrap Percolation While topology is one side of the coin - dynamics is the other. Dynamics on such graphs, where each node has a state and its future state depends only on the local neighborhood, have been found to be useful and insightful to investigate and understand real-world dynamics. They are extensively used to model rumor/infection/meme spreading, technological-adaptation, neighbor-based voting and more. One models for these dynamics is the α-bootstrap percolation model (shortly α-BP) for some $\alpha \in [0, 1]$. In α-BP every node has state either black or white and at each (discrete) time-round t, a node becomes black if at least α-fraction of its neighbors were black at time $t-1$. Otherwise it keeps its state. α-BP and variations thereof have been fruitful models for real world network dynamics such as rumor spreading or technological-adaptation - and they are also interesting from a theoretical analysis point of view.

1.2 Aim of the thesis

The thesis aims to contribute to a better understanding of the dynamics of α-bootstrap percolation (shortly α-BP) on preferential attachment graphs. Concretely, we are interested in dynamic monopolies. A dynamic monopoly is a set of black nodes in a graph which under the dynamics of α-BP eventually makes the entire graph black regardless of the initial state of the remaining nodes. An interesting question is: What is the size of the *smallest* dynamic monopoly in a graph? We aim to give an answer to this question for the class of preferential attachment graphs - and share insights and intermediate results of this investigation.

1.3 Outline

In Chapter 2 we begin by giving concrete definitions and summarizing related prior work. Chapter 3 is the main contribution of the thesis where we present the approaches we have taken and share some insights and partial results. Finally in Chapter 4 we discuss the results and make a conclusion.

1.4 Notation

The following general conventions will be used in this thesis. Most of them will be re-introduced when first used. All graphs are undirected multi-graphs unless mentioned otherwise.

Notation	Definition
$\mathbb{N}_0$	natural numbers $\{0, 1, 2, \dots\}$
$\mathbb{N}^+$	natural numbers $\{1, 2, 3, \dots\}$
$[n]$	$\{1, \dots, n\}$ for $n \in \mathbb{N}^+$
$\mathbb{1}(c)$	indicator function with $\mathbb{1}(c) = 1$ if c and $\mathbb{1}(c) = 0$ otherwise
$\mathrm{mult}(x, A)$	multiplicity of element x in multiset A. For example given $A = \{a, b, a\}$ we have $\mathrm{mult}(a, A) = 2$, $\mathrm{mult}(b, A) = 1$ and $\mathrm{mult}(x, A) = 0$ for all $x \in \mathcal{U} - \{a, b\}$ in universe $\mathcal{U}$
$\|A\|$	the cardinality of the multiset A defined as $\sum_{x \in \mathcal{U}} \mathrm{mult}(x, A)$
$f(n) \in \mathcal{O}(g(n))$	g is an asymptotic upper bound on f - i.e. $\limsup_{n \to \infty} \frac{f(n)}{g(n)} < \infty$
$f(n) \in \mathrm{o}(g(n))$	asymptotically f grows slower than g - i.e. $\limsup_{n \to \infty} \frac{f(n)}{g(n)} = 0$
$f(n) \ll g(n)$	alternate notation for $f(n) \in \mathrm{o}(g(n))$
$f(n) \in \Omega(g(n))$	$g(n) \in \mathcal{O}(f(n))$
$f(n) \in \omega(g(n))$	$g(n) \in \mathrm{o}(f(n))$
$f(n) \in \Theta(g(n))$	$g(n) \in \mathcal{O}(f(n))$ and $f(n) \in \mathcal{O}(g(n))$
$\log n$	natural logarithm of n
w.h.p.	an event occurs *with high probability* if the probability is $1 - \mathrm{o}(1)$ for $n \to \infty$
$G_{n,p}$	Erdős-Rényi graph model (binomial random graph) where each pair of nodes is independently connected with probability p
$\mathbb{H}^d$	d-dimensional hypercube - i.e. $\mathbb{H}^d = (V, E) = (\{0,1\}^d,\ \{\ (x, y) \mid \|x - y\| = 1\ \})$
$T_{m,n}$	undirected torus (toroidal mesh) with $T_{m,n} = ([m] \times [n], E)$ and $m \leq n$ without loss of generality where $E = \bigcup_{(i,j) \in [m] \times [n]} \{((i,j), (i+1 \mod m, j)),\ ((i,j), (i, j+1 \mod m))\}$

Chapter 2

Theory and Background

In this chapter, the preferential attachment graph model and bootstrap percolation will be precisely defined. Then after summarizing established results for both models, we will give our results regarding α-BP on preferential attachment graphs.

2.1 Preferential Attachment Graphs

We will first present the sequential preferential attachment model (shortly PA) equivalently defined in [10] and then provide an alternative model that is technically easier to work with. The model is also known as the *Barabási-Albert model*.

Definition 2.1 (Preferential Attachment Graph [10]) *Let $m \geq 1$ be a constant integer. We define the random undirected multigraph $G_m^n = (V, E)$ on the node set $V = [n] = \{1, .., n\}$ inductively. E is a multiset of edges $e \in V \times V$. G_m^n is a random graph from the space $\mathcal{G}_m^n$ of multigraphs over the node set $V = [n]$ with $|E| = nm$. The sampling method defines a probability distribution on $\mathcal{G}_m^n$:*

Let $G_{m,j}^t$ denote the intermediate graph that contains t nodes and where the last node t has connected $j \leq m$ edges. We start with $G_{m,m}^1$ - the graph containing only the node $v_1 = 1$ with m self-loops. Then, from $G_{m,m}^{t-1}$, one constructs $G_{m,0}^t$ by adding a new node $v_t = t$. Next, m edges are added sequentially. Let $\deg(k, G)$ denote the number of edges incident to node k in the graph G and let $D_{m,j}^t = \sum_{i=1}^t \deg(i, G_{m,j}^t)$ be the sum of all degrees in $G_{m,j}^t$. From $G_{m,j-1}^t$, one constructs $G_{m,j}^t$ by adding a new edge $e_{t,j} = \{t, v\}$ where $v \in [t]$ is chosen randomly with probability

$$\mathbb{P}(v = k) = \frac{\deg(k, G_{m,j-1}^t) + \mathbb{1}(k = t)}{D_{m,j-1}^t + 1}$$

Essentially, we count the degree of each node before connecting to it. The indicator $\mathbb{1}(k = t)$ and the $+1$ in the denominator counts the out-going 'half' of the currently added edge.

The sequential model, though easy to understand, is difficult from a technical point of view, because every random choice is dependent on previous outcomes. Next, we shall present an alternative model also introduced in [10] and [9] which is technically easier to handle.

Definition 2.2 (Preferential Attachment Graph, alternative model) *Let $m \geq 1$. For each $i \in [n]$ and $j \in [m]$ let $(x'_{i,j}, y'_{i,j})$ be independent and uniform samples from $[0,1] \times [0,1]$ with $x'_{i,j} < y'_{i,j}$. With probability 1 all numbers are distinct. Let $(x_{i,j}, y_{i,j})_{i,j}$ be a relabeling where the $y_{i,j}$ are lexicographic sorted - i.e. $y_{1,1} < \cdots < y_{1,m} < y_{2,1} < \cdots < y_{2,m} < \cdots < y_{n,m}$. Let $G^n_m = (V, E)$ denote the undirected multigraph on $V = [n]$. Also, let $W_0 := 0$, $W_i := y_{i,m}$ and $w_i := W_i - W_{i-1}$ be weights assigned to each node i. The weights will be tightly linked to their eventual degrees. For each $x_{i,j}$ an edge between node i and node k is added where k is chosen to satisfy $W_{k-1} < x_{i,j} < W_k$. Or equivalently formulated, we have $\operatorname{mult}(\{i,k\}, E) = |\{\ \{i,j\} \mid W_{k-1} < x_{i,j} < W_k \text{ for } j \in [m]\ \}|$ for $i \geq k$. This sampling method defines a probability distribution over $\mathcal{G}^n_m$ that is equivalent to the sequential model.*

Given $(y_{i,j})_{i,j}$ the random variables $(x_{i,j})_{i,j}$ are independently distributed $x_{i,j} \sim U(0, y_{i,j})$. When the proof statements allow for it, we will make the further simplification, that $x_{i,j} \sim U(0, W_i)$. This only slightly increases the probability of a loop.

A few key properties, established in [10] and [21], are satisfied with high probability for the weigths w_i and W_i from the definition.

Lemma 2.3 ([10], [21]) *Let $m \geq 2$. Let $s = 2^a$ be the smallest power of 2 satisfying $s > \log^7 n = (\log n)^7$ and let 2^b be the largest power of 2 satisfying $2^b < \frac{2}{3}n$. We restrict our attention to large graphs G^n_m where $a < b$.* [1] *For $a \leq t \leq b$ we consider the intervals of nodes $I_t = [2^t + 1, 2^{t+1}]$. The following events hold with probability tending to 1 as $n \to \infty$:*

The W_i have a certain distribution:
$E_1 = \left\{ \left| W_i - \sqrt{\frac{i}{n}} \right| \leq \frac{1}{10}\sqrt{\frac{i}{n}} \text{ for } s \leq i \leq n \right\}$

Many weights w_i are lower-bounded in their intervals:
$E_2 = \left\{ I_t \text{ contains at least } 2^{t-1} \text{ nodes with } i \text{ with } w_i \geq \frac{1}{10\sqrt{in}} \text{ for } a \leq t < b \right\}$

Early or late nodes have high or respectively low weights:
$E_3 = \left\{ w_1 \geq \frac{4}{\log n \sqrt{n}} \right\}$

[1] $a < b$ is satisfied for all $n \geq 3 \cdot 10^{10}$

$E_4 = \left\{ w_i \geq \frac{\log^2 n}{n} \textit{ for } i < n^{1/5} \right\}$
$E_5 = \left\{ w_i \leq \frac{\log^2 n}{n} \textit{ for } i \geq \frac{n}{2} \right\}$

Since these properties hold with high probability, we can simply establish results assuming $W_1, \ldots, W_n$ are given and satisfy $E_1, \ldots, E_5$. Results proven for these therefore hold for the general case with high probability.

2.2 Bootstrap Percolation and Dynamic Monopoly

In this section, we will introduce a general framework to formalize dynamical processes on graphs and give precise definitions for r-BP, α-BP and the majority model. The definitions are generalizations of those presented in [29] and [47], taking multi-edges and loops into account. Finally, we will present two-way variants of these processes.[2]

Definition 2.4 (Bootstrap Percolation Process) *Let $G = (V, E)$ be a multigraph. A configuration $\mathcal{C} : V \to \{b, w\}$ assigns a binary state, black or white, to each node. Let $\varphi : V \times \{b, w\} \times \mathbb{N}_0 \to \{b, w\}$ be a given function. Let $N(v)$ denote the multiset of neighbors of node v which satisfies $\mathrm{mult}(u, N(v)) = \mathrm{mult}(\{v, u\}, E)$. Each node u is counted with multiplicity equal to the number of edges between u and v. Let $N_c^{\mathcal{C}}(v)$ be the multiset of neighbors of v with state c in configuration $\mathcal{C}$ - i.e. $\mathrm{mult}(u, N_c^{\mathcal{C}}(v)) = \mathbb{1}(\mathcal{C}(u) = c) \cdot \mathrm{mult}(u, N(v))$. We will define the sequence of configurations $(\mathcal{C}_t)_{t \in \mathbb{N}_0}$ starting at a given initial configuration $\mathcal{C}_0$. For all $t \geq 1$ and $v \in V$ we inductively define*

$$\mathcal{C}_t(v) = \varphi(v, \mathcal{C}_{t-1}(v), |N_b^{\mathcal{C}_{t-1}}(v)|).$$

Thus the state of node v depends on the previous state of itself and the number of black neighbors as dictated by φ. We define $\mathcal{B}_t := \{ v \in V : \mathcal{C}_t(v) = b \}$ as the set of black nodes at time t. If a black node stays black forever - i.e. $\forall v, k.\ \varphi(v, b, k) = b$ - then we call it monotone bootstrap percolation or irreversible bootstrap percolation.

Having this definition in hand, we are ready to formally define r-BP, α-BP and their two-way variants as well as the majority model.

Definition 2.5 *Using definition 2.4, we define φ for each model on a graph G. Let $r \in \mathbb{N}^+$ and $\alpha \in (0, 1]$ be constants.*

[2] also called *reversible bootstrap percolation*

Name	Definition
monotone r-BP, φ_r	$\varphi_r(v,c,k) = \begin{cases} b & \textit{if } k \geq r \textit{ or } c = b \\ w & \textit{otherwise} \end{cases}$
two-way r-BP, φ_{2w-r}	$\varphi_{2,r}(v,c,k) = \begin{cases} b & \textit{if } k \geq r \\ w & \textit{otherwise} \end{cases}$
monotone α-BP, φ_α	$\varphi_\alpha(v,c,k) = \begin{cases} b & \textit{if } k \geq \alpha \cdot \deg(v,G) \textit{ or } c = b \\ w & \textit{otherwise} \end{cases}$
two-way α-BP, $\varphi_{2w-\alpha}$	$\varphi_{2,\alpha}(v,c,k) = \begin{cases} b & \textit{if } k \geq \alpha \cdot \deg(v,G) \\ w & \textit{otherwise} \end{cases}$
majority model, φ_{maj}	$\varphi_{\text{maj}}(v,c,k) = \begin{cases} c & \textit{if } k = \frac{1}{2}\deg(v,G) \\ b & \textit{if } k > \frac{1}{2}\deg(v,G) \\ w & \textit{otherwise} \end{cases}$

When speaking of r-BP and α-BP we refer to the monotone variants - unless explicitly mentioned. In literature, r-BP is sometimes called *k-complex contagion* [23] or seen as a special case of the *threshold model* [33]. Note that the only difference between α-BP with $\alpha = \frac{1}{2}$ and the majority model is the way they handle ties. In case of a tie the majority model keeps the current state whereas $\frac{1}{2}$-BP chooses black. Majority model is invariant under the transformation $b \leftrightarrow w$, making it symmetric. In contrast, r-BP and α-BP are always non-symmetric.

We are interested in the smallest set of initially black nodes that eventually make all nodes black. Such a set is called a *dynamic monopoly* [36], a *contagious set* [20,23] or a *target set* [18]. We present a definition following [36].

Definition 2.6 (Dynamic Monopoly) *Let $G = (V,E)$ be a multigraph and φ be a function as in definition 2.4. Let $I \subseteq V$ be a set of nodes and $\mathcal{B}_0 = I$. I is called a dynamic monopoly of model φ on G if there exists $T_0 \in \mathbb{N}_0$ such that $\forall t \geq T_0$. $\mathcal{B}_t = V$ - i.e. if the nodes I are black initially then eventually every node in G becomes black and stays so regardless of their initial color. A minimum dynamic monopoly, denoted $I \in \text{dyn}^*(G,\varphi)$, is a smallest set I that is a dynamic monopoly of model φ on G.*[3]

In the next section, we will present previous work on preferential attachment graphs and bootstrap percolation.

[3] If $\varphi$and G are clear from the context, then we simply write $I \in \text{dyn}^*$

2.3 Prior Work

Preferential Attachment Graphs The preferential attachment model has been a well-known and investigated complex network model in the past decades since its introduction in [6]. It was constructed to resemble imitate the real world processes that lead to complex networks. It was built to exhibit front and fore-most a power-law degree-distribution similar to those of real world networks. Its power-law degree distribution, i.e. the fraction of nodes with degree d being proportional to $d^{-\gamma}$ for some $\gamma \in \mathbb{R}_{>0}$, was formally proved shortly after in [9]: With high probability the fraction of nodes with degree d is approximated by $f(d) = \frac{2m(m+1)}{d(d+1)(d+2)}$ for $d \in \{0, 1, \ldots, n^{1/15}\}$. Since $f(d) \sim d^{-3}$ we have $\gamma = 3$. Real world networks haven often found to be well approximated with exponents $2 < \gamma < 3$. The attachment procedure in the model can be slightly adjusted to exhibit exponents smaller than 3 as argued in [6]. In [9] it was informally argued that w.h.p. the maximum degree is $\Theta(\sqrt{n})$. Another relevant property is the diameter - that is the length of the longest *shortest path* between two nodes in the graph. In [10] it was established, that for $m \geq 2$ with high probability a PA graph G_m^n is connected and the diameter is approximately $\frac{\log n}{\log \log n}$. For $m = 1$ the graph is w.h.p. not connected - making a forest consisting of $\mathcal{O}(\log n)$ trees. The longest shortest path between two nodes in the same tree is approximately given by $c \log n$ for a constant $c \in [3, 4]$. This is in good consistency with real-world networks. Finally consider indicators of clustering for random graphs. Among these is the clustering coefficient C_v of node v defined as the fraction of neighbors of v that are directly connected to each other [46]. For v having degree d we have $C_v = \frac{2l_v}{d(d-1)}$ where l_v is the number of edges between neighbors of v. PA graphs have small average clustering $\langle C \rangle = \frac{1}{n} \sum_v C_v$ with $\langle C \rangle \approx \frac{m}{8} \frac{\log n}{n}$ as proven in [39]. The small clustering of PA graphs is one of the few properties where it differs from real-world complex networks - latter usually having a large clustering coefficient. Over the years many more properties have been discovered. A good starting point for getting an overview is the book [7].

***r*-Bootstrap Percolation** Bootstrap percolation was first introduced in questions arising in statistical physics in [11]. r-BP and variants have been extensively investigated on various random graph models - including PA graphs. Among the important questions investigated is the probability of the set of black nodes of a random configuration being a dynamic monopoly and bounds on the size of the minimum dynamic monopoly for a family of random graphs. Concretely, given $\rho \in [0, 1]$, if initially every node is black with probability ρ independently of others, one asks what the eventual evolution of the model becomes on graphs of a given class. For the d-dimensional hypercube, it was proved [3], that under 2-BP the probability of a random initial configuration being a dynamic monopoly tends to 1 and $d \to \infty$ for $\rho \geq \frac{c_1}{d^2 4^{\sqrt{d}}}$ - or tends to 0 for $\rho \leq \frac{c_2}{d^2 4^{\sqrt{d}}}$ - for constants c_1, c_2.[4] They also showed, that the hypercube demonstrates *sharp threshold behavior* (also called *critical phenomena*) - meaning that the probability of a dynamic monopoly sharply transitions from 0 to 1 at the above probability bounds for ρ. As we shall see, such sharp threshold behavior are common for dynamic processes on graphs. For the Erdős-Rényi graph $G_{n,p}$, an investigation in [35] revealed sharp threshold behavior where - depending on the parameters n, p - a random initial configuration either makes at-least $n - o(n)$ or at-most $o(n)$ nodes black eventually. The d-dimensional torus also gained considerable amount of attention due to its application in statistical physics [47]. Both monotone r-BP and two-way r-BPwere found to show sharp threshold behavior by [4] and [47]. Other families of graphs investigated include random regular graphs [16] and infinite trees [37].

The idea of giving bounds on the minimum size of a dynamic monopoly - where the initial configuration is not random anymore - was introduced in [38] and [42]. Research interest in dynamic monopolies grew, and as a consequence, already established random graphs models and bootstrap percolation processes were investigated for dynamic monopolies. Starting with the binomial random graph $G_{n,p}$ it was proved, by [25], that assuming $r \geq 2$ and $p = d/n$ where $1 \ll d \ll \left(\frac{n \log \log n}{\log^2 n}\right)^{(r-1)/r}$ the size of the minimal dynamic monopoly $|I|$, for $I \in \text{dyn}^*(G_{n,p}, \varphi_r)$, satisfies $|I| \in \Theta\left(\frac{n}{d^{r/(r-1)} \log d}\right)$. d is the average degree of the binomial graph - therefore the results demonstrates minimum dynamic monopolies linear in n - while also being smaller for larger average degree d. The case for the n-dimensional hypercube $\mathbb{H}^d$ was conjectured by [3] and recently proven in [41]. Concretely, they showed that $|I| = \frac{1+o(1)}{r} \binom{d}{r-1}$ for $I \in \text{dyn}^*(\mathbb{H}, \varphi_r)$. Thus the size of the minimum dynamic monopoly is essentially polynomial in d with exponent $r-1$.This is considerably small given that the hypercube consists of $n = 2^d$ nodes leading to a poly-logarithmic size in n. Dynamic monopolies under r-BP have been investigated for many other graphs such as random regular graphs [34], multidimensional tori [36], dense graphs with dynamic monopoly of size r [28] and on graphs with a given degree-sequence [44].

Now let us focus on social network models - such as PA graphs, other power-law random graph models etc. Random graphs with power-law degree-distribution can be generated in many ways - among which PA is only one. A popular alternate model for power-law random graphs is the Chung-Lu model introduced in [19]. In the Chung-Lu model, one begins with a se-

[4] Note that, we have $d = \log_2 n$ in the number of nodes n.

quence of weights $= (w_1, \ldots, w_n)$ assoociented with the nodes $[n]$ representing an expected degree-sequence. Then every pair of nodes i, j is independently connected with probability proportional to $w_i w_j$. With a corresponding choice of the weights $(w_1, \ldots, w_n)$ the generated graph is power-law. r-BP was investigated for the Chung-Lu model in [2]. They determined a critical sublinear function $a_c(n)$, such that if $a_c(n) \ll |\mathcal{B}_0|$ then a linear fraction of the nodes will become black w.h.p. This is in contrast to the binomial graph $G_{n,p}$ where a linear number of nodes is nesessary. The investigation of PA graphs was recently settled in [23]. They showed that the size of the minimum dynamic monopoly for PA graphs is exactly r with high probability - i.e. $|I| = r$ for $I \in \mathrm{dyn}^*(G_m^n, \varphi_r)$. They showed, that w.h.p. eventually all nodes become black if the *first* r nodes are infected ($I = [r]$). This is surprising, as it stands in strong contrast to the linear size of minimum dynamic monopolies for the binomial graph (or poly-logarithmic for the hypercube). They also established a polynomial bound $f(n)$ where a uniformly-random set of initially black nodes of size $o(f(n))$ will not be a dynamic monopoly w.h.p. However, if one instead chooses $\Omega(f(n) \log n)$ nodes black initially, then the entire graph will become black w.h.p. Therefore, while a constant number of nodes r is sufficient to make all nodes black, the choice heavily matters as randomly infecting sub-polynomially many nodes does not suffice. Finally, r-BP was also investigated for dynamic monopolies on various other social network models - see [32] and [24].

Majority Model and α-bootstrap percolation Related to r-BP is α-BP and the majority model. Both have been fairly well-studied. We start by giving an overview of the main results for the majority model and then continue to α-BP. Starting with the binomial graph $G_{n,p}$, it was recently proved, that $G_{n,p}$, under the majority model, shows a phase transition at $p^* = \frac{\log n}{n}$. Concretely, if p_b is the probability that a node is colored black initially and $p_b \leq \frac{1}{2} - \omega(\frac{1}{\sqrt{np}})$, then for $p > p^*$ all nodes will become white in constantly many rounds w.h.p. This is interesting, as, in the realm $p > (1+\varepsilon)p^*$ for some $\varepsilon > 0$, a very small majority of a color initially will lead to an eventual uniform configuration. For $p < (1-\varepsilon)p^*$, we have w.h.p. co-existence of both colors if p_b is large enough ($p_b \in \omega(e^{np}/n)$). If the density of p_b is lower ($p_b \in o(e^{np}/n)$), then we get a fully black configuration w.h.p. Majority model therefore can be considered a good density classifier on $G_{n,p}$ graphs. Dynamic monopolies were also investigated in the same paper. It was established, that every dynamic monopoly $I \in \mathrm{dyn}^*(G_{n,p}, \varphi_{\mathrm{maj}})$ has to contain almost half of the nodes - $|I| \geq n(\frac{1}{2} - \frac{c}{\sqrt{np}})$ for some constant $c > 0$ to be exact. This result is in a similar spirit as the case for initially random configurations. In the more general setting of α-BP, a result was given in [14] - namely, that $|I| \in \Omega(\alpha n)$ for $I \in \mathrm{dyn}^*(G_{n,p}, \varphi_\alpha)$. This is easily explained, as the independent and uniform connectivity of $G_{n,p}$ leads to the result, that at least α-fraction of the nodes need to be black for all nodes to become black. The similarity of α-BP to r-BP is equally intuitive, as $G_{n,p}$ has a concentrated binomial degree distribution. This homogenity leads to similar behavior of r-BP to α-BP for $\alpha np \approx r$. Next, let us consider grids and tori. Two-dimensional grids and tori were investigated in [30] under the majority model for 4-neighborhood. If the nodes are black with probability $p_b \leq 1/2$ again, then there is threshold behavior - at $f(n) = n^{\frac{1}{2}}$. Concretely, if $p_b \ll f(n)$, then there are too few black nodes and eventually every node will become white w.h.p. Otherwise for $f(n) \ll p_b$ both graphs show co-existence w.h.p. Dynamic monopolies on tori have been investigated in [27] - proving upper and lower bounds for the torus $T_{m,n}$. Concretely, they showed that given $I \in \text{dyn}^*(T_{m,n}, \varphi_{\text{maj}})$[5], we have $\frac{mn+1}{3} \leq |I| \leq \lceil \frac{n}{3} \rceil (m+1)$. The torus, like $G_{n,p}$, needs a linear number of black nodes for a dynamo. Other relevant results include general graphs [42], random regular graphs [31], hypercube [26] [40], multidimensional cube [5], power-law graphs [43] and many more as listed in [15] [17] and [13]. An interesting variant of majority model was investigated in [1] where k random neighbors are chosen before applying the majority rule on graphs with given degree sequences. There is sparse literature regarding the general α-BP. An investigation for general undirected and connected graphs was done in [12] - proving a linear upper bound $|I| \leq (2\sqrt{2}+3)\alpha n$ for $I \in \text{dyn}^*(G, \varphi_\alpha)$. A similar bound was shown in [15] for directed graphs when the influence of only spreads in direction of the edges. The constant was improved later in [29] to 2 - yielding $|I| < 2\alpha n$. In the same paper, they gave an upper bound for graphs G' with cycles of length atleast 5 - i.e. girth ≥ 5. Concretely, they showed $|I| < (1+\varepsilon)\alpha n$ for $I \in \text{dyn}^*(G', \varphi_\alpha)$ and any $\varepsilon > 0$. Both results were shown to be optimal bounds in their respective range of applicability. An early general investigation of a variant of α-BP where the threshold is different for each node and it drawn from a probability distribution is given in [45].

[5] where $m \leq n$ without loss of generality

Chapter 3

Bootstrap Percolation on Preferential Attachment Graphs

In this section, we share the experience gained in investigating α-BP on PA graphs. From the many approaches tried out, only a few of them were fruitful or could be finalized into a proper intermediate results. Since we were not able to get final answers on the size of the dynamics monopoly, we are instead going to give a few main partial results and insights. After the Preliminaries, we will share the main learnings from working on the model directly (page 15), from using Computer simulations (page 24) and from working on the alternative model (page 28). Finally we will present a few notable approaches and ideas.

3.1 Preliminaries

We list a series of useful lemmas we will use for the subsequent proofs.

Lemma 3.1 *Let f be a continuous decreasing function on $[a,b]$. A sum over f is bounded by the following integrals:*

$$\int_a^{b+1} f(t)\mathrm{d}t \le \sum_{i=a}^{b} f(i) \le f(a) + \int_a^b f(t)\mathrm{d}t$$

If f is increasing, we instead have

$$f(a) + \int_a^b f(t)\mathrm{d}t \le \sum_{i=a}^{b} f(i) \le \int_a^{b+1} f(t)\mathrm{d}t.$$

Proof Let f be a decreasing function on $[a,b]$. We focus on f decreasing as the case of an increasing f follows immediately by investigating $-f$. Consider the graph of f on the interval $[a,b]$. The area covered by every term

$f(i)$ in the sum is lower bounded by the corresponding integral $\int_i^{i+1} f(t)\mathrm{d}t$. Summing them all together yields the lower bound $\int_a^{b+1} f(t)\mathrm{d}t \leq \sum_{i=a}^{b} f(i)$. Similarly, every term $f(i)$ of the sum is upper bounded by the integral on the shifted function - namely by $\int_i^{i+1} f(t-1)\mathrm{d}t = \int_{i-1}^{i} f(t)\mathrm{d}t$. Applying this bound on every term but the first (namely $f(a)$), we get the upper bound $\sum_{i=a}^{b} f(i) \leq f(a) + \int_a^b f(t)\mathrm{d}t$. □

We are using the following Chernoff bound on the sum of independent Bernoulli variables as given in [10].

Lemma 3.2 (Chernoff Bound) *Let $X = \sum_{i\in[n]} X_i$ be the sum of independent Bernoulli random variables where $X_i \sim Ber(p_i)$ for $p_i \in [0,1]$. Let $\mu = \mathbb{E}[X] = \sum_i p_i$. Then for $t \geq 0$, we have*

$$\mathbb{P}(X \geq \mu + t) \leq \exp\left(-\frac{t^2}{2(\mu + t/3)}\right)$$

and

$$\mathbb{P}(X \leq \mu - t) \leq \exp\left(-\frac{t^2}{2\mu}\right).$$

Azumas Inequality Given the expectation of a random variable (e.g. the expected degree of the k-th node $\mathbb{E}[\deg(k, G_m^n)]$) - a variant of Azumas Inequality can be used to show a concentration result if the evolution of the random variable throughout the generation process $G_m^1, ..., G_m^t, ..., G_m^n$ has bounded differences. For that, let us first introduce conditional probabilities and martingales. Conditional probabilities model the probability of a random variable conditioned on an event having a specific realization. A conditional probability $\mathbb{P}(Z = z|\mathcal{E})$ denotes the probability of the event $Z = z$ given that event $\mathcal{E}$ happens. It is defined as $\mathbb{P}(Z = z|\mathcal{E}) := \mathbb{P}(Z = z, \mathcal{E})/\mathbb{P}(\mathcal{E})$. The conditional expectation is defined similarly to usual expectations as $\mathbb{E}[Z|\mathcal{E}] := \sum_{z\in\mathrm{domain}(Z)} \mathbb{P}(Z = z|\mathcal{E})$.

Conditional probabilities enable us to model the evolution of a random variable of interest. For example, a random variable $Z = \deg(k, G_m^n)$ can be investigated for its evolution by considering the sequence of random variables $X_0, \ldots, X_n$ with $X_t := \mathbb{E}[Z|G_m^t]$ where the expectation goes over the domain of all remaining graphs $G^t, \ldots, G_m^n$ in the generation process. Note that, if we used a conditioned event $\mathcal{E} = \{G_m^n = g_m^n\}$, which denotes the event that the graph G_m^n takes the realization g_m^n (for a concrete graph g_m^n), then X_t were numbers as the expectation would remove the randomness. Since we want to make general statements, we instead condition on the event without fixing a concrete graph g_m^n - instead we treat g_m^n as a random variable itself. This leads to X_t becoming a random variable of interest. Since both random graphs g_m^n and G_m^n are going to be equal, we simply use G_m^n in the conditioned event directly.

Given this definition, in the scope of this thesis, a martingale is a sequence of random variables $X_1, ..., X_t$ with the property $\mathbb{E}[X_{t+1}|G_m^1, ..., G_m^t] = X_t$ - i.e. in expectation the future is equal to the current value.[1]

Lemma 3.3 (Azumas Inequality) *Let $(X_t)_{t=1}^n$ be a martingale with $|X_t - X_{t-1}| \leq h(t)$ for $2 < t \leq n$. Then the probability of the deviation of X_n from X_1 exceeding $a(n)$ is given by the following inequality:*

$$\mathbb{P}(|X_n - X_1| \geq a(n)) \leq \exp\left(-\frac{a(n)^2}{2\sum_{i=0}^{n} h(i)^2}\right)$$

In the context of this thesis, we are going to be using $X_t = \mathbb{E}[f(G_m^1, ..., G_m^n)|G_m^t]$ where $f(G_m^1, .., G_m^n)$ expresses the random varible of interest as a function of the graphs. Such a sequence of random variables is a martingale (as shown in Prop. 5.3. [22]), and therefore we can apply Azumas Inequality to derive a concentration. For instance using $f(G_m^1, ..., G_m^n) = \deg(k, G_m^n)$ for $1 \leq k \leq n$ we can observe the evolution of the expected degree of the k-th node throughout the graph creation process. Note that, by definition we have $X_1 = \mathbb{E}[\deg(k, G_m^n)]$ and $X_n = \deg(k, G_m^n)$. With a bound on $|X_t - X_{t-1}|$ Azumas Inequality can be applied to derive a concentration. This mathematical tool was used in the prevalent paper [9] to prove the d^{-3} degree distribution for PA graphs.

3.2 Direct model

In this section, we a a summary of our results obtained from working directly on the model 2.1. This section shows the earliest approaches that uses tools and inspiration from [9] which allow us to make statements integrating the inter-dependencies between the random decisions that make mathematical results more difficult usually. The section starts with a results regarding the degrees of an arbitrary node as well as the cumulative degree of the first $z \in [n]$ nodes. We need the former to prove the latter. We are interested in the latter, as the first z nodes are a natural candidate for a minimum dynamic monopoly. They have the highest degrees and thus the strongest influence in the network. To calculate how likely they can make a node black, we need will want to compute the fraction of edges of node $k \in [n]$ that connect to $[z]$. In the second part of this section, we will describe how we tried to to compute this fraction - but came across difficulties and increasing complexity. The second part is going to be less formal and rather a description of our approach with ideas on how to overcome them. They may be relevant in a full investigation for either continuing this approach or trying a different one.

[1] Strictly, speaking $(X_t)_{t=1}^n$ is a martingale with respect to the sequence of random graphs $(G_m^t)_{t=1}^n$.

3.2.1 Degree of node k

The following lemma gives an upper and lower bound on $\mathbb{E}[\deg(k, G_m^n)]$.

Lemma 3.4 *Let G_m^n denote the random graph obtained from definition 2.1. The expected degree of the k-th node ($k \in [n]$) in the final graph G_m^n is bounded as follows:*

$$m\sqrt{\frac{n}{k}} \leq \mathbb{E}[\deg(k, G_m^n)] \leq 2m\sqrt{\frac{n}{k-1+m^{-\frac{1}{2}}}}$$

.

Proof We begin by establishing an exact formula for the expectation $\mathbb{E}[\deg(k, G_1^{nm})]$ for graphs of the form G_1^N with $N = nm$ and then use the a correspondence $G_1^{nm} \leftrightarrow G_m^n$ to derive the results for $\mathbb{E}[\deg(k, G_m^n)]$. The beginning of this proof until 3.1 is similarly derived in [9]. Consider the expected degree of a node t which just arrived in G_1^t for some $t \in [N]$. We have $\mathbb{E}[\deg(t, G_1^t)] = 1 + \frac{1}{2t-1}$, because there is a $\frac{1}{2t-1}$ probability of the half-edge leaving t to hit itself - hence increasing the expectation. Furthermore, given the degree of a node $k \in [N]$ in G_1^{t-1}, we can calculate the expectation in G_1^t. With probability $\frac{\deg(k,G^{t-1})}{2t-1}$ the edge from node t hits k and increases it's degree by 1. Thus, for $t > k$ we have

$$\begin{aligned}
\mathbb{E}[\deg(k, G_1^t) \mid \deg(k, G_1^{t-1})] &= \frac{\deg(k, G_1^{t-1})}{2t-1}(\deg(k, G_1^{t-1}) + 1) \\
&\quad + \frac{2t-1-\deg(k, G_1^{t-1})}{2t-1}\deg(k, G_1^{t-1}) \\
&= \deg(k, G_1^{t-1}) + \frac{\deg(k, G_1^{t-1})}{2t-1} \\
&= \frac{2t}{2t-1}\deg(k, G_1^{t-1}).
\end{aligned}$$

Applying $\mathbb{E}[\cdot]$ on both sides we get then the following recurrence relation:

$$\mathbb{E}[d_{G_1^t}(k)] = \begin{cases} \frac{2t}{2t-1} & \text{if } t = k \\ \frac{2t}{2t-1}\mathbb{E}[d_{G_1^{t-1}}(k)] & \text{if } t > k \end{cases}$$

From that, we get a product form for the final graph G_1^N:

$$\begin{aligned}
\mathbb{E}[\deg(k, G_1^N)] &= \prod_{t=k}^{N} \frac{2t}{2t-1} \\
&= \frac{4^{N-k+1}N!^2(2k-1)!}{(2N)!(k-1)!^2} && \text{([9])} \quad (3.1) \\
&= \sqrt{\frac{N}{k}}\frac{2k}{2k-1} && \text{(Stirlings formula)} \quad (3.2)
\end{aligned}$$

Note, that $\frac{2k}{2k-1} \in (1, 2]$. Using this, we can now turn to G_m^n for general m. [9] established a useful correspondence between G_1^{nm} and G_m^n that is equivalent with our definition 2.1. Concretely, a graph G_m^n can be created from the graph G_1^N by merging consecutive ranges of m nodes. Concretely, for all $0 \leq i < n$ the nodes $\{im+1, im+2, \ldots, im+m\}$ in G_1^N are merged to form node $i+1$ in G_m^n. Using this correspondence, we have

$$\mathbb{E}[\deg(k, G_m^n)] = E\left[\sum_{j=1}^{m} \deg((k-1)m + j, G_1^{nm})\right] \qquad (G_1^{nm} \leftrightarrow G_m^n)$$

since the k-th node in G_m^n, is created by merging m-nodes $\{(k-1)m + j \mid 1 \leq j \leq m\}$ in G_1^{nm}. Furthermore, with $f(i) := i^{-1/2}\frac{2i}{2i-1}$ we have

$$\begin{aligned}
&\mathbb{E}[\deg(k, G_m^n)] \\
&= \sum_{j=1}^{m} \mathbb{E}[\deg((k-1)m + j, G_1^{nm})] \\
&= \sqrt{nm}\sum_{j=1}^{m}((k-1)m+j)^{-1/2}\frac{2((k-1)m+j)}{2((k-1)m+j)-1} \quad \text{(from 3.2)} \\
&= \sqrt{nm}\sum_{j=1}^{m} f((k-1)m+j) \qquad \text{(def. } f(\cdot)).
\end{aligned}$$

Since $f(\cdot)$ is a decreasing function we lower (upper) bound the sum by simply replacing each term by the minimum at $(k-1)m + m = km$ (maximum at $(k-1)m+1$) and using $1 < \frac{2i}{2i-1} \leq 2$. Hence, we obtain the bounds:

$$\begin{aligned}
\sqrt{nm}\,\frac{m}{\sqrt{km}} &\leq \mathbb{E}[d(k)] \leq \sqrt{nm}\,\frac{2m}{\sqrt{(k-1)m+1}} \\
\Rightarrow m\sqrt{\frac{n}{k}} &\leq \mathbb{E}[d(k)] \leq 2m\sqrt{\frac{n}{k-1+m^{-\frac{1}{2}}}}
\end{aligned} \qquad \square$$

3.2.2 Degree of the first z-nodes

This section will derive the expectation of and a concentration result on the degree of the first z-nodes in a PA graph G_m^n. Let $\deg(S, G) = \sum_{s \in S} \deg(s, G)$ denote the degree over sets of nodes. Then the lemma on the expectation states:

Lemma 3.5 *Let G_m^n be a graph from the definition 2.1 and let $z \in [n]$ The expectation of the sum of degrees of the first z nodes in G_m^n is bounded as follows:*

$$1.5m\sqrt{n}(\sqrt{z} - 0.5) \leq \mathbb{E}[\deg([z], G_m^n)] \leq 4m\sqrt{n}(\sqrt{z} + m^{\frac{1}{4}})$$

Thus:

$$\mathbb{E}[\deg([z], G_m^n)] \in \Theta\left(m\sqrt{nz}\right).$$

Proof We have $\mathbb{E}[\deg([z], G_m^n)] = \sum_{j=1}^{z} \mathbb{E}[\deg(j, G_m^n)]$ by definition. Using Lemma 3.4, with $c := m^{-\frac{1}{2}}$, we get an expression for node j and use Lemma 3.1 to bound the sum:

$$\begin{aligned}
\sum_{j=1}^{z} m\sqrt{\frac{n}{j}} \quad & \leq \mathbb{E}[\deg([z], G_m^n)] \leq \sum_{j=1}^{z} 2m\sqrt{\frac{n}{j-1+m^{-\frac{1}{2}}}} \\
\Rightarrow m\sqrt{n}\sum_{j=1}^{z} j^{-1/2} \quad & \leq \mathbb{E}[\deg([z], G_m^n)] \leq 2m\sqrt{n}\sum_{j=c}^{z-1+c} j^{-1/2} \\
\Rightarrow m\sqrt{n}\int_{1}^{z+1} x^{-1/2}\mathrm{d}x \quad & \leq \mathbb{E}[\deg([z], G_m^n)] \leq 2m\sqrt{t}(c^{-\frac{1}{2}} + \int_{c}^{z-1+c} x^{-1/2}\mathrm{d}x) \\
\Rightarrow 2m\sqrt{n}(\sqrt{z+1} - \sqrt{1}) & \leq \mathbb{E}[\deg([z], G_m^n)] \leq 4m\sqrt{n}(m^{\frac{1}{4}} + (\sqrt{z-1+c} - \sqrt{c})) \\
\Rightarrow 1.5m\sqrt{n}(\sqrt{z} - 0.5) \quad & \leq \mathbb{E}[\deg([z], G_m^n)] \leq 4m\sqrt{n}(\sqrt{z} + m^{\frac{1}{4}})
\end{aligned}$$

In the last line, we use $\sqrt{a+b} \leq \sqrt{a} + \sqrt{b}$ and $2(\sqrt{z+1} - 1) \geq \frac{3}{2}(\sqrt{z} - \frac{1}{2})$. In the limit, we thus get $\mathbb{E}[\deg([z], G_m^n)] \in \Theta\left(\sqrt{nz}\right)$. □

Having the expectation, we can use Azumas Inequality to derive a concentration result.

Lemma 3.6 *Let G_m^n be a random graph from definition 2.1 and let $z \in [n]$ and $f(n) = 4m\sqrt{cz}\ln n$ for some $c \in \mathbb{N}$, let . with high probability $(1 - n^{-c})$ the sum of the degrees of the first z nodes lies within $f(n)$ of its expectation - i.e.*

$$0.75m\sqrt{z}(\sqrt{n} - 11.5m^{1.25}\log n) \leq \deg([z], G_m^n) \leq 4m^{1.25}\sqrt{z}(\sqrt{n} + 4.2m\log n).$$

Hence the deviation of the sum of the degrees are bounded by $\mathcal{O}(\log n/\sqrt{n})$ - i.e. $\deg([z], G_m^n) \in \Theta\left(\sqrt{nz}\left(1 \pm \frac{\log n}{\sqrt{n}}\right)\right)$

Proof We prove the concentration using Azumas Inequality (3.3) on the sequence of random variables $(X_t)_{t=1}^{n}$ with $X_t = \mathbb{E}[\deg([z], G_m^n)|G_m^t]$ by considering the evolution throughout the graph generation process. Since $\deg([z], G_m^n)$ is a function of G_m^n, X_t is a martingale as explained in the Preliminaries. Note that, from $X_1 = \mathbb{E}[\deg([z], G_m^n)]$, $X_n = \deg([z], G_m^n)$ and a bound on the increase of the sum of the degrees $|X_{t+1} - X_t|$ the result will follow.

We are want to establish a tight bound $h(t)$ where $|X_{t+1} - X_t| \leq h(t)$ for all $t \in [n]$. From the definition, we have

$$|X_{t+1} - X_t| \leq \mathbb{E}[\deg([z], G_m^n) \mid G_m^{t+1}] - \mathbb{E}[\deg([z], G_m^n) \mid G_m^t] = (*)$$

This difference denotes the increase of the expectation of the eventual degree $\deg([z], G_m^n)$ (when the graph is fully generated) when node $t+1$ is added to the graph. For that, let $X_{[m]}^{t+1} = (X_1^{t+1}, \ldots, X_m^{t+1})$ denote the random variables corresponding to the random choices of the m edges of node $t+1$ - so that we have event equality as $\mathcal{E} = \{G_m^{t+1}\} = \{X_{[m]}^{t+1}, G_m^t\}$. From that we conclude

$$\begin{aligned}
(*) &= \mathbb{E}[\deg([z], G_m^n) \mid X_{[m]}^{t+1}, G_m^t] - \mathbb{E}[\deg([z], G_m^n) \mid G_m^t] \qquad \text{(event equality)} \\
&= \mathbb{E}[\ \mathbb{E}[\deg([z], G_m^n) \mid X_{[m]}^{t+1}] \mid G_m^t] - \mathbb{E}[\deg([z], G_m^n) \mid G_m^t] \\
&= \mathbb{E}[\ \mathbb{E}[\deg([z], G_m^n) \mid X_{[m]}^{t+1}] - \deg([z], G_m^n) \mid G_m^t]
\end{aligned}$$

where the inner $\mathbb{E}[\cdot]$ goes over the domain of $X_{[m]}^{t+1}$. Intuitively speaking, the increase in expectation is upper bounded the expected number of edges of node $t+1$ that will connect to $[z]$. $[z]$.[2] Let $H(t+1, [z], G_m^{t+1})$ be the random variable denoting the number of edges of node $t+1$ that connected to any node in $[z]$ in the graph G_m^{t+1}. We hence have

$$|X_{t+1} - X_t| \le \mathbb{E}[H(t+1, [z], G_m^{t+1}) \mid G_m^t].$$

Next, we can decompose the random variable into the contributions from each of the edges and subsequently sum over an upper bound to get an expression. So, first we get

$$\begin{aligned}
&E[H(t+1, [z], G_m^{t+1}) \mid G_m^t] \\
&= \sum_{i=1}^{m} \mathbb{E}[\mathbb{1}(i\text{-th edge leaving } t+1 \text{ in } G_m^{t+1} \text{ connects to } [z]) \mid G_m^t] \qquad (3.3)
\end{aligned}$$

- and from $\mathbb{E}[\mathbb{1}(c)] = \mathbb{P}(c)$ follows $\mathbb{E}[\mathbb{1}(c)] = \mathbb{E}[\mathbb{E}[\mathbb{1}(c)]] = \mathbb{E}[\mathbb{P}(c)]$ (by applying $\mathbb{E}[\cdot]$ on both sides) to obtain

$$\begin{aligned}
&\sum_{i=1}^{m} \mathbb{E}[\mathbb{1}(i\text{-th edge leaving } t+1 \text{ in } G_m^{t+1} \text{ connects to } [z]) \mid G_m^t] \\
=&\mathbb{E}[\sum_{i=1}^{m} \mathbb{P}(i\text{-th edge leaving } t+1 \text{ in } G_m^{t+1} \text{ connects to } [z] \mid G_m^t)].
\end{aligned}$$

For the first edge leaving $t+1$ we have $\frac{\deg([z], G_m^t)}{2(t+1)m-1}$ as the probability to connect to $[z]$. The probability of the last edge to connect to $[z]$ is upper bounded by $\frac{\deg([z], G_m^t)+m-1}{2(t+1)m-1}$ as at-most $m-1$ edges have been added to $[z]$ beforehand.

[2]While this is intuitive in the point of view of the author, he tried his best in expressing this intuition in the above derivation. However, there is a risk, that the statement is wrong due to the authors limited experience with such statistical tools and false intuition that may follow from that.

This probability is also the upper bound for each of the m edges leaving $t+1$, thus we have

$$\mathbb{P}(i\text{-th edge leaving } t+1 \text{ in } G_m^{t+1} \text{ connects to } [z] \mid G_m^t) \leq \frac{\deg([z], G_m^t) + m - 1}{2(t+1)m - 1}.$$

Furthermore, note that $\deg([z], G_m^t) + m - 1 \leq 2\deg([z], G_m^t)$ since $\forall k.\ \deg(k, G_m^t) \geq m$ as each node sends m outgoing edges.

Bringing it all together, we then get a bound $h(t)$ we can work with.

$$\begin{aligned}
|X_{t+1} - X_t| &\leq \mathbb{E}[\textstyle\sum_{i=1}^m \mathbb{P}(i\text{-th edge leaving } t+1 \text{ hits } [r] \mid G_m^t)] && (3.4)\\
&\leq \mathbb{E}[m \frac{\deg([z], G_m^t) + m - 1}{2(t+1)m - 1}] \\
&\leq 2m \frac{\mathbb{E}[\deg([z], G_m^t)]}{2(t+1)m - 1} \\
&\leq 4m^2 \frac{\sqrt{t}(\sqrt{z} + m^{\frac{1}{4}})}{(t+1)m - 1} && \text{(Lemma 3.5)}\\
&\leq 4m^{2.25} \frac{\sqrt{tz}}{(t+1)m - 1} && (m^{2.25}\sqrt{z} \geq m^2(\sqrt{z} + m^{\frac{1}{4}}))\\
&:= h(t)
\end{aligned}$$

To apply this bound, we need to first upper bound $\sum_{t=1}^n h(t)^2$ as that is the denominator in the exponential in Azumas Inequality (Lemma 3.3). We have

$$\begin{aligned}
\sum_{i=1}^n h(i)^2 &\leq 4^2 m^{4.5} z \sum_{t=1}^n \frac{t}{(tm + m - 1)^2} \\
&\leq 4^2 m^{4.5} z \left(0 + \int_{t=1}^n \frac{t}{(tm + m - 1)^2}\right) && \text{(Lemma 3.1)}\\
&\leq 4^2 m^{4.5} z \left[\frac{m-1}{m^2(tm + m - 1)} + \frac{\log(tm + m - 1)}{m^2}\right]_{t=1}^{t=n} && \text{(integration tool)}\\
&\leq 17 m^{4.5} z \left(\frac{m-1}{m^2(nm + m - 1)} + \frac{\log(nm + m - 1)}{m^2}\right) \\
&\leq 17 m^{4.5} z \left(\frac{1}{2(n+1)m^2} + \frac{\log((n+1)m)}{m^2}\right) && (3.5)
\end{aligned}$$

Now we are ready to find the deviation $a(n)$ that is allowed by the bound $h(\cdot)$. For Azumas Inequality (Lemma 3.3) we set $\mathbb{P}(|X_n - X_1| \geq a(n)) \leq$

$\exp\left(-\frac{a^2(n)}{2\sum_{t=1}^{n} j(t)}\right) \le n^{-2}$ and conclude the form of $a(n)$:

$$
\begin{aligned}
& n^{-2} \ge \exp\left(-\frac{a^2(n)}{2\sum_{t=1}^{n} h(t)}\right) \\
\Rightarrow \quad & \frac{a^2(n)}{2\sum_{t=1}^{n} h(t)} \ge 2\log n \\
\Rightarrow \quad & a^2(n) \ge 34m^{4.5}z\left(\frac{1}{2(n+1)m^2} + \frac{\log((n+1)m)}{m^2}\right)\log n && \text{(from 3.5)} \\
\Rightarrow \quad & a^2(n) \ge 68m^{4.5}z\log^2 n && \text{(simplification)} \\
\Rightarrow \quad & a(n) = 8.5m^{2.25}\sqrt{z}\log n
\end{aligned}
$$

Putting this together: From $X_n = \deg([z], G_m^n)$, $X_1 = \mathbb{E}[\deg([z], G_m^n)]$ and Lemma 3.6, we conclude that the degree of the first z nodes is with $a(n)$ of its expectation. In other words, we have

$$0.75m\sqrt{nz} - 8.5m^{2.25}\sqrt{z}\log n \le \deg([z], G_m^n) \le 4m^{1.25}\sqrt{nz} + 8.5m^{2.25}\sqrt{z}\log n$$

and equivalently

$$0.75m\sqrt{z}(\sqrt{n} - 11.5m^{1.25}\log n) \le \deg([z], G_m^n) \le 4m^{1.25}\sqrt{z}(\sqrt{n} + 4.2m\log n)$$

which is our desired result. □

3.2.3 Influence of the first z nodes on a node k

We can now turn our attention to investigate the fraction of edges of a node k that connect with z. This gives us a first impression of the intensity of influence per node. High degree early nodes need many black neighbors for them to become black - whereas late low degree nodes can become black easily. However, high degree nodes equally can make many more other nodes black (if α is small), whereas one needs many low degree nodes to achieve the same effect. This is an interesting situation regarding the balance - and that is our interest. As mentioned previously, this section presents our approach and ideas rather than concrete results.

To get started, we would like to give a statement on the expected fraction of edges. Let $H(k, [z], G_m^n)$ be the random variable denoting the number of edges from node k connecting to a node in $[z]$ in the random graph G_m^n. We would like to give compute or give bounds for

$$\mathbb{E}\left[\frac{H(k, [z], G_m^n)}{\deg(k, G_m^n)}\right]$$

. With the expectation calculated, the concentration result follows from Azumas Inequality like in the previous proof. To compute the expectation, we try a similar recursive approach as in Lemma 3.4 for the proof of the degree of a node k. We will first begin with the base case $\mathbb{E}[H(k,[z],G_m^t)/\deg(k,G_m^t)]$ for $t = k$ and then turn to the recursive case

$$\mathbb{E}\left[H(k,[z],G_m^t)/\deg(k,G_m^t) \mid H(k,[z],G_m^{t-1})/\deg(k,G_m^{t-1})\right]$$

for $t > k$.

Base Case The base case can be bounded easily with a similar approach as was used to count the expected number of edges connecting to z in the previous proof (Equations 3.3 and 3.4). We only need to bound $\mathbb{E}[H(k,[z],G_m^t)]$ to bound $\mathbb{E}[H(k,[z],G_m^t)/\deg(k,G_m^t)]$ because $m \leq \deg(k,G_m^t) \leq 2m$ since $t = k$ as every node sends out m edges in the iteration when they arrive. Using the same argumentation as used in Equations 3.3 and 3.4, we conclude that

$$\mathbb{E}[H(t,[z],G_m^t)] = \mathbb{E}\left[\textstyle\sum_{i=1}^m \mathbb{P}(i\text{-th edge leaving } t \text{ in } G_m^t \text{ connects to } [z])\right]$$

Carrying over from the previous proof, we get the upper bound as

$$\mathbb{E}[H(t,[z],G_m^t)] \leq 4m^{2.25}\frac{\sqrt{tz}}{tm-1}.$$

Regarding the lower bound, consider the fact, that the connection probability is lower bounded by taking the largest denominator - that is $2(t+1)m-1$ when all m edges have been added - and the smallest denominator - that is $\deg([z],G_m^{t-1})$ when no edges of t had been added. Hence, we conclude

$$\begin{aligned}
\mathbb{E}[H(t,[z],G_m^t)] &= \mathbb{E}\left[\textstyle\sum_{i=1}^m \mathbb{P}(i\text{-th edge leaving } t \text{ in } G_m^t \text{ connects to } [z])\right] \\
&\geq \mathbb{E}[m\frac{\deg([z],G_m^t)}{2(t+1)m-1}] \\
&\geq \frac{\mathbb{E}[\deg([z],G_m^t)]}{2(t+1)} \\
&\geq \frac{1.5\sqrt{t}(\sqrt{z}-0.5)}{2(t+1)} && \text{(Lemma 3.6)} \\
&\geq \frac{0.75m\sqrt{tz}}{2(t+1)}
\end{aligned}$$

Recursive Case With the base case bounded, we would like to give an expression for the recursive case

$$\mathbb{E}\left[H(k,[z],G_m^t)/\deg(k,G_m^t) \mid H(k,[z],G_m^{t-1})/\deg(k,G_m^{t-1})\right] = (*)$$

for $t > k$. Intuitively speaking[3], the expectation denotes the change in the fraction when node t arrives. Every node t that arrives after k can only increase the denominator by connecting itself to k. Hence, the above should be equal to

$$(*) = H(k, [z], G_m^{t-1}) / (\deg(k, G_m^{t-1}) + C_{t,k})$$

where $C_{t,k}$ is the random variable that counts the number of edges from node t to k. Essentially, we have an initial positive contribution to the fraction then node $t = k$ arrives. After that, in each iteration the expected fraction decreases due to additional edges from nodes $t > k$. Hence, after applying $\mathbb{E}[\cdot]$ on both sides, we have

$$\mathbb{E}[H(k, [z], G_m^t) / \deg(k, G_m^t)] = \mathbb{E}\left[\frac{H(k, [z], G_m^{t-1})}{\deg(k, G_m^{t-1}) + C_{t,k}}\right].$$

This is the point, where the statistical dependencies make it difficult to progress. Ideally, we would like to apply the expectation on $C_{t,k}$, as it can be computed (or bounded) as a fraction containing $\deg(k, G_m^{t-1})$ from previous calculations. Essentially, we have $\mathbb{E}[C_{t,k}] \propto (\deg(k, G_m^{t-1})/t)$. At this point, we are unclear on whether this quantity can be further expressed in a meaningful way - as the statistical dependencies make progress difficult. If we work with the approximation $\mathbb{E}[C_{t,k}]$ instead of $C_{t,k}$ in the denominator itself - then we could conclude

$$\begin{aligned}
&\mathbb{E}[H(k, [z], G_m^t) / \deg(k, G_m^t))\\
&\approx \mathbb{E}\left[\frac{H(k, [z], G_m^{t-1})}{\deg(k, G_m^{t-1}) + \mathbb{E}[C_{t,k}]}\right]\\
&= \mathbb{E}\left[\frac{H(k, [z], G_m^{t-1})}{\deg(k, G_m^{t-1}) + \Theta(\deg(k, G_m^{t-1})/t)}\right]\\
&= \mathbb{E}\left[\frac{H(k, [z], G_m^{t-1})}{\deg(k, G_m^{t-1}) + c\deg(k, G_m^{t-1})/t}\right] && \text{for some constant } c\\
&= \mathbb{E}\left[\frac{H(k, [z], G_m^{t-1})}{\deg(k, G_m^{t-1})(1 + c/t)}\right]\\
&= \mathbb{E}\left[\frac{H(k, [z], G_m^{t-1})}{\deg(k, G_m^{t-1})}\right]\frac{1}{1 + c/t}
\end{aligned}$$

[3] This argument is intuitive in the point of view of the author - however, he was unable to make the argument more concrete. Hence the conclusion is at the risk of being incorrect or incomplete due to the authors limited experience with these statistical tools.

This would be very useful, as it leads to a product form of the expected fraction

$$\mathbb{E}[\mathbb{E}[H(k,[z],G_m^n)/\deg(k,G_m^n))] = \mathbb{E}[H(k,[z],G_m^t)/\deg(k,G_m^t)] \prod_{t=k+1}^{t} \frac{1}{1+c/t}.$$

The base case is expressed using our previous calculations and the product can be evaluated by using $\prod(\cdot) = \exp(\log \sum(\cdot))$ and bounding the sum using Lemma 3.1. We did not find any useful closed form like in the proof for the expected degree of node k (Lemma 3.4). We did reach some approximate answers however and in a full investigation one could use these results to get an intuition on the fraction of edges per node k to nodes $[z]$. However, the complexity and statistical dependencies make it harder to progress and to prove statements. Especially, it is not clear, how one might prove a concrete statement that avoids approximating $C_{t,k}$ by $\mathbb{E}[C_{t,k}]$ in the previous calculation. One possibility is to split the dependencies by making independent statements on the expectation of the number of edges of node k that connect to $[z]$. Concretely, by establishing $l(k,z,n,m) \leq \mathbb{E}[H(k,[z],G_m^n)] \leq u(k,z,n,m)$ for some bounds l,u, one could derive w.h.p. bounds on $H(k,[z],G_m^n)$. By equally deriving w.h.p. bounds on the degree of node k, we can overcome the statistical dependencies by focus on the concentration result. As both statements hold w.h.p, they are true w.h.p combined as well. Overcoming the statistical dependencies by externalizing concrete bounds into functions may be a fruitful approach. We will however not continue this approach as - after a while - we found other slightly simpler approaches that lead results more easily.

3.3 Computer Simulations

With the aforementioned difficulties with the previous approach, we used computer simulations to get better insights on the dynamics of α-BP on PA graphs. Having a few computer simulations and visualizations of the dynamics can greatly aid in building a intuition for the dynamics as well as giving a general overview of the situation. We will first present the setup of the simulation - that includes the algorithm for generating PA graphs, the choice of the node-set $\mathcal{B}_0$ as well as the parameters of the simulation. Afterwards, we will discuss the results and insights.

3.3.1 Setup, Algorithms and Visualization

The main motivation for the simulation was to experimentally find the size of the minimum dynamic monopoly for different ranges of α. An equally important goal was to trying out various approaches in algorithms to look for an interesting observation.

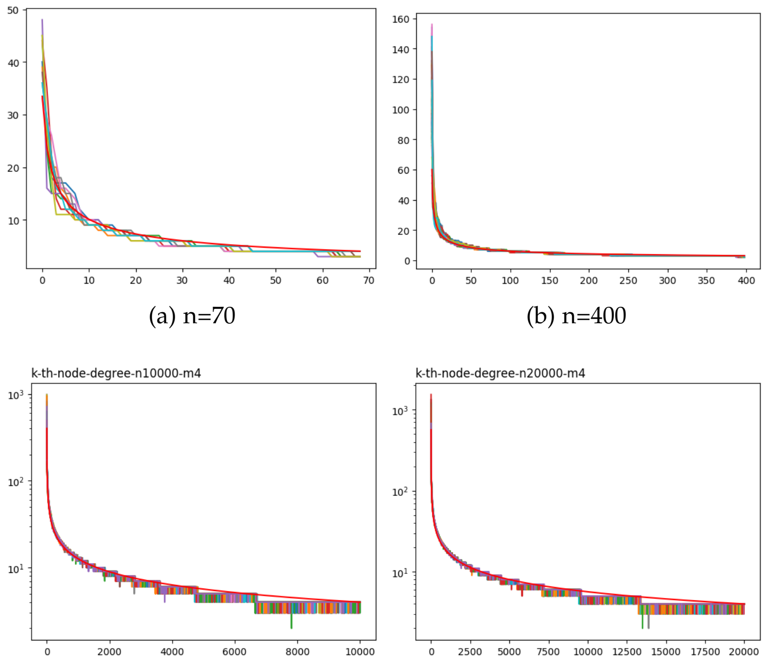
We started with an implementation of an algorithm for sampling graphs from the definition of PA graphs 2.1. [8] gives an efficient $\Theta(nm)$ time algorithm for sampling such graphs which was very useful as it can generate graphs with 10^5 nodes in seconds. We generated $N = 25$ graphs of various sizes $n \in [70, 400, 700, 1 \cdot 10^4, 1.5 \cdot 10^4, 2 \cdot 10^4, 2.5 \cdot 10^4, 3 \cdot 10^4]$ with $m = 4$. After sampling, all nodes in a graph were sorted degree with highest degree first. This way the most influential nodes were also the first ones.

Before implementing α-BP dynamics we first visualized a few relevant properties of the graphs. These included graphical representation of the graphs as well as plots of the degree distribution and the degrees of the (sorted) nodes. They were insightful and will be presented in the next section.

α-BP was implemented directly and as a first approach, we decided to test the first z nodes (in the *degree-sorted sequence*) for dynamic monopolies. Got get an overview of the situation, we wanted to find the smallest number $z(\alpha, n)$ of highest degree nodes, such that when only these nodes are black initially, then the entire graph will become black eventually under α-BP. For 80 equidistant $\alpha \in [0, 1]$ and 60 equidistant $z \in \{1, n\}$, we ran α-BP with the first z nodes (in the sorted-degree sequence) for all $N = 25$ graphs for $n \in [1 \cdot 10^4, 1.5 \cdot 10^4, 2 \cdot 10^4, 2.5 \cdot 10^4, 3 \cdot 10^4]$. For each run, the number of rounds a node needed to become black (if it became black at all) was plotted. For each n we also gathered all runs to see what the minimum number of black highest degree nodes is for it to become dynamic monopoly. Selecting the highest degree nodes may not yield the minimum dynamic monopoly - but it will probably be close to the minimum in many cases.

3.3.2 Results and Insights

Running the simulations was merely a matter of a few hours upto a day of computation time. Among the first findings is the accuracy of our result on the expected degree of the k-th node (Lemma 3.4, assuming the nodes are degree-sorted). Consider the Figure 3.1. The sorted degree sequence visualized for various n show a great resemblance to $\mathbb{E}[\deg(i, G_m^n)] \approx m\sqrt{n/i}$. While the theorem is about the i-th arriving node, the plot shows the nodes in order of decreasing degree. The first ordering is often not the same as the latter ordering, as due to randomness and luck, a later node might very well have a higher degree than an earlier one. Furthermore, preferential attachment favors the higher degree node even more then. If nodes are not sorted then the plots have shown to have a high variance - i.e. the resemblance between both vanishes quickly - and often even the first small fraction of nodes get unlucky and have a low degree in the end. It seems, that once a node is lucky or unlucky to get a higher or lower degree, then it takes simply the role of an earlier or later node in the graph. It might be very well possible to prove a variant of Lemma 3.4 that is not on the expectation, but on the degrees of a degree-sorted node sequence. One could prove that property by looking at the degree of each node during the generation of the graph and with the assumption that the k-th node has bounded degree. Once the degree of a node exceed its bounds, we could swap it with a low degree node to restore balance. This is an interesting result as it gives us a slightly different perspective on the nodes and their degrees.



(a) n=70 (b) n=400

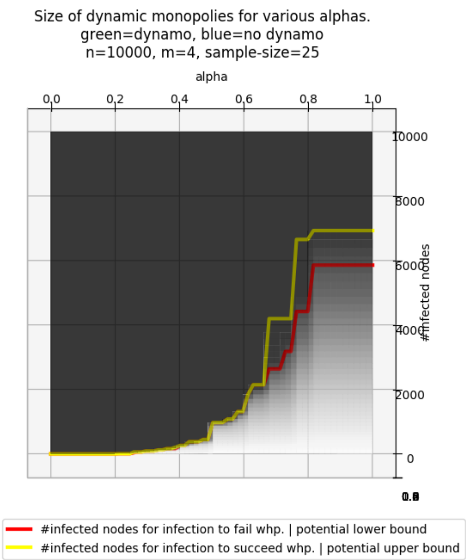
(c) n=$1 \cdot 10^4$ (d) n=$2 \cdot 10^4$

Figure 3.1: The x-axis denotes the k-th node having the k-highest degree. The y-axis is the degree of the k-th node in linear scale (a,b) and logarithmic scale (c,d). For each of the $N = 25$ graphs a separate color was used to plots its nodes degree. The red line is a plot of the function $m\sqrt{n/k}$.

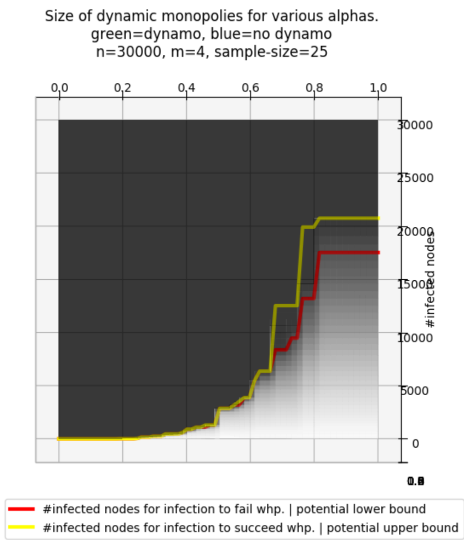
The results from the simulations were very useful in getting an overview of the situation. Consider Figure 3.2. We can immediately see, that the choice of α greatly influence the size z of highest-degree nodes that are necessary for a dynamic monopoly. The yellow line displays the smallest size z for which all nodes became black in all $N = 25$ graphs (for a concrete n). The size starts out small for small α and grows quickly bigger once it comes close to $\alpha = 1/2$. After that, it increases step-wise with two noticeable jumps at $z \approx 41\%n$ and $z \approx 70\%n$ and plateaus at the second jump. The size of a dynamic monopoly seems to be very small for small $\alpha \leq \frac{1}{m} = 0.25$. Sublinear z are intuitive and suspected in this domain. This is also consistent with prior work of r-BP on PA graphs. It was shown by [23], that r-BP making the first r nodes black is sufficient for a dynamo w.h.p.. For very small $\alpha \leq \max_k \deg(k, G_m^n) \approx \Theta(\frac{1}{m\sqrt{n}})$, when every node needs only a single black neighbor to become black, then α-BP can be reduced to r-BP with $r = 1$. Hence for very small α we have $z = 1$ as the minimum dynamic monopoly. On the other hand, for $\alpha \geq 1/2$ (and especially for large $\alpha \geq 1 - 1/m = 0.75$), the dynamic monopoly z seems to be linear in the size of n. This is supported by the fact, that *all* choices for n in these simulations revealed an almost identical picture as those shown in Figure 3.2. This is very intuitive, as for large $\alpha \geq 1/2$ each newly black node decreases the number of remaining edges that can be used to make other remaining nodes black. It seems, that we can roughly break down the picture into three regions of interest. The first region is $\alpha \leq \frac{1}{m}$. We expect very few nodes to be sufficient to make all nodes black - hence, one might want to establish an upper bound for the minimum size of a dynamic monopoly by explicit construction. The second region of interest is $\frac{1}{m} \leq \alpha \leq 1 - \frac{1}{m}$. The size of the minimum dynamic monopoly increases gradually to (suspected) linear size - and making that concrete would be interesting and insightful. However it might be more difficult to establish results here compared to the first and the third region. The third region of interest is $\alpha \geq 1 - \frac{1}{m}$. Here one might be able to derive lower bounds with low effort, as the large α makes it difficult for small sets to take over the graph. This may be established in a Lemma by investigating the number of edges between black and white nodes throughout the progress of the dynamics. As every newly black node will decrease this number of edges, eventually the process would suffocate as the number of remaining edges diminishes. It is also interesting to vary m and observe the change of the picture. As it is now, the α-BP dynamics for small degree nodes is strongly influenced by the size of m. Such a variation would also improve on the previous regions of interest or put them into another perspective. Another relevant factor is the choice of our first z highest degree nodes. Even though, finding the minimum size of a dynamic monopoly under α-BP is NP-complete[4], one might use different and better algorithms to decide which nodes to make black initially.

All in all, the simulations were useful in getting an overview of the situation and they one might take them as a reason to distinguish between regions of interest in the values of α regarding the minimum size of a dynamic monopoly and related questions.

[4]The size of the minimum dynamic monopoly for $\alpha = 1$ is a reformulation of the definition of a Vertex Cover Problem



(a) n=$1 \cdot 10^4$



(b) n=$3 \cdot 10^4$

Figure 3.2: The smallest number of nodes highest-degree nodes z that were necessary to be black to make the entire graph black for two choices of n and all α. The x-axis displays α while the y-axis denotes the number of nodes necessary for a dynamo (yellow line). The background color of a patch corresponds to the fraction of nodes that were black after once the dynamics settled down. The red line shows a potential lower bound on the dynamic monopoly if the choice of z highest-degree nodes were optimal.

3.4 Alternative model

In this section, we give a summary of our results obtained from using the alternative model 2.2 also give directions on how to continue the investigation using this model. We begin with a small demonstrative proof to show one can use the alternative model. After another Lemma for bounds on degrees, we are going to investigate the connectivity regarding the dynamics of α-BP. We start with prove the following Lemma that lower bounds the degree of the first node.

3.4.1 Degree bounds

Lemma 3.7 *Let G_m^n denote a graph sampled from the alternative definition 2.2 and let $n \geq 3 \cdot 10^{10}$. The degree of the first node is lower-bounded as*

$$\deg(1, G_m^n) \geq m \frac{20}{11} \frac{\sqrt{n}}{\log n} - 2\sqrt{mn}^{\frac{1}{4}}$$

with high probability $(1 - n^{-1})$.

Proof Consider the random variables $x_{i,j}, y_{i,j}, W_i$ and w_i for $i \in [n], j \in [m]$ from the definition 2.2. From lemma 2.3, we know, that $E_1, \ldots, E_5$ hold w.h.p. Thus w.h.p, from E_3 it follows that $w_1 \geq \frac{4}{\log n\sqrt{n}}$. We will now consider all variables $(x_{i,j})_{i,j}$ and count the number of incoming edges - which corresponds to $x_{i,j}$ being smaller than w_1. According to the definition, given $(y_{i,j})_{i,j}$ we can work with the simplification, that $x_{i,j} \sim \mathcal{U}(0, W_i)$ as the slight increase of the probability of a loop still leads to a lower bound on the degree of the first node. The W_i can w.h.p. be upper bounded as $W_i \leq \frac{11}{10}\sqrt{\frac{i}{n}}$ using E_1 for $i \geq s$ (lemma 2.3). Thus, putting it all together and using $y_{*,*}$ as a short form for $(y_{i,j})_{i,j}$, we have

$$\begin{aligned}
&\mathbb{P}(x_{i,j} \leq w_1 \mid y_{*,*}) \\
&\geq P\left(x_{i,j} \leq \frac{4}{\log n\sqrt{n}} \mid y_{*,*}\right) && \text{(from } E_3\text{)} \\
&\geq \begin{cases} 1 & \text{if } W_i \leq \frac{4}{\log n\sqrt{n}} \\ \frac{4}{\log n\sqrt{n}}/W_i & \text{if } W_i > \frac{4}{\log n\sqrt{n}} \end{cases} && (\text{using } x_{i,j} \sim \mathcal{U}(0, W_i)) \\
&\geq \min\left(\frac{4}{W_i \log n\sqrt{n}}, 1\right) \\
&\geq \min\left(\frac{4}{\frac{11}{10}\sqrt{\frac{i}{n}}\log n\sqrt{n}}, 1\right) && \text{(from } E_1\text{)} \\
&= g(i) && \left(\text{where } g(i) := \frac{40}{11}\frac{1}{\log n\sqrt{\max(i,s)}}\right)
\end{aligned}$$

for all $x_{i,j}$ where $i \geq s$. Even though E_1 only holds for nodes $i \geq s$ but not for very early nodes $i \in \{2, \ldots, s-1\}$, we can use $g(i)$ as a bound for them as well since $W_{i-1} < W_i$ for all i and $W_s \leq \frac{11}{10}\sqrt{\frac{s}{n}}$. To count the edges, let X be the random variable denoting the number of edges from nodes $i \in \{2, \ldots, n\}$ to the first node 1. We can decompose X into the contributions from each $(x_{i,j})_{i,j}$ as they are independent given $y_{*,*}$. Thus, we have

$$X = \sum_{2 \leq i \leq n, j \in [m]} X_{i,j}$$

as a lower bound where $X_{i,j} \sim \mathrm{Ber}(g(i))$. Since X is a the sum of independent Bernoulli random variables, we can use the Chernoff bound from lemma 3.2 to derive a lower bound. For that, we first lower bound the expectation of X as

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{2\le i\le n, j\in[m]} g(i) \\
&= m \sum_{2\le i\le n} g(i) \\
&= \frac{40}{11}\frac{m}{\log n} \sum_{2\le i\le n} \frac{1}{\sqrt{\max(i,s)}} && \text{(definition of } g(i)\text{)} \\
&= \frac{40}{11}\frac{m}{\log n} \left((s-2)s^{-1/2} + \sum_{s\le i\le n} i^{-1/2} \right) && \text{(split at } s\text{)}
\end{aligned}$$

Furthermore, bounding the sum using an integral we get a lower bound on the expectation

$$\begin{aligned}
\mathbb{E}[X] &\ge \frac{40}{11}\frac{m}{\log n} \left(\sqrt{s} + \int_s^{n+1} t^{-1/2} \mathrm{d}t - 2s^{-1/2} \right) && \text{(lemma 3.1)} \\
&= \frac{40}{11}\frac{m}{\log n} \left(\sqrt{s} + 2(\sqrt{n+1} - \sqrt{s}) - 2s^{-1/2} \right) \\
&= \frac{80}{11}\frac{m}{\log n} \left(\sqrt{n+1} - \sqrt{s} - 2 \right) && (s^{-1/2} \le 1) \\
&\ge \frac{40}{11} m \frac{\sqrt{n}}{\log n}. && \text{(for } n \ge 3\cdot 10^{10}\text{)} \quad (3.6)
\end{aligned}$$

Consider the probability that the degree is too small.

$$\begin{aligned}
\mathbb{P}(X \le \mathbb{E}[X] - t) &\le \exp\left(-\frac{t^2}{2\mathbb{E}[X]} \right) && \text{(Lemma 3.2)} \\
&\le \exp\left(-\frac{t^2}{2\frac{40}{11} m \frac{\sqrt{n}}{\log n}} \right) := h(n,t) && \text{(from 3.6)}
\end{aligned}$$

Setting a low probability - namely $h(n,t) \le n^{-c}$ for some $c > 0$ - yields

$$\begin{aligned}
& & h(n,t) &\le n^{-c} \\
&\Rightarrow & \exp\left(-\frac{t^2}{m}\frac{11}{80}\frac{\log n}{\sqrt{n}} \right) &\le n^{-c} \\
&\Rightarrow & -\frac{t^2}{m}\frac{11}{80}\frac{\log n}{\sqrt{n}} &\le -c\log n \\
&\Rightarrow & t &\ge \sqrt{cm\frac{80}{11}\sqrt{n}} \\
&\Rightarrow & t &\ge 2\sqrt{cm} n^{1/4}.
\end{aligned}$$

Setting $t = 2\sqrt{m}n^{1/4}$ for $c = 1$, we then conclude from the lower bound on $\mathbb{E}[X]$ that

$$\mathbb{P}\left(X \le \frac{40}{11}m\frac{\sqrt{n}}{\log n} - 2\sqrt{m}n^{1/4}\right) \le \mathbb{P}(X \le \mathbb{E}[X] - 2\sqrt{m}n^{1/4}) \le n^{-1}.$$

Thus, with high probability, the number of edges from nodes $i \in \{2, \dots, n\}$ is larger than $\frac{40}{11}m\frac{\sqrt{n}}{\log n} - 2\sqrt{m}n^{1/4}$. Adding the m self-edges that the first node sends to itself yields $\frac{40}{11}m\frac{\sqrt{n}}{\log n} + 2m$ for the first term - for which $m\frac{40}{11}(\frac{11}{20} + \frac{\sqrt{n}}{\log n})$ and then $m\frac{20}{11}\frac{\sqrt{n}}{\log n}$ are lower bounds as a quick calculation shows. □

Next, we are going to provide an useful lemma that is used in subsequent larger proofs. The lemma lower and upper bounds the degree of a node a bound on its weight.

Lemma 3.8 (Bounded Degree) *Let G_m^n be a graph from the definition 2.2 and $n \ge 10^{10}$. If a node $k \le \gamma n$, where $0 \le \gamma < 1$, has bounded weight w_k, then it has bounded degree with high probability. Concretely, if $w_k \le u(n)$ for some $u(n) \in \omega(\log n/n)$ then w.h.p.*

$$\deg(k, G_m^n) \le m\left(1 + \frac{40}{9}u(n)n(1 - \sqrt{(k-1)/n})\right)$$

Similarly if $w_k \ge l(n)$ for some $l(n) \in \omega(\log n/n)$ then w.h.p.

$$m\left(1 + \frac{10}{11}l(n)n(1 - \sqrt{k/n})\right) \le \deg(k, G_m^n).$$

Proof We will first establish a decomposition of the degree of node k into independent Bernoulli random variables (like in the previous proof) - which can then be easily bounded using lemma 3.2.

Again we consider the random variables $x_{i,j}, y_{i,j}, W_i$ and w_i for $i \in [n], j \in [m]$ as well as the graph $G_m^n = (V, E)$ from the definition 2.2. From the definition, the degree of node k is given by

$$\deg(k, G_m^n) = m + \sum_{i \in \{k,\dots,n\}} \mathrm{mult}(\{i,k\}, E)$$

where the first term corresponds to the m out-going edges of node k and the remaining sum corresponds to the contributions from all nodes that could connect with k. From $\mathrm{mult}(\{i,k\}, E) = |\{\ \{i,j\} \mid W_{k-1} < x_{i,j} < W_k \text{ for } j \in [m]\ \}|$, $x_{i,j} \sim \mathcal{U}(0, y_{i,j})$ (given $y_{*,*}$), $W_{i-1} \le y_{i,j} \le W_i$ and $w_i = W_i - W_{i-1}$ we can conclude bounds on the probability of a connection. Concretely, we have

$$
\begin{aligned}
\mathbb{P}(W_{k-1} < x_{i,j} < W_k | y_{*,*}) &= \mathbb{P}(x_{i,j} \leq W_k - W_{k-1} | y_{*,*}) \\
&= \mathbb{P}(x_{i,j} \leq w_k | y_{*,*}) \\
&= w_k / y_{i,j} & (x_{i,j} \sim \mathcal{U}(0, y_{i,j})) \\
&\geq w_k / W_i & (y_{i,j} \leq W_i) \\
&\geq l(n) / W_i & (w_k \geq l(n))
\end{aligned} \tag{3.7}
$$

and similarly

$$
\begin{aligned}
\mathbb{P}(W_{k-1} < x_{i,j} < W_k | y_{*,*}) &= w_k / y_{i,j} & (x_{i,j} \sim \mathcal{U}(0, y_{i,j})) \\
&\leq w_k / W_{i-1} & (y_{i,j} \geq W_{i-1}) \\
&\leq u(n) / W_{i-1} & (w_k \leq u(n))
\end{aligned} \tag{3.8}
$$

for all i, j. Hence we have for random variables $L_i \sim \mathrm{Ber}(l(n)/W_i)$, $U_i \sim \mathrm{Ber}(u(n)/W_{i-1})$, we can bound the degree of node k as follows:

$$
\sum_{i \in \{k, \ldots, n\}, j \in [m]} L_i \leq \deg(k, G_m^n) - m \leq \sum_{i \in \{k, \ldots, n\}, j \in [m]} U_i
$$

To derive high probability bounds, we proceed similarly as in the previous proof. With $D = \deg(k, G_m^n) - m$, we first bound the expectation as

$$
\begin{aligned}
\mathbb{E}[D] &\leq m \sum_{i \in \{k, \ldots, n\}} \mathbb{E}[U_i] \\
&\leq m \sum_{i \in \{k, \ldots, n\}} \frac{u}{W_{i-1}} & (\text{def. } \mathbb{E}) \\
&\leq m \sum_{i \in \{k, \ldots, n\}} u(n) \frac{10}{9} \sqrt{\frac{n}{i}} & (\text{w.h.p. } E_1) \\
&\leq \frac{10}{9} u(n) m \sqrt{n} \int_{k-1}^{n} t^{-1/2} \mathrm{d}t & (\text{lemma 3.1}) \\
&\leq \frac{20}{9} u(n) m \sqrt{n} (\sqrt{n} - \sqrt{k-1}) \\
&\leq \frac{20}{9} u(n) m n \left(1 - \sqrt{\frac{k-1}{n}}\right)
\end{aligned} \tag{3.9}
$$

for the upper bound and similarly

$$
\frac{20}{11} l(n) m n (1 - \sqrt{k/n}) \leq \mathbb{E}[D] \tag{3.10}
$$

for the lower bound.

Using Chernoff (lemma 3.2) we can bound the probability that the degree is smaller than a lower bound. With $t = \left(\frac{10}{11} l(n) mn (1 - \sqrt{k/n})\right)$, We have

$$\begin{aligned}
&\mathbb{P}\left(D \leq \frac{10}{11} l(n) mn (1 - \sqrt{k/n})\right) \\
&\leq \mathbb{P}\left(D \leq \mathbb{E}[D] - t\right) && \text{(from 3.10)} \\
&\leq \exp\left(-\frac{t^2}{2\mathbb{E}[D]}\right) && \text{(lemma 3.2)} \\
&\leq \exp\left(-\frac{10 l(n) mn (1 - \sqrt{k/n})}{22}\right), && \text{(from 3.10)} \\
&\leq \exp\left(-\frac{10 l(n) mn (1 - \sqrt{\gamma})}{22}\right), && (k \leq \gamma n)
\end{aligned}$$

which is in $\mathrm{o}(1)$ for $l(n) \in \omega(\log n/n)$. Similarly, with $t = \frac{20}{9} u(n) mn \left(1 - \sqrt{(k-1)/n}\right)$ for the upper bound we have

$$\begin{aligned}
&\mathbb{P}\left(D \geq \frac{40}{9} u(n) mn \left(1 - \sqrt{\frac{k-1}{n}}\right)\right) \\
&\leq \mathbb{P}(D \geq \mathbb{E}[D] + t) && (3.9) \\
&\leq \exp\left(-\frac{t^2}{2(\mathbb{E}[D] + t/3)}\right) && \text{(lemma 3.2)} \\
&\leq \exp\left(-\frac{20 u(n) mn (1 - \sqrt{\gamma} + \mathrm{o}(1))}{21}\right), && (3.9 \text{ and } k \leq \gamma n)
\end{aligned}$$

which is in $\mathrm{o}(1)$ for $u(n) \in \omega(\log n/n)$. Putting both together, we conclude following bounds on D that hold with probability $1 - \mathrm{o}(1)$:

$$\begin{aligned}
&\frac{10}{11} l(n) mn (1 - \sqrt{k/n}) \\
&\leq D = \deg(k, G_m^n) - m \\
&\leq \frac{40}{9} u(n) mn (1 - \sqrt{(k-1)/n})
\end{aligned}$$

These bounds are equivalent to the desired result. □

3.4.2 α-BP dynamics

With these results, we would like to get an understanding of how α-BP behaves over a few rounds on PA graphs. We will present some approximate and rough results on these dynamics by direct step-wise calculation. The general idea is, that we start with the first nodes $[z]$ as black and then investigate how many other nodes get black as well. We know most intermediate and early nodes beyond z - namely for example nodes $\{z+1, \ldots, n/2\}$ have high degree. Thus while trying to start with the smallest dynamic monopoly as possible - it is easier to make late nodes (such as $\{n/2, \ldots, n\}$) black - as they have low degrees on average - whereas early nodes are be much harder. As a first try, we will therefore investigate the number of late nodes which get black from $[z]$ in the first round. Another reason we tried late nodes first, was because event E_5 from lemma 2.3 already gives upper bounds on the weights of late nodes which can be directly translated into bounds on their degree.

Approach 1 Starting with black nodes $\mathcal{B}_0 = [z]$, we want to first calculate the cumulative weight that these nodes have in the alternative model. This can then be used to calculate how many edges later node connect to there initially black nodes. For that, we consider a graph G_m^n given by definition 2.2 as well as the random variables $x_{i,j}, y_{i,j}, W_i$ and w_i for $i \in [n], j \in [m]$. Since we take the first z nodes, we have cumulated weight W_z by definition. From E_5 we have that w.h.p. the nodes $i \in \{n/2, \ldots, \gamma n\}$ (for some $1/2 < \gamma < 1$) have bounded weight $w_i \leq \frac{\log^2 n}{n} \in \omega(\log n/n)$. From the lemma 3.8, we conclude, that these late nodes w.h.p. have a bounded degree

$$\deg(i, G_m^n) \leq m \left(1 + \log^2 n \frac{40(1 - 1/\sqrt{2})}{9}\right) := f(n, m) \tag{3.11}$$

for $i \in \{n/2, \ldots, n\}$ (by setting $k = n/2$ for the lemma). Next we want to lower bound the number of late nodes that get black - i.e. the number of nodes for which

$$H_i/f(n, m) \geq \alpha \tag{3.12}$$

for random variable H_i denoting the number of edges from node i to $\mathcal{B}_0 = [z]$. At this point, progress on this path of direct calculation became unfruitful. A big reason for difficulties was the fact, that the logarithmic term in 3.11 always increases the number of edges one would need to connect to $\mathcal{B}_0$ to make that node black. However, as we only have constantly many edges (namely m) which could connect to $\mathcal{B}_{\prime}$, we cannot prove[5], that there will be enough edges connecting to make that node black. For intermediate and early nodes the situation is even worse as they have even higher degree and thus cannot become black early on. Intuitively, speaking however, a certain fraction of late nodes is suspected to become black in the first round as some of them may have low degree and therefore their few edges to $\mathcal{B}_0$ can make

[5] for large n

them useful. This roadblock mainly demonstrates the difficulties of a direct attempt using E_5 to bound the edges of late nodes.

Approach 2 At this point, one could try to use a different bound on the degree of certain nodes to gain some insights and to calculate how many of them could get infected. Unfortunately, none of the remaining events $E_1, \ldots, E_4$ from [10] offer an upper bound on the weights to conclude a degree upper bound. However, in a step further, we can investigate the proofs of these events in their paper [10] and make slight adjustments to yield the necessary lower bounds. With these new bounds, one could try to continue similarly from the roadblock above.

While the proofs in the paper have a certain technical difficulty (from the point of view of the author), some parts of the proofs are rather isolated and easily understood. In their proofs for $E_1, \ldots, E_5$ they establish a few central statistical properties of the random variables which then get translated into these events in a few steps. Ideally, we would like to have a variant of

$$E_2 = \left\{ I_t \text{ contains at least } 2^{t-1} \text{ nodes } i \text{ with } w_i \geq \frac{1}{10\sqrt{in}} \text{ for } a \leq t < b \right\},$$

where $I_t = [2^t + 1, 2^{t+1}]$, because it focuses on the node weights w_i and gives specialized bounds depending on the node placement. This is in contrast to $E_3, \ldots, E_5$ which refer to the first, and early and late nodes separately. While they are technically easier to use initially, they might also be too general or generous to be useful in our situation. This is the problem we encountered in the previous approach.

E_2 has a special structure. It considers intervals of exponentially increasing sized intervals I_t and each interval has half of its nodes with weights bounded by a function dependent on $i \in I_t$. The dependency is useful, as it provides a finer structure in its statement. Fortunately, a slight adjustment in the proof of E_2 gives us a variant with an upper bound instead of a lower bound. This will is established in the next lemma. After that, we will return to our approach.

Lemma 3.9 *Let w_i, a, b for $i \in [n]$ be defined according to definition 2.2 as well as $n \geq 10^{10}$ and $I_t := [2^t + 1, 2^{t+1}]$. With high probability the following complementary event holds*

$$E_{2,compl} = \left\{ I_t \textit{ contains at least } 2^{t-2} \textit{ nodes with } i \textit{ with } w_i \leq \frac{1}{10\sqrt{in}} \textit{ for } a \leq t < b \right\}.$$

Proof We will rely on the following statistical statement in the proof of E_2 cited from [10]:

> "Restricting our attention to $i \in I_t$ for some $a \leq t < b$, it follows that
> $$Z_t = \left| \left\{ i \in I_t : w_i < \frac{1}{10\sqrt{in}} \right\} \right| \times \mathbb{I}_{E_1} \tag{3.13}$$
> is stochastically dominated by a [binomial] $\mathrm{Bi}(2^t, 9/20)$ distribution, ..."

Z_t counts the number of low weight nodes in a certain interval I_t. They prove E_2 (given E_1), by then applying lemma 3.2 on $\mathbb{P}(Z_t > 2^{t-1})$ and showing that a union bound over all $a \leq t < b$ still leads to a low probability ($o(1)$). From that E_2 follows. We can use the statement to also give a bound on a complementary event. Since w.h.p. E_1, for the rest of the proof, we condition on it. From 3.13, we have $Z_t \sim \mathrm{Bi}(2^t, 9/20)$. Now we consider the probability that there are too few low weight nodes and apply lemma 3.2:

$$\begin{aligned}
&\mathbb{P}\left(Z_t \leq 2^{t-2}\right) \\
&= \mathbb{P}\left(Z_t \leq \mathbb{E}[Z_t] - 2^t 16/36\right)\right) && \text{(def. } \mathbb{E}[\cdot] \text{ and eq. 3.13)} \\
&= \exp\left(-2^t 160/9\right) && \text{(lemma 3.2)}
\end{aligned}$$

Like in the paper, we proceed by showing, that this probability of a lower than bound weight is small overall when we make a union bound over $a \leq t < b$. We have

$$\mathbb{P}\left(E^c_{2,compl} | E_1\right) \leq \sum_{t=a}^{b-1} \exp\left(-2^{t-2}\right) \in o(1),$$

where $\cdot^c$ is set complement, for $a \to \infty$ as $n \to \infty$ using the same argument as in the paper [10]. $\square$

Having established $E_{2,compl}$ we can proceed our approach. We want to investigate which (late) nodes get black in the first round - and therefore we focus on intervals I_t for large t - i.e. $I_{t_1} = [\frac{10}{31}n + 1, \frac{20}{31}n]$ with $t_1 = \log_2(\frac{10}{31}n)$. Applying lemma 3.8, we can establish that the 2^{t-2} nodes i in the interval I_t with $w_i \leq 1/(10\sqrt{in})$ - denoted by $I_{t,bounded}$ - have bounded degrees as given by

$$\begin{aligned}
&\deg(i, G^n_m) \\
&\leq m\left(1 + \frac{4}{9}(\sqrt{n/i} - \sqrt{(i-1)/i})\right) \\
&\leq \frac{3}{2}m. && \text{(if } t = t_1\text{)}
\end{aligned}$$

For $t = t_1$, we fortunately get a constant bound which we can use in the next step. Note, that if we had instead chosen a sub-linear range of nodes i - i.e. $i \in o(n)$ when $t \in o(\log n)$ - then the upper bound would have reduced to the form $m(1 + \omega(1))$. This would have been a problem as we would get a non-constant upper bound for the degree leading to the same difficulties as in approach 1. In this sense, looking at nodes in the linear range is the only option. To keep things simple, we will consider the concrete example $t = t_1$. In a further investigation, one would look at a general linear range of the form $t_1 = \beta n$ for some constant $\beta < 1/3$. Since we have atmost $3m/2$ edges, a node $i \in I_{t_1}$ is gets black if atleast $\lceil 3\alpha m/2 \rceil$ of its edges connect to $\mathcal{B}_0$. For that let $J = \sum_i J_i$ be the random variable denoting the number of nodes in $I_{t_1,bounded}$ that get black in the first round where J_i is a binary random variable indicating whether node i got black. Using an analogous argument as in the proof of lemma 3.8 (specifically with from 3.7), we conclude that each edge of nodes $i \in I_{t_1,bounded}$ has probability of connecting to $\mathbb{B}_0$ lower bounded by W_z/W_i (given $y_{*,*}$). Therefore, we can calculate a lower bound on the probabilities by treating each edge as a $\text{Ber}(W_z/W_i)$ random variable. Hence, the number of edges per node i connecting to $\mathcal{B}_0$ is a binomial distribution lower bounded by $\text{Bin}(m, W_z/W_i)$. We could continue by calculating the probability of $J_i = 1$.

$$
\begin{aligned}
\mathbb{P}(J_i = 1 | y_{*,*}) &\geq \mathbb{P}(X_i \geq \lceil 3\alpha m/2 \rceil \, | y_{*,*}) \qquad\qquad (X_i \sim \text{Bin}(m, W_z/W_i)) \\
&\geq 1 - \mathbb{P}(X_i \leq \lceil 3\alpha m/2 \rceil \, | y_{*,*}) \qquad\qquad (3.14)
\end{aligned}
$$

However, to make things simpler and to progress further more easily - we proceed by focusing on approximations and ideas rather than rigorous results (due to the increasing complexity). In that sense, the rest of this section should be regarded as guidelines for further investigation.

Focusing on approximations, we can start by simplifying J. Concretely, we simply add together all the contributing edges to B_0 while not exactly inspecting, whether a node above α-fraction of edges. Note that, we have $\mathbb{E}[X_i] \propto \sqrt{z/n}$ which follows from the definition of $\mathbb{E}[\cdot]$,E_1 and $i \geq 10n/31$ if $t = t_1$. For this approximation to work we might want to assume that $\mathbb{E}[X_i] \propto \sqrt{z/n} \in \Theta(1)$. If this condition is not met then, as n tends to ∞, the fraction of edges per node connecting to $\mathcal{B}_0$ tends to 0 leading to a situation with a diminishing fraction of late nodes becoming black in the first round. The main reason for this is the fact, that α is a constant independent of n as $n \to \infty$. If we instead focus on large but constant n, then one might also investigate dependencies such as $\alpha \leq \frac{1}{m \log n}$. We will focus on this case, that n is large but constant, as the contrary case leads to the conclusion that $z \in \Theta(n)$ for arbitrary α. This is consistent with the observations in the simulations where for large n where the upper bounds on the dynamic monopolies were also linear in size for $\alpha > 1/3$.

Continuing with our approximation, we would like to focus on the case, when a single edge connecting to $\mathcal{B}_0$ will be sufficient to become black. Since the nodes $i \in I_{t_1,bounded}$ have degree less equal $3m/2$, we conclude that $\alpha \leq 2/3m$.[6] Since now a single of the m edges per node is sufficient, we can approximate J by calculating the probability that neither of the m edges connects. Thus, we have

$$\begin{aligned}
J &\geq \sum_{1 \in I_{t_1,bounded}} \mathbb{P}(\{\text{at least one edge of node } i \text{ connects to } \mathcal{B}_0\}) \\
&= \sum_{1 \in I_{t_1,bounded}} 1 - \mathbb{P}(\{\text{no edge of node } i \text{ connects to } \mathcal{B}_0\}) \\
&= \sum_{1 \in I_{t_1,bounded}} 1 - (1 - W_z/W_i)^m && (E_1) \\
&\geq \sum_{1 \in I_{t_1,bounded}} 1 - \left(1 - 9/11\sqrt{z/i}\right)^m && (E_1) \\
&\geq \sum_{1 \in I_{t_1,bounded}} 1 - \left(1 - \sqrt{z/n}\right)^m && (i \leq 20n/31) \\
&\geq \frac{10}{31} n(1 - (1 - \sqrt{z/n})^m). && (|I_{t,bounded}| = 2^{t_1 - 2})
\end{aligned}$$

To get a better understanding, we give the value of $\frac{10}{31} n(1 - (1 - \sqrt{z/n})^m) =: g(n, m)$ for different m:

$$\begin{aligned}
g(n,1) =& \frac{10}{31} n\sqrt{zn} && \in \mathcal{O}(\sqrt{zn}) \\
g(n,2) =& \frac{10}{31}(2\sqrt{zn} - z) && \in \mathcal{O}(\sqrt{zn} - z) \\
g(n,3) =& \frac{10}{31}(3\sqrt{zn} - 3z + \sqrt{z^3/n}) && \in \mathcal{O}(\sqrt{zn} - z + \sqrt{z^3/n}) \\
g(n,4) =& \frac{10}{31}(4\sqrt{zn} - 6z + 4\sqrt{z^3/n} - z^2/n) && \in \mathcal{O}(\sqrt{zn} - z + \sqrt{z^3/n} - z^2/n)
\end{aligned}$$

Morever, consider table 3.1 which summarizes in $\mathcal{O}$-Notation the sizes of the lower bound for three sizes of z. For constantly many initially black nodes, we make at-least $\sqrt{n}$ nodes black in the first round. Making more nodes black initially does lead more nodes becoming black - as is the case for $z \in \Theta(\log n)$. The lower bound is useful for $m \leq 3$ - after that, the larger powers slowly control and lead to fluctuations in the lower bound. The closer z is to being linear, the less useful the lower bound becomes for large m. However, close to linear z is not important, we are concerned with small

[6] This is independent of n as we as using t_1 instead of a general t.

	$z \in \mathcal{O}(1)$	$z \in \Theta(\log n)$	$z \in \Theta(n)$
$g(n,1)$	$\sqrt{n}$	$\sqrt{n \log n}$	n
$g(n,2)$	$\sqrt{n}$	$\sqrt{n \log n} - \log n$	n
$g(n,3)$	$\sqrt{n}$	$\sqrt{n \log n} + \log^{1.5}(n)/n$	n
$g(n,4)$	$\sqrt{n}$	$\sqrt{n \log n} - \log^2(n)/n$	$n - n^2$

Table 3.1: The $\mathcal{O}$-classes of the lower bound $g(n,m)$ for $m \in [4]$ for three classes of z: constant (1st column), logarithmic (3rd cloumn) and linear (4th column)

$\alpha \leq \frac{2}{3m}$ - where linear z are not an interesting result. We will be working with the lower bound $\frac{10}{31}\sqrt{zn} \approx g(n,m) \in \Theta(\sqrt{zn})$ as this is accurate for small m and small z.

Second Round With $\Theta(\sqrt{zn})$ black low degree late nodes (which we will call $\mathcal{B}_{1,late}$ shortly), we can proceed to investigate how many nodes will get black in the second round. For that, we want to consider another interval I_t - now with a slightly smaller t - and see how many of them get infected. For that, we consider $t = t_2$ for some $t_2 < t_1$. The idea is that, as more low degree nodes are black now, more nodes with slightly higher degree can also become black. In this sense, our intuition is that in each round nodes of larger and larger degrees will be added - until the highest degree nodes (i.e. $z + \mathcal{O}(1)$) - that need the most nodes to become black. Using $E_{2,compl}$ we can again compute an upper bound for the node degrees. For $i_2 \in I_{t_2,bounded}$ we have

$$\deg(i_2, G_m^n) \leq m\left(1 + \frac{4}{9}\sqrt{n2^{-t_2}} - \sqrt{1 - 2^{-t_2-1}}\right). \tag{3.15}$$

With the degree bounded, we would like to use the newly black nodes $\mathcal{B}_{1,late}$ to make these nodes black and get an understanding of their weight and relevance. We investigated this and used the procedure we describe in the next paragraph. In a full treatment however, the contribution from the initially black nodes $[z]$ should be counted as well - as their weight is suspected to be relevant.

To calculate how many nodes in $I_{t_2,bounded}$ get black, we first need to calculate the number of edges from the $\mathcal{B}_{1,late}$ nodes that could potentially connect to $I_{t_2,bounded}$. For that we need to know, how many edges of each node $i_1 \in \mathcal{B}_{1,late}$ did not connect to $\mathcal{B}_0$. While we do know, that each node in $\mathcal{B}_{1,late}$ send at least one edge to $\mathcal{B}_0$, we don't know how many will be left that are not. For the moment, we will proceed with the guess, that almost all nodes in $\mathcal{B}_{1,late}$ have at most *half* of their m edges available for the second round. In contrast, in a complete investigation, one would not only lower bound the number of newly black nodes in the first round (as we did with $J \geq \mathcal{B}_{1,late}$) but also count how many edges each node sends to $\mathcal{B}_0$. As the m

edges are independent (given $y_{*,*}$), for sub-linear z, the probability for each additional edge going to $\mathcal{B}_0$ will be diminishing in total. Since, for example for $z = 1$, on average we have $\Theta(\sqrt{n})$ nodes with one at least one edge to $\mathcal{B}_0$ - then $\Theta(\sqrt{n})$ of these will have two edges - and $\Theta(\sqrt{n})$ of these again, will have three edges etc. This analysis would be easily upper bounded in a complete investigation - and therefore we continue now with the guess that the upper bound is $m/2$ per node in $\mathcal{B}_{1,late}$.

With $\frac{10}{31}\sqrt{zn} \cdot m/2 = \frac{10m\sqrt{zn}}{62} := \Delta$ edges from $\mathcal{B}_{1,late}$ available, we would like to see, how many nodes in $I_{t_2,bounded}$ get black. For that, let us first focus on a single node $i_2 \in I_{t_2,bounded}$. If we knew a lower bound on the weight of i_2 - i.e. $w_{i_2} \geq d$ for some d - then we need σ edges of the total Δ edges to connect with node i_2 to make this node black. Regardng the number of edges required to make i_2 black, we have

$$\alpha \mathrm{deg}(i, G_m^n) \leq 2/3 \left(1 + \frac{4}{9}\sqrt{n2^{-t_2}} - \sqrt{1 - 2^{-t_2-1}}\right) =: \sigma.$$

The probability, that a node $i_1 \in \mathcal{B}_{t_1,late}$ connects with i_2 were then at least $q := d/W_{i_1} \leq w_{i_2}/W_{i_1}$ - using by same argument as in lemma 3.8. In other words, with $\mathrm{Bin}(x; n, p) = \mathbb{P}(X \leq x)$ denoting the cumulative distribution function of a random variable $X \sim \mathrm{Bin}(n, p)$, the probability of node i getting black is at least $\mathrm{Bin}(\sigma; \Delta, q)$. To make i_2 black w.h.p., we simply set

$$\mathrm{Bin}(\sigma; \Delta, q) \geq 1 - \mathrm{o}(1)$$

where the binomial distribution function can be bounded using existing statistical results. To make this calculation however, we would need the lower bound $w_{i_2} \geq q$ we supposed previously.

To find such a lower bound, we would need another variant of E_2 and $E_{2,compl}$ - one which lower and upper bounds the weights of the nodes. We suspect such an event $E_{2,lower+upper}$ to be able to be proved with the framework given. However, as it seems to need a more elaborate approach than what was necessary for proving $E_{2,compl}$ (as far as we have tried) - we will instead continue with a guess on $E_{2,lower+upper}$ and leave the proof for a further investigation. For the remaining of this section, we assume it is possible to prove that this event (or a variant with different constants) is true w.h.p.:

$$E_{2,lower+upper} = \left\{ I_t \text{ contains } 2^{t-2} \text{ nodes } i \text{ with } \frac{1}{10\sqrt{in}} \leq w_i \leq \frac{1}{5\sqrt{in}} \text{ for } a \leq t < b \right\}$$

With that assumed, we proceed with by it for bounding the weights of the nodes $I_{t_1,bounded}$ with $w_i \geq \frac{1}{5\sqrt{in}} = d$. From that, $i_1 \leq 20n/31$, E_1 and $i_2 \leq$

2^{t_2+1} we get the connection probability $q = \frac{2}{18\sqrt{2^{t_2}n}}$. With this, let us now extend our focus on all nodes $i_2 \in \mathrm{I}_{t_2,bounded}$. Instead of using all our edges to make only one node black, we would like to infect as many nodes as is possible with still having a w.h.p. result. For that, let $B \in [|I_{t_2,bounded}|]$ denote the number of nodes in $I_{t_2,bounded}$ we will make black in the next calculation. We want to invest as few edges into a node i_2 as possible such that it still gets black. A simple way is to distribute our Δ edges evenly onto B (arbitrarily chosen) nodes from $\mathcal{B}_{1,bounded}$. In other words, a node among these B nodes becomes black w.h.p. if

$$\mathrm{Bin}(\sigma; \Delta/B, q) \geq 1 - \mathrm{o}(1) \tag{3.16}$$

Thus our goal is finding the largest B such that equation 3.16 is satisfied. We will now proceed with a first try to calculate such a B. Concretely, we aim for $\mathrm{Bin}(\sigma; \Delta/B, q) = 1 - \mathbb{P}(X < \sigma) \geq 1 - n^{-2}$ for $X \sim \mathrm{Bin}(\Delta/B, q)$. Thus, we set $\mathbb{P}(X < \sigma) \leq n^{-2}$. We can apply lemma 3.2 if we can show that $\sigma < \mathbb{E}[X]$. Having $\sigma < \mathbb{E}[X] = \Delta q/B$ is an intuitive precondition, as otherwise the probability of making a node black would be less than $\frac{1}{2}$. Unfortunately however, for large n the opposite is the case for sub-linear z:[7]

$$\begin{aligned}
\mathbb{E}[X] &= \Delta q \\
&= \frac{10m\sqrt{zn}}{62} \frac{2}{18\sqrt{2^{t_2}n}} \\
&= \Theta(\sqrt{z/2^{t_2}}) \\
&\ll \Theta(1) + \Theta(\sqrt{n/2^{t_2}}) && (z \in \mathrm{o}(n)) \\
&= \frac{2}{3}\left(1 + \frac{4}{9}\sqrt{n/2^{t_2} - \sqrt{1 - 2^{-t_2-1}}}\right) \\
&= \sigma
\end{aligned}$$

In other words, for large n and a non-trivial sub-linear z the nodes $\mathbb{B}_{1,late}$ are not enough to make even a single node black. This is the point, where the investigation could be continued by re-introducing the $[z]$ black nodes in the process of deciding how many nodes get black in the second round. The procedure form the previous paragraph need to be adjusted slightly. We cannot simply increase q and Δ to count the contributions from the $[z]$ nodes as this would falsely imply that edges to $\mathcal{B}_0$ and edges from $\mathcal{B}_{1,late}$ had the same probability to connect with $I_{t_2,bounded}$. Hence it is advisable to instead consider the sum of two binomial distributions. As the sum of binomial distributions are well-studied statistical cases, we believe is it possible to conclude relevant results. From this perspective, it is possible to continue

[7] We focus on sub-linear z as we focus on small $\alpha \leq \frac{2}{3m}$ where a linear dynamic monopoly size z wouldn't be interesting.

this approach in a full investigation. After a few rounds, one might be able to succinctly abstract over the core calculations and statements which need to be satisfied to compute the progress in the next round. Ideally, one would be able to give a statement that relates $z = \mathcal{B}_0$ to the sizes of all the nodes in all considered intervals $(I_{t_1}, I_{t_2}, \dots, I_{t_k})$ for $t_1 > t_2 > \cdots > t_k \in \Theta(s)$ that were bounded in degree and eventually got black. This way, we modeled the progression of the black nodes from late nodes to earlier and earlier nodes until they reach to very early nodes close to s. The fact, that it only concerns a constant fraction of nodes in each considered interval I_t is not a problem as, once the process has changed nodes close to s, we will have linearly many nodes changed black - from where we suspect it to be not too difficult to make all remaining nodes black as well. Each remaining node will have linearly many nodes before it and after it which could potentially connect to it. Aside form this, finalizing the above process would be very useful since the direct relationship between the z and how many nodes get black in which round would form a basis for investigation of further relevant questions regarding α-BP on PA graphs.

3.5 Abstractions and Approaches

In this section, we will give a short overview of a few remaining ideas and approaches, that we tried in the few months of investigating r-BP and α-BP on preferential attachment graphs. For each approach and abstraction, we provide a short description and certain insights.

Vertex Cover Reduction For $\alpha > 1 - \frac{1}{m}$ the size of a dynamic monopoly is found to be equal to the size of a minimal vertex cover. Perhaps there are existing results on the size of the minimum vertex cover for general graphs of preferential attachment graphs, that can be used in this special case. We tried a few ways to give good lower bounds on the minimal size of a vertex cover for preferential attachment graphs as we did not find any existing literature for vertex cover on preferential attachment graphs. By counting the degrees, we got the approximate result that $VC \geq (\frac{N}{4} - \frac{\sqrt{N}}{2})/m$. If the proof can be made rigorous, then one can deduce a linear lower bound for the size of the minimumy dynamic monopoly for $\alpha \geq \frac{1}{\max_k \deg(k, G_m^n)}$.

Suffocation of Edges We can look at evolution of the number of edges between the black node and white nodes as the dynamic process progresses. For $\alpha > 0.5$, each newly black node will decrease the number of edges between black nodes and white nodes. If the size of initially black nodes is too small, then eventually the number of edges between the black nodes and white nodes is insufficient to make a few remaining nodes black. This may yield a lower bound on the size of a dynamic monopoly for large α. With a good model, such as the alternative model, one might reach a result.

Labeled Branching Process In [23] a mathematical tool called *Label Branching Process* was used to settle the case for r-BP on PA graphs. Given the usefulness of that tool in that paper, one might try to use a similar adapted approach for α-BP. Labeled Branching Process might be as fruitful of an abstraction as was the alternative model.

Chapter 4

Discussion and Conclusion

The investigation of minimum dynamic monopolies under α-BP on PA graphs revealed a few insights regarding the dynamics in the topology of PA graphs as well as giving a picture on how the situation looks like for various α. Working on the model directly was simple in the beginning - however, it leads to a dramatic increase in complexity later and became unfruitful in gaining an understanding of the situation. The computer simulations gave a first tangible overview of the situation and lead to various regions of interest as well as putting the choice of nodes in the order of arrival in Lemmas (rather than degree-based choice) into another perspective. Working with the alternative model, while having a learning curve initially, was the most fruitful as it enabled us to mathematically express two rounds of the dynamics - and it did not seem to lead to a roadblock. Many of the remaining approaches, while underdeveloped, might become very useful if done with more experience in mathematical proof making. A big problem during most of the investigation was the difficulty or inexpressibility of certain ideas and the intuition from the point of view and experience of the author. Much progress was only possible, when (towards a late stage of the investigation) the alternative model was used instead. A future and full investigation can find intermediate results, insights, guidance and ideas in this thesis.

4.1 Final Words by the Author

The investigation for this thesis was, while challenging at many points, an interesting learning experience. Theoretical research is less predictable in its outcome than is practical research or engineering. Much of the challenges were due to the lack of practise with the mathematical tools. While it was possible to give intuitive arguments and proof sketches, the difficulties in rigorously proving them or sometimes even expressing them made progress often halt. I believe, having a more solid mathematical background in the tools used here (by practise) and having a generally diverse mathematical knowledge can greatly help in deriving results. After all, I believe, the extent of the mathematical tools one knows (or can develop them themselves) is directly linked to the universe of ideas and intuitions one can think and express in established mathematical language.

Bibliography

[1] Mohammed Amin Abdullah and Moez Draief. Global majority consensus by local majority polling on graphs of a given degree sequence. *Discrete Applied Mathematics*, 180:1–10, 2015.

[2] Hamed Amini and Nikolaos Fountoulakis. Bootstrap percolation in power-law random graphs. *Journal of Statistical Physics*, 155(1):72–92, 2014.

[3] József Balogh and Béla Bollobás. Bootstrap percolation on the hypercube. *Probability Theory and Related Fields*, 134(4):624–648, 2006.

[4] József Balogh, Béla Bollobás, Hugo Duminil-Copin, and Robert Morris. The sharp threshold for bootstrap percolation in all dimensions. *Transactions of the American Mathematical Society*, 364(5):2667–2701, 2012.

[5] József Balogh and Gábor Pete. Random disease on the square grid. *Random Structures & Algorithms*, 13(3-4):409–422, 1998.

[6] Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *science*, 286(5439):509–512, 1999.

[7] Albert-László Barabási et al. *Network science*. Cambridge university press, 2016.

[8] Vladimir Batagelj and Ulrik Brandes. Efficient generation of large random networks. *Physical Review E*, 71(3):036113, 2005.

[9] B´ ela Bollobás, Oliver Riordan, Joel Spencer, and Gábor Tusnády. The degree sequence of a scale-free random graph process. *Random Structures & Algorithms*, 18(3):279–290, 2001.

[10] Béla Bollobás and Oliver Riordan. The diameter of a scale-free random graph. *Combinatorica*, 24(1):5–34, 2004.

[11] John Chalupa, Paul L Leath, and Gary R Reich. Bootstrap percolation on a bethe lattice. *Journal of Physics C: Solid State Physics*, 12(1):L31, 1979.

[12] Ching-Lueh Chang. Triggering cascades on undirected connected graphs. *Information Processing Letters*, 111(19):973 – 978, 2011.

[13] Ching-Lueh Chang and Yuh-Dauh Lyuu. Bounding the number of tolerable faults in majority-based systems. In *International Conference on Algorithms and Complexity*, pages 109–119. Springer, 2010.

[14] Ching-Lueh Chang and Yuh-Dauh Lyuu. Spreading of messages in random graphs. *Theory of Computing Systems*, 48(2):389–401, 2011.

[15] Ching-Lueh Chang and Yuh-Dauh Lyuu. Triggering cascades on strongly connected directed graphs. In *Parallel Architectures, Algorithms and Programming (PAAP), 2012 Fifth International Symposium on*, pages 95–99. IEEE, 2012.

[16] Ching-Lueh Chang and Yuh-Dauh Lyuu. Bounding the sizes of dynamic monopolies and convergent sets for threshold-based cascades. *Theoretical Computer Science*, 468:37–49, 2013.

[17] Ching-Lueh Chang and Chao-Hong Wang. On reversible cascades in scale-free and erdős-rényi random graphs. *Theory of Computing Systems*, 52(2):303–318, 2013.

[18] Chun-Ying Chiang, Liang-Hao Huang, Bo-Jr Li, Jiaojiao Wu, and Hong-Gwa Yeh. Some results on the target set selection problem. *Journal of Combinatorial Optimization*, 25(4):702–715, 2013.

[19] Fan Chung and Linyuan Lu. Connected components in random graphs with given expected degree sequences. *Annals of combinatorics*, 6(2):125–145, 2002.

[20] Amin Coja-Oghlan, Uriel Feige, Michael Krivelevich, and Daniel Reichman. Contagious sets in expanders. In *Proceedings of the twenty-sixth annual ACM-SIAM symposium on discrete algorithms*, pages 1953–1987. Society for Industrial and Applied Mathematics, 2015.

[21] Benjamin Doerr, Mahmoud Fouz, and Tobias Friedrich. Asynchronous rumor spreading in preferential attachment graphs. In *Scandinavian Workshop on Algorithm Theory*, pages 307–315. Springer, 2012.

[22] Devdatt P Dubhashi and Alessandro Panconesi. *Concentration of measure for the analysis of randomized algorithms*. Cambridge University Press, 2009.

[23] Roozbeh Ebrahimi, Jie Gao, Golnaz Ghasemiesfeh, and Grant Schoenbeck. How complex contagions spread quickly in preferential attachment models and other time-evolving networks. *IEEE Transactions on Network Science and Engineering*, 4(4):201–214, 2017.

[24] Roozbeh Ebrahimi, Jie Gao, Golnaz Ghasemiesfeh, and Grant Schoenebeck. Complex contagions in kleinberg's small world model. In *Proceedings of the 2015 Conference on Innovations in Theoretical Computer Science*, pages 63–72. ACM, 2015.

[25] Uriel Feige, Michael Krivelevich, Daniel Reichman, et al. Contagious sets in random graphs. *The Annals of Applied Probability*, 27(5):2675–2697, 2017.

[26] Paola Flocchini, Rastislav Královič, Peter Ružička, Alessandro Roncato, and Nicola Santoro. On time versus size for monotone dynamic monopolies in regular topologies. *Journal of Discrete Algorithms*, 1(2):129–150, 2003.

[27] Paola Flocchini, Elena Lodi, Fabrizio Luccio, Linda Pagli, and Nicola Santoro. Dynamic monopolies in tori. *Discrete applied mathematics*, 137(2):197–212, 2004.

[28] Daniel Freund, Matthias Poloczek, and Daniel Reichman. Contagious sets in dense graphs. *European Journal of Combinatorics*, 68:66–78, 2018.

[29] Frederik Garbe, Richard Mycroft, and Andrew McDowell. Contagious sets in a degree-proportional bootstrap percolation process. *Random Structures & Algorithms*, 53(4):638–651, 2018.

[30] Bernd Gärtner and Ahad N Zehmakan. Color war: Cellular automata with majority-rule. In *International Conference on Language and Automata Theory and Applications*, pages 393–404. Springer, 2017.

[31] Bernd Gärtner and Ahad N Zehmakan. Majority model on random regular graphs. In *Latin American Symposium on Theoretical Informatics*, pages 572–583. Springer, 2018.

[32] Golnaz Ghasemiesfeh, Roozbeh Ebrahimi, and Jie Gao. Complex contagion and the weakness of long ties in social networks: revisited. In *Proceedings of the fourteenth ACM conference on Electronic Commerce*, pages 507–524. ACM, 2013.

[33] Mark Granovetter. Threshold models of collective behavior. *American journal of sociology*, 83(6):1420–1443, 1978.

[34] Alberto Guggiola and Guilhem Semerjian. Minimal contagious sets in random regular graphs. *Journal of Statistical Physics*, 158(2):300–358, 2015.

[35] Svante Janson, Tomasz Łuczak, Tatyana Turova, Thomas Vallier, et al. Bootstrap percolation on the random graph $g_{\{n,p\}}$. *The Annals of Applied Probability*, 22(5):1989–2047, 2012.

[36] Clemens Jeger and Ahad N Zehmakan. Dynamic monopolies in reversible bootstrap percolation. *arXiv preprint arXiv:1805.07392*, 2018.

[37] Yashodhan Kanoria, Andrea Montanari, et al. Majority dynamics on trees and the dynamic cavity method. *The Annals of Applied Probability*, 21(5):1694–1748, 2011.

[38] David Kempe, Jon Kleinberg, and Éva Tardos. Maximizing the spread of influence through a social network. In *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 137–146. ACM, 2003.

[39] Konstantin Klemm and Victor M Eguiluz. Growing scale-free networks with small-world behavior. *Physical Review E*, 65(5):057102, 2002.

[40] Fabrizio Luccio, Linda Pagli, and Hermineh Sanossian. Irreversible dynamos in butterflies. In *SIROCCO*, pages 204–218, 1999.

[41] Natasha Morrison and Jonathan A Noel. Extremal bounds for bootstrap percolation in the hypercube. *Journal of Combinatorial Theory, Series A*, 156:61–84, 2018.

[42] David Peleg. Local majority voting, small coalitions and controlling monopolies in graphs: A review. In *Proc. of 3rd Colloquium on Structural Information and Communication Complexity*, pages 152–169, 1997.

[43] Luiz FC Pereira and FG Brady Moreira. Majority-vote model on random graphs. *Physical Review E*, 71(1):016123, 2005.

[44] Daniel Reichman. New bounds for contagious sets. *Discrete Mathematics*, 312(10):1812–1814, 2012.

[45] Duncan J Watts. A simple model of global cascades on random networks. *Proceedings of the National Academy of Sciences*, 99(9):5766–5771, 2002.

[46] Duncan J Watts and Steven H Strogatz. Collective dynamics of 'small-world'networks. *nature*, 393(6684):440, 1998.

[47] Ahad N Zehmakan. Two phase transitions in two-way bootstrap percolation. *arXiv preprint arXiv:1809.10764*, 2018.

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Declaration of originality

The signed declaration of originality is a component of every semester paper, Bachelor's thesis, Master's thesis and any other degree paper undertaken during the course of studies, including the respective electronic versions.

Lecturers may also require a declaration of originality for other written papers compiled for their courses.

I hereby confirm that I am the sole author of the written work here enclosed and that I have compiled it in my own words. Parts excepted are corrections of form and content by the supervisor.

Title of work (in block letters):

Bootstrap Percolation on Complex Networks

Authored by (in block letters):
For papers written by groups the names of all authors are required.

Name(s):	**First name(s):**
Sahoo	Chandrakant Swaneet Kumar

With my signature I confirm that
- I have committed none of the forms of plagiarism described in the 'Citation etiquette' information sheet.
- I have documented all methods, data and processes truthfully.
- I have not manipulated any data.
- I have mentioned all persons who were significant facilitators of the work.

I am aware that the work may be screened electronically for plagiarism.

Place, date	**Signature(s)**
Zurich, 1.3.19	Chan Swa Kum Sahoo

For papers written by groups the names of all authors are required. Their signatures collectively guarantee the entire content of the written paper.