UNIVERSITÄT PADERBORN
Die Universität der Informationsgesellschaft

Epidemic Spreading and Information Dissemination in Technological and Social Systems

---

Adrian Ogierman

Dissertation in Computer Science

Submitted to the

Faculty of Computer Science,
Electrical Engineering and Mathematics,
University of Paderborn

in partial fulfillment of the requirements for the degree of

doctor rerum naturalium
(Dr. rer. nat.)

Paderborn, 2014

Supervisors:

Prof. Dr. Robert Elsässer	**University of Salzburg**
Prof. Dr. Christian Scheideler	**University of Paderborn**

Date of submission: **July 28, 2014**

Acknowledgments

First and foremost, I have to thank my supervisors, Robert Elsässer and Christian Scheideler. Without their assistance and dedicated involvement in every step throughout the process, this thesis would have never been accomplished. It was always a pleasure to work with them, and I am really grateful for the opportunity to study and evolve under their influence. As it always is, some periods during that time were more difficult than others. However, I could always count on their support. Even during the very difficult times of funding problems, I was always positive not to be let down. In this context I am especially thankful for the continuous support of Christian Scheideler, who gave me a position in his group for the last three years and therefore the liberty to fully concentrate on my research.

At this point, I must not forget about one very special person: Ulf-Peter Schroeder, the counselor I trust. Whenever I struggled on my way due to non-research related problems, he was there to help. It was a priceless service he offered me (and many other students) in terms of advise. In times of need he was the one who got us back on track. I would like to thank you very much for your outstanding support, patience, understanding, and all the discussions I bothered you with over these past five years.

I would also like to show gratitude to the persons of the research group I was working in. Of course, I start with one of the most important persons in said group: Marion Bewermeyer. She keeps this group alive by preventing it from being buried under tons of paperwork. Her skill to handle the bureaucracy and her knowledge about who to address (or to blame) when something does not proceed as intended have always been of great value to me. And although I took advantage of it occasionally, I hope she did not mind. Without a doubt, another person deserving a separate reference is Kálmán Graffi. Our frequent discussions considering the very essence of this thesis during the last two years continuously helped me to improve the structure of this paper. Then there is Dominik Dumrauf, a former colleague who had to listen to all of my complaints and problems I have had for almost three years while sharing an office with me. Since a lot of unpleasant and shocking events occurred during that time, I must not forget his moral support here.

Of course, I have to thank all of my colleagues during that time. I am very grateful for every single person I got the chance to get acquainted with, all the laughter and tears we shared, and even the (occasionally boring) duties we had to fulfill.

Obviously, getting through my dissertation required more than academic support, and I have many, many people to thank for listening to and, at times, having to tolerate me over the past five years. Most importantly, none of this could have happened without my family. To my Grandmother, who kept reminding me that there is more than just work. My parents, who offered their encouragement through phone calls. Every time I was ready to quit, you did not let me and I am forever grateful. This dissertation stands as a testament to your unconditional love and encouragement.

Zusammenfassung

In dieser Arbeit betrachten wir Probleme aus dem Bereich der Nachrichten- und Krankheitsverbreitung in unterschiedlichen Netzwerktypen, motiviert durch das stetig wachsende Interesse an *sozialen Netzwerken*. Hierbei beschäftigen wir uns mit dynamischen als auch statischen Strukturen aus dem Gebiet der technologischen Netzwerke, also Netzwerke, die hauptsächlich von technischen Geräten etabliert und aufrecht erhalten werden, und der sozialen Netzwerke, also Netzwerke, die hauptsächlich durch menschliche Verhaltensweisen und Eigenheiten gebildet werden.
Als erste Fragestellung untersuchen wir, ob ein verteiltes Protokoll zur Nachrichtenverbreitung in Netzwerken mit Power Law Knotengradverteilung existiert, so dass sich die Knotengradverteilung nicht negativ bemerkbar macht. Als Antwort präsentieren wir ein Protokoll, welches mit hoher Wahrscheinlichkeit nur $\mathcal{O}(\log n)$ viele Runden mit $\mathcal{O}(n \log \log n)$ vielen Nachrichten benötigt um alle n Knoten zu informieren. Dieses Protokoll stellt eine Abwandlung des Random Phone-Call Modells dar, in der jeder Knoten in jeder Runde eine konstante Anzahl an Kommunikationspartnern wählt.
Wenn sich jedoch Nachrichten in solchen Netzwerken schnell ausbreiten können, wie könnten dann Strategien, welche ebenfalls einen ähnlichen Ausbreitungsprozess nutzen, zur Eindämmung dieser Ausbreitung aussehen? Um dies zu untersuchen betrachten wir folgendes Modell. Sei $\mathcal{V}$ der für die Ausbreitung der schädlichen Nachricht/Infektion verantwortliche Prozess. Weiterhin lassen wir $\mathcal{V}$ sich von jedem infizierten Knoten über eine konstante Anzahl von Verbindungen verbreiten. Unsere Strategie zur Bekämpfung von $\mathcal{V}$ wird an jedem infizierten Knoten nach einer konstanten Anzahl von Zeitschritten aktiviert. Im Falle eines minimalen Knotengrades von $\Omega(\log \log n)$ zeigen wir, dass es bereits genügt nur die direkte Nachbarschaft eines jeden infizierten Knoten einmalig zu immunisieren um die Infektion mit hoher Wahrscheinlichkeit zu eliminieren. Ist der minimale Knotengrad jedoch eine Konstante, so ändern wir die Strategie wie folgt. Jeder infizierte Knoten v immunisiert alle Knoten in seiner $\mathcal{O}(\log^2(\deg_v))$-Nachbarschaft, wobei $\deg_v$ den Knotengrad von Knoten v bezeichnet. So wird $\mathcal{V}$ mit hoher Wahrscheinlichkeit nicht mehr als $o(n)$ viele Knoten infizieren.
Doch wie verhält es sich, wenn wir ein Netzwerk mit mobilen Entitäten haben? Zur Untersuchung dieser Fragestellung betrachten wir eine Epidemie in einer städtischen Umgebung mit mobilen Einwohnern. Jedem Ort der Umgebung wird eine Attraktivität gemäß einer Power Law Verteilung zugewiesen. Jedes Individuum wählt in jeder Runde seinen Zielort zufällig nach der Attraktivität aus. Werden nun keinerlei Gegenmaßnahmen getroffen, so bleibt dennoch mit hoher Wahrscheinlichkeit ein polynomieller Anteil der Population von der Epidemie unberührt. Werden jedoch Gegenmaßnahmen, welche z.B. das Verhalten der Individuen beeinflussen, getroffen, so werden mit Wahrscheinlichkeit $1 - o(1)$ nur polylogarithmisch viele Individuen infiziert und die Epidemie innerhalb von $\mathcal{O}((\log \log n)^4)$ vielen Runden ausgerottet.

Abstract

In this thesis we consider the problems of information dissemination and epidemic spreading in different types of networks, motivated by the recent hype of social networks and their relation to human behavior. We divide the types of dissemination processes into information dissemination processes eager to spread a (malicious) piece of data within the network, and epidemiological processes struggling to survive by infecting new hosts. The types of networks we examine in detail, which may be highly dynamic as well as static, are technological networks, i.e., networks created and maintained mainly by artificial devices, and social networks, i.e., networks mainly driven by human behavior and actions.

We start by wondering if there might be a fast decentralized dissemination protocol, such that a power law degree distribution does not slow down the dissemination process in the network. It turns out that an adaptation of the Random Phone-Call model, where each node calls on a constant amount of randomly chosen communication partners in each round, leads to a protocol that informs all n nodes within $\mathcal{O}(\log n)$ many rounds using $\mathcal{O}(n \log\log n)$ many transmissions with high probability. Naturally, a subsequent question arises.

How do we design a counteracting dissemination process to combat the malicious one, denoted by $\mathcal{V}$, keeping the potentially fast dissemination of $\mathcal{V}$ in mind? Suppose $\mathcal{V}$ uses a constant number of randomly chosen connections of each infected node to infect others for one time only and suppose that the counteracting dissemination process is activated on each infected node after a constant delay. Under this model we present two counteracting strategies. We show that it suffices to immunize the neighborhood of each infected node to eliminate the infection with high probability, provided the minimum degree of the network is $\Omega(\log\log n)$. Otherwise, if the minimum degree of the network is constant, we propose to immunize every node within $\mathcal{O}(\log^2(\deg_v))$ many hops of each infected node v, where $\deg_v$ denotes the degree of node v. Executing this strategy we prove that $\mathcal{V}$ does not infect more than $o(n)$ many nodes until it is eliminated with high probability.

Finally, we take mobility into account and examine an epidemic outbreak in an urban environment inhabited by mobile individuals on a small and on a large scale. On a small scale, in each round each individual chooses a location independently at random according to a power law distribution describing the attractiveness of the locations. We show that at least a polynomial fraction of the individuals remains uninfected even if they do not respond to the epidemic outbreak in any way. However, if the epidemic outbreak does influence the individual's behavioral pattern and certain countermeasures are applied, then only a polylogarithmic amount of individuals is infected until the epidemic is embanked after $\mathcal{O}((\log\log n)^4)$ many rounds with probability $1 - o(1)$. Furthermore, we analyze such epidemic outbreaks empirically on a large scale and find that countermeasures used in the real world have a comparable impact in our model.

Table of Contents

Chapter 1

Introduction

Communication and interaction are two of the most important keys our society is built upon. Our ancestors already used these skills to rise to power and become the predominant species on this planet. Nowadays, more than ever, we rely on them on a daily basis in our private and professional life. They were and still are fundamental for the enormous technical progress we have made in recent time, thus implicitly fueling our economy. In the meantime an entire industry has formed around these skills, focusing on research, education, and development. Facilities like universities and companies all over the world dedicate a fortune to develop new technologies, improve the old ones, and deliver ground-breaking ideas that would not have been possible otherwise.

Throughout the decades we experienced many changes on a more personal level as well. Each of us takes telecommunication systems for granted and uses the Internet as an easily accessible source of information nowadays. It is common to carry a cellphone or laptop at any time we want and thus we stay connected all day long. But all these devices themselves rely on networks, wired as well as wireless ones. Obviously, keeping all these networks working is a non-trivial but viable task. However, we first need to understand their essence in order to be able to diagnose, secure, and improve them. Some of them may change dynamically over time, others may not. Either way, due to our behavior, our habits, and cultural circumstances, we do influence them. Furthermore, many of these networks serve one purpose and one purpose only: to make our life easier. This implies a strict coupling and adaptation of the technology to us as persons. Interestingly, this kind of influence works both ways, especially since we use so many wireless devices (e.g., cellphones, tablets, or laptops) on a daily basis.

On the other hand, due to the technical progress, our behavioral patterns and social structures evolved as well. Our cities are still growing and are becoming more crowded. We live in an industrialized, well-connected world. Traveling and trading around the globe is easier and cheaper than ever before. However, we also became more vulnerable to attacks exploiting these social structures. For example, rumors circulating in social networks, stating false accusations, can do much harm to a person's reputation and can be very hard to defeat. Some epidemics spread very quickly throughout the population, even without being easily imported by plane. Understanding the way and reasons why

such processes spread through our social structures so quickly is an important step towards counteracting them.

1.1 Scope of the Thesis

We handle the following types of networks motivated by the above statements:

1. Technological networks, i.e., artificial networks (such as the Internet) primarily created through and maintained by artificial devices (such as laptops or smartphones), and

2. Social networks, i.e., networks (such as person-to-person communication networks) created by social tasks performed by humans.

Obviously, a social network may be created with the aid of a technological network (e.g., Facebook, Google+). Note, however, that person-to-person contacts, for example, can be modeled as a social network too. This implies that depending on the situation each one of the two types may utilize the other one but, in general, does not necessarily rely on this.

In the context of technological networks we mainly focus on graphs providing a property following a power law distribution. Such a property may be the degree distribution, for instance. A degree distribution follows a power law distribution if the amount of nodes in the network with degree k is proportional to $k^{-\alpha}$, for sufficiently large k and α. Technological networks, depending on their nature, can be highly dynamic or entirely static. However, our main interest here is to examine information dissemination by using a (randomized) dissemination protocol/process. In Section 5.1 we state a distributed information dissemination protocol and analyze its pace in a network with power law degree distribution. In Section 5.2 then we present two similar information dissemination processes in a comparable setting. Hereby one dissemination process tries to cause as much harm to the network as possible, whereas the other one aims to stop it.

In the context of social networks we focus on information dissemination and epidemic spreading driven by social interactions. One of the main driving forces here is the individuals' movement. Especially in urban environments, some locations tend to be visited more frequently than others. Many interesting questions arise when such property is taken into account. For example, one could wonder if a dissemination process could be stopped at all in such a highly dynamic scenario or what countermeasures may be necessary to achieve this goal. Amongst others, we investigate these kind of questions analytically in Section 6.1 on a small scale and experimentally in Section 6.2 on a large scale.

1.2 Organization

The remainder of this thesis is organized as follows. In Chapter 2 we give an overview of the most recent related work relevant for this thesis. Both main directions, i.e., technological and social networks, are covered extensively with respect to information dissemination, epidemic spreading, and human behavior and movement patterns. If reasonable, we state theoretical and empirical results with respect to the aforementioned areas of interest. Since this thesis includes extensive experimental analyses, a brief survey of the different types of simulations is given at the end of this chapter.

A condensed, yet detailed, overview of the provided contributions is given in Chapter 3. The most often used basic definitions and techniques throughout our theoretical analyses are stated, and proven if necessary, in Chapter 4.

In Chapter 5 we consider information dissemination in networks/graphs experiencing at most minor changes in their structure over time. The presentation mainly suggests technological networks, although most of the insights can be easily transfered to the social context[1]. On the other hand, in Chapter 6 we examine epidemics in dynamic social networks. However, similarly as before, all statements can be easily transfered to the technical context[1].

Finally, we conclude and point out some interesting directions worth further studies in Chapter 7.

[1]See Section 2.1 for a more detailed description.

Parts of this chapter have been published in:

[35, 36, 37]

Chapter 2

Related Work

This chapter is dedicated to the most relevant related work in the scope of this thesis. The state of the art is presented according to a very basic classification system. Said classification system distinguishes between the following:

- Technical Context
- Social Context

We refer to the above classification in Section 2.1 in more detail.

2.1 Networks/Graphs

In the following we describe the previously mentioned classification system in more detail and present the most relevant related work.

In general, all networks relevant for this thesis can be classified in at least one of the following types:

1. technological networks, i.e., artificial networks (e.g., the Internet) primarily created through and maintained by artificial devices (such as laptops or smartphones), and
2. social networks, i.e., networks (e.g., person-to-person communication networks) created by social tasks performed by humans.

Within these types of networks we mainly focus on two domains:

a) information dissemination using a (randomized) dissemination protocol, and

b) information dissemination and epidemic spreading driven by social interactions.

Obviously, both domains can be found in both types of networks. For example, in technological networks clever dissemination protocols are used to maintain consistency

of databases, whereas vehicular ad hoc networks act with respect to the current situation provoked by human behavior (e.g., a traffic jam during rush hour). On the other hand, networks like Facebook and Google+ and also real-world gossiping are natural examples for social networks with respect to both domains.

In the following we categorize the related work as follows:

Classification
1. Technical Context
a. Information Dissemination Protocols
b. Power Law Networks/Graphs
c. Evolving Networks/Graphs
2. Social Context
a. Social (Interaction) Networks
b. Epidemic Spreading

Since we study a variation of the Random Phone-Call model as an information dissemination protocol in Section 5.1, we restrict the scope of the information dissemination protocols for the related work to (variations of) said model.

Definition 2.1 (Random Phone-Call Model)**.** *In the Random Phone-Call model, in every round each node calls on a neighbor chosen uniformly at random, and opens a communication channel to this node. Then, in this round, every channel can be used for bidirectional communication. We distinguish between two types of transmissions depending on the direction of the communication:* push *transmission, and* pull *transmission. In the case of a push transmission, the caller actively sends a message to the callee. In contrast, the caller actively requests a (so far unknown) message from the callee in the case of a pull transmission. The callee then responds, if possible, by sending the desired message back to the caller.*

In order to start with the presentation, we need to introduce some basic graph (construction) models.

Definition 2.2 (Preferential Attachment Model)**.** *The Preferential Attachment Model basically creates the network using the following construction rules, where the nodes are considered one after the other: 1.) the very first node forms the graph G_1^m with one node and m self loops, 2.) G_n^m is constructed from G_{n-1}^m by adding a new node w with m edges to the network. The neighbors of w, i.e., the endpoints incident to the m edges connecting w to the rest of the network, are chosen with probability proportional to their degree in G_{n-1}^m.*

Definition 2.3 (Chung-Lu Model)**.** *Let $w = (w_1, ..., w_n)$ be an expected degree sequence in the Chung-Lu model. Then an edge between two nodes v_i and v_j in the graph is created with probability $w_i \cdot w_j / \sum_i w_i$.*

Definition 2.4 (The Erdös-Renyi Model)**.** *The Erdös-Renyi graph $G(n, p)$ is a random graph with n nodes where an edge between two nodes is created independently uniformly at random with probability p.*

For the sake of completeness, we give a short overview of the most relevant simulation models in Section 2.2.

2.1.1 Technical Context

In the following we differentiate between related work concerning:

- the Random Phone-Call model,
- Power Law Graphs, and
- Evolving Graphs.

Since extensive research was conducted in these areas over the past decades, we only mention the very most important and recent results for the scope of this thesis. For the purpose of a more general point of view, we refer the reader to the excellent surveys of Hedetniemi et al. [52], Hromkovic et al. [54], and Fraigniaud and Lazard [44] with respect to broadcasting, rumor spreading, and gossiping on various kinds of network types and setups. However, for the sake of completeness, we give a short overview of these surveys before starting our own.

Hedetniemi et al. [52] provide an excellent review of results considering broadcasting and gossiping. The authors cover deterministic gossiping in various types of graphs including the complete graph, arbitrary connected graphs, grid graphs, directed graphs, and hypergraphs. Additionally, results for modifications like unrestricted calls as well as restricted ones, one-way communication, conference calls, and gossiping with randomly placed calls are listed as well. Deterministic broadcasting is being studied on graphs like trees, arbitrary connected graphs, grid graphs, and hypergraphs including results concerning the amount of tolerable faults and the dissemination utilizing randomly placed calls.

Fraigniaud and Lazard [44] provide a similar, yet updated, survey of deterministic broadcasting and deterministic gossiping in several different graphs. These include, amongst others, the complete graph, the torus, the hypercube, the cube-connected cycles, the undirected deBruijn, and the butterfly graph. All graphs are studied with a focus on methods utilizing either half duplex or full duplex transmissions.

The book by **Hromkovic et al. [54]** gives, amongst others, an extensive and excellent overview of deterministic broadcasting and deterministic gossiping (including ideas and methods concerning the development of such algorithms) in different scenarios. The first part of the book focuses on synchronous communication and models. Broadcasting is

studied for degree-bounded networks like cube-connected cycles, butterfly, and deBruijn, whereas gossiping is mainly studied for hypercube-like networks with constant degree and complete graphs. The authors also consider fault tolerance, i.e., permanent link faults, permanent node faults, transmission faults, and probabilistic faults. The second part of the book focuses on similar settings in the case of asynchronous models. That is, deterministic broadcasting in arbitrary networks, tori, and hypercubes with and without fault tolerance.

2.1.1.1 The Random Phone-Call Model: Push-Pull

Dissemination protocols have been studied extensively in the past and many approaches and protocols were invented and analyzed during that time. Two of these are the *push*-model and the *pull*-model. At first, these two models were analyzed separately until Karp et al. [59] proposed the combination of them, the so-called Random Phone-Call model. In the following we present some recent results for variations of the (randomized) Push-Pull protocol. Note that we restrict the presentation to this kind of protocols, since we study a variation of the Random Phone-Call model in this thesis.

Chierichetti et al. **[20]** study randomized rumor spreading in the Preferential Attachment model (cf. Definition 2.2). The authors show that: 1.) regardless of the starting node, the push strategy requires polynomially many rounds to successfully spread the rumor with constant probability, 2.) there are starting nodes, such that the pull strategy requires polynomially many rounds to successfully spread the rumor with constant probability, and 3.) regardless of the starting node, the Push-Pull strategy requires $\mathcal{O}(\log^2 n)$ many rounds to successfully spread the rumor with probability $1 - o(1)$.

Doerr et al. **[28]** study the performance of randomized rumor spreading in the Preferential Attachment model. Their findings, amongst others, refine the one of Chierichetti et al. [20]. To be more specific, they show that the Push-Pull strategy delivers a message to all nodes within $\Theta(\log n)$ rounds with high probability. Further, by modifying the protocol such that contacts are chosen uniformly from all neighbors but the one contacted in the previous round, the necessary amount of rounds to complete the randomized rumor spreading shrinks to $\Theta(\log n / \log \log n)$.

Berenbrink et al. **[14]** study randomized broadcasting in d-regular networks with suitable good expansion properties. The authors present a communication algorithm, i.e., a modification of the Random Phone-Call model, where each node has a (randomly ordered) cyclic list of all of its neighbors. Then in round i each node chooses the ith entry of its list to communicate with. This algorithm provides the following properties: 1.) after at most $\mathcal{O}(\log n)$ many rounds the rumor is spread with high probability, and 2.) during these rounds each node produces $\mathcal{O}(\sqrt{\log n})$ many transmissions on average.

Berenbrink et al. [13] consider random graphs and hypercubes in the setting of the Quasi-Random Phone-Call model, which is a modification of the Random Phone-Call model. Here each node provides a list of neighbors whose order is specified by an adversary. In each round i, starting from a random position, each node opens a channel to its ith neighbor according to his list. Then messages can be exchanged in a bidirectional manner. For random graphs, the authors present an address-oblivious algorithm[1] with runtime $\mathcal{O}(\log n)$, which uses at most $\mathcal{O}(n \log \log n)$ many message transmissions. On the other hand, in the case of hypercubes with dimension $\log n$, they present an address-oblivious algorithm with runtime $\mathcal{O}(\log n)$, which uses at most $\mathcal{O}(n(\log \log n)^2)$ many message transmissions.

In [43] **Fountoulakis et al.** study randomized rumor spreading on random graphs with power law degree distribution, created according to the Chung-Lu model (cf. Definition 2.3), with an arbitrary power law exponent $\beta > 2$. The degree distribution of the resulting graph follows a power law degree distribution if the expected degree sequence used to create the graph followed a power law distribution. A degree distribution follows a power law distribution if the amount of nodes in the network with degree k is proportional to $k^{-\beta}$, for sufficiently large k and β. In this setting the authors show two main results: 1.) if $2 < \beta < 3$, then for any $\epsilon > 0$ the rumor spreads to a $(1-\epsilon)$-fraction of all nodes within $\Theta(\log \log n)$ many rounds with high probability, and 2.) $\Omega(\log n)$ many rounds are necessary in the case of $\beta > 3$.

In [9] **Avin and Elsässer** present a modification of the Random Phone-Call model that uses pointer jumping to spread a rumor in the complete graph. The main idea behind the algorithm is as follows: 1.) use push-transmissions to build a sufficiently large amount of informed nodes, 2.) build a meta structure by choosing a leader, i.e., nodes representing information sources for all its connectors/clients, i.u.r.[2] and connectors, i.e., nodes trying to be informed by connecting to a leader node, and 3.) inform all remaining nodes by the standard Push-Pull method.

The concept of pointer jumping is hereby used in the process of sending the addresses of previous communication partners to current communication partners, thus forcing them to use the most recently received addresses as their next communication partners. Utilizing this method, each node will be connected to at least two leaders in the corresponding phase of the algorithm, enabling the meta structure of leader and connector to efficiently spread the rumor. The authors show that a rumor spreads to all nodes within only $\mathcal{O}(\sqrt{\log n})$ many rounds with high probability under these circumstances. Additionally, their algorithm is able to cope with at most $o(n/2^{\sqrt{\log n}})$ many node failures.

Giakkoupis [47] focuses on the connection between the conductance of a graph and the rate in which a rumor spreads in said graph using the traditional Push-Pull model.

[1] An algorithm is called address-oblivious if the decisions of each node v do not depend on the IDs (i.e., some sort of identifier) of the nodes that v has communicated with in any previous round.

[2] independently uniformly at random

The conductance of a graph is defined as the minimum conductance over all possible cuts $(S, \bar{S})$, whereby the conductance of a cut relates the amount of edges of the cut and the minimum amount of edges incident to the vertices in S and $\bar{S}$. In other words, a connected graph with high conductance is well connected, whereas a graph with low conductance is not. The author shows that rumor spreading in the traditional Push-Pull model is successful within $\mathcal{O}(\phi^{-1} \log n)$ many rounds with high probability for any n-node graph with conductance ϕ.

In [**48**] **Giakkoupis** studies the relationship between the vertex expansion of an arbitrary graph and rumor spreading in the traditional Push-Pull model. The vertex expansion α of a graph describes the minimum ratio of the neighbors that a set of nodes has over the size of the set. The author shows that, with high probability, no more than $\mathcal{O}(\log^2 n/\alpha)$ many rounds are necessary to spread the rumor to all nodes in the graph for any graph with vertex expansion at least α.

2.1.1.2 Power Law Networks/Graphs

The recent appearance of (large) social networks refueled the interest in power law networks and keeps the momentum of these networks going [11]. Power law networks were and still are quite popular and widely used due to their usefulness in a variety of research areas. In the case of the World Wide Web, Albert et al. [4] found that the probability for a web page of the nd.edu domain to have k outgoing links follows a power law distribution. The same holds true for incoming links, i.e., for the amount of web pages linking to a specific one. Newman observed in [66] that many real-world networks such as the Internet, World Wide Web, as well as various social and biological networks follow a power law degree distribution. Aiello et al. [2] demonstrated that the degree sequence of the so-called call graphs, i.e., graphs representing calls handled by a telephony carrier for a specific period of time, can be approximated by a power law distribution. Cohen et al. [23] showed that the network connectivity in scale-free networks remains intact under random node removals. This explains the high robustness of real-world networks against random failures [5]. It was also shown that many peer-to-peer networks appear to follow a power law degree distribution. A quite well-known example is Gnutella [58].

In the case of social networks, Redner [69] discovered that the networks of "actors linked by a movie and scientific papers linked by citations" follow a power law as well. Further, in the context of randomized rumor spreading, social networks following a power law degree distribution possess the property of spreading a rumor very quickly (cf. Section 2.1.1.1). On the other hand, Satorras et al. [67] showed that the epidemic threshold[3] converges to zero in the case of a scale free network. Interestingly, for certain graphs

[3]The epidemic threshold, with respect to a network, defines the point where the effective spreading rate, i.e., the ratio of the rate that nodes are infected and the rate they become susceptible again, is high enough such that the epidemic can spread and become persistent.

with a power law degree distribution, there is no constant threshold for the epidemic outbreak as long as the power law exponent is less than 3 [66] (which is, in fact, the case in most real-world networks, e.g. [40, 1, 7, 72]). Further, the results of Jeong et al. [57] imply that metabolic networks do comply with the main design principles of robust and error-tolerant scale-free networks.

In **[32] Draief and Ganesh** study the coincidence time of susceptible and infected individuals, i.e., the total time during which both individuals are at the same vertex, and the infection probability on various graph classes including the Erdös-Renyi (cf. Definition 2.4) and power law graphs (Chung-Lu Model, cf. Definition 2.3). The model itself is actually quite simple. Let $G = (V, E)$ be a network and P be a set of mobile particles moving on G. Then the movement of each $p \in P$ is characterized by independent continuous-time Markov chains on the finite state space $V = 1, ..., n$.

The authors prove that the coincidence time up to time t is inversely proportional to the number of nodes n in the graph in the case of homogeneous graphs, whereas the coincidence time heavily depends on the power law exponent γ for power law random graphs. For $\gamma > 3$ said coincidence time is $\mathcal{O}(t/n)$, for $\gamma = 3$ it is $\mathcal{O}(\log m/n)$ with m being the maximum expected degree, and for $2 < \gamma < 3$ it scales polynomially in the maximum expected degree and the average expected degree with parameter γ.

Grabowski et al. [49] studied the properties of five different social networks:

- Grono, an Internet community where one can join after a direct invitation only,
- Allseron, a MMORPG[4] including a social network,
- LastFM, a (social) system based on the individuals' musical taste where persons with similar preferences can meet, make friends, and gather into groups,
- XFire, a gamer community enabling inside and outside game chat without quiting a game and a whole system to maintain the individuals' acquaintances, and
- Shelfari, a web system where users may create virtual shelves with books and write recommendations for users with similar taste.

Data was collected for the above social networks over up to five years and further analyzed with respect to the list of friends, groups, creation date, and last login of each individual. Despite some differences in the exponent, mainly due to the type of activity performed in the corresponding network, all these networks could be very well described using a power law.

[4]Massively Multiplayer Online Role-Playing Game

2.1.1.3 Evolving Networks

Graphs and networks have always been highly interesting for the computer science community. We implement them in areas like telecommunications or utilize them as virtual-structures over real ones (e.g., VPNs and social networks). Especially evolving graphs or networks that obey a certain kind of change over time are particularly interesting to us. Various approaches have been developed. For example, Eubank et al. [39] model inter-personal contact patterns at various physical locations within an urban environment (Portland, Oregon) with a bipartite graph. Clementi et al. [21], on the other hand, consider dynamic Markovian graph models transitioning from one state to another depending on the current one. We refer the reader to Section 2.1.2 for related work focusing on simulations.

Obviously, often only a part of the graph, rather than the whole structure, evolves at a specific point in time. However, the main idea remains the same: in each unit of time the graph or network only represents a snapshot of the current situation. Note that a series of these graphs enables us to examine the evolution of the graph over time.

In **[22] Clementi and Pasquale** consider flooding in evolving graphs. An evolving graph is basically a series of graphs, where the graph corresponding to each time step is considered fixed during said time step, built according to some constraints. To be more specific, the authors mainly consider two types of evolving graphs: 1.) edge Markovian evolving graphs, i.e., graphs where each existing edge in a time step t of the graph vanishes in time step $t+1$ with some probability q, whereas each non-existing edge in time step t is created in time step $t+1$ with probability p, and 2.) geometric Markovian evolving graphs, i.e., graphs where all nodes perform a random walk on a two-dimensional grid and an edge between two nodes exists if and only if both nodes are within an euclidean distance of at most r, where r is a parameter.

Amongst many others, the authors obtain the following results: 1.) the flooding time of any edge Markovian evolving graph is bounded by $\mathcal{O}\left({}^{\log n}/_{\log(1+np)}\right)$ with high probability, 2.) starting with an empty edge Markovian evolving graph, i.e., $E_0 = \emptyset$, the flooding time is bounded by $\Omega\left({}^{\log n}/_{np}\right)$ with high probability, 3.) if the resolution coefficient[5] ϵ is at most 1 and $c\sqrt{\log n} \leq r \leq \sqrt{n}$ with a sufficiently large constant c, then the flooding time of the corresponding stationary geometric Markovian evolving graph is bounded by $\mathcal{O}\left({}^{\sqrt{n}}/_{r} + \log\log r\right)$ with high probability, and 4.) for any adaptive worst case adversary there exists a randomized broadcasting protocol such that broadcasting completes within $\mathcal{O}({}^{n^2}/_{\log n})$ time steps with high probability.

Clementi et al. [21] consider the edge Markovian evolving graph with respect to the task of rumor spreading (Push protocol). Hereby an edge Markovian evolving graph is a series of graphs with the following two properties at time step $t \geq 0$: 1.) each edge

[5] Each node can assume only positions whose coordinates are an integer multiple of a resolution coefficient.

that does not exist in time step t appears in time step $t+1$ with probability p, and 2.) each existing edge in time step t vanishes in time step $t+1$ with probability q.

The authors show three main results: 1.) if $1/n \leq p < 1$ and $q \in \Omega(1)$, then the Push protocol informs all nodes within $\Theta(n)$ many time steps, 2.) the completion time of the Push protocol does not exceed $\mathcal{O}\left({}^{\log n}/_{n \cdot \min\{p,1/n\}}\right)$ with high probability in the case of $q = 1 - p$ and arbitrary p, and 3.) the Push protocol finishes successfully in $\mathcal{O}(\log n)$ many time steps for $p = c/n$, for some constant $c > 0$ and arbitrary q.

2.1.2 Social Context

In this section we mainly focus on related work that considers the following topics:

- Social (Interaction) Networks,
- (Urban) Environmental Modeling and Human Behavior, and
- Epidemic Spreading and Countermeasure Approaches.

Similarly as before, all of the topics were extensively analyzed from many different perspectives including medical and biological science (e.g., [75, 63, 25, 57]), physics (e.g., [68, 79, 74]), urban planning (e.g., [27, 6, 71]), and sociology (e.g., [19, 77, 62]). Consequently, we list only a very small selection of the most recent findings below.

2.1.2.1 Social (Interaction) Networks

Social networks are complex systems but crucial for understanding how dissemination processes perform in environments affected by human behavior and interaction and/or communication. This section is dedicated to these types of networks. Note, however, that said networks may not necessarily be based on urban environments.

Doerr et al. [29] focus on the question why rumors spread so quickly in social networks. In short, they simulate an information spreading process in different network topologies and analyze how news spreads in such social networks. These topologies include the Preferential Attachment model (cf. Definition 2.2), the complete graph, the Random Attachment model[6], and samples from the Twitter and Orkut social network. The spreading process itself is modeled by a slight variation of the traditional Push-Pull model: each node chooses a communication partner from its neighborhood uniformly at random but excludes the one used in the last round.

[6] This model is similar to the Preferential Attachment model. The main difference is that each node chooses m neighbors uniformly at random.

The outcome is quite interesting. News spreads significantly faster in existing social-network topologies than in other types of networks. The key to this behavior, according to the authors, lies in the properties of such social networks: namely the power law degree distribution. The authors state that shortcuts between nodes with large degrees are built by the ones with relatively few neighbors, since high degree nodes talk less often to each other directly due to their large neighborhood. That is, a rumor spread over such a shortcut travels from one hub to another in a short amount of time. Each hub then spreads this rumor throughout its large neighborhood rather quickly.

Tripathy et al. [76] study the possibility of combating a rumor (dissemination process) with an anti-rumor in the context of social networks with a process similar to the original one. The authors used data of the Twitter social network and a synthetically generated Watts-Strogatz model for their experiments. The Watts-Strogatz model [80] starts with n nodes distributed evenly on a circle. Then each node creates edges to its k nearest neighbors on either side, with k being a parameter. Now a rewiring process starts. Each edge (u, v) of the graph is replaced with uniform probability p' by an edge (u, v'), where v' is chosen uniformly at random amongst all nodes $V \setminus \{u\}$.

The dissemination process used by the authors is quite simple and starts with 10 randomly infected nodes. Then in every round each infected node tries to infect all its neighbors, whereby each of its neighbors is independently infected with probability p. Concerning the anti-rumor process, two basic strategies are presented: 1.) the delayed start model in which the rumor is discovered after some time and a local authority starts the anti-rumor process, and 2.) the beacon model in which vigilant agents within the network start the anti-rumor process at their position after they discover the rumor. Then the anti-rumor spreads in a similar manner as the rumor itself but with a different probability parameter q. The key feature of this paper is the possibility to combat the rumor with a different message, which in turn can be further disseminated by participants who are not aware of the original rumor themselves. This property is crucial, since actions are not limited to reactive ones, but may also be preemptive.

The authors gain some basic insights. They found that rumors persisted longer in the Twitter network than in the Watts-Strogatz network: i.e., the rumor was harder to combat in the real-world network. Further, the delay between the start of the rumor and its discovery was crucial. The lifetime of the rumor spreading process increased at least linearly with a rising delay.

Eubank et al. [39] explore the use of bipartite graphs to model the physical contact patterns resulting from movements of persons between specific locations. The authors use individual-based simulations incorporating actual census, land-use, and population mobility data and present a case study of Portland, Oregon.

One particular thing they observed is that the number of locations accommodating k different visitors in a unit of time follows a power law with exponent ≈ 2.8. The same holds true for the in and out degree distributions of the locations network: i.e., the

network where a directed edge (i, j) exists if and only if there is at least one person traveling directly from location i to j during the day (unit of time).

The implication is that "attempting to shatter the contact graph by vaccinating the most gregarious people in a population would essentially be equivalent to mass vaccination". Similarly, just closing locations with high capacities or vaccinating all visitors does not lead to an efficient embankment of the epidemic until a large fraction of the population is affected. Additionally, their data suggests that fast detection is necessary if one wants to use targeted vaccination. Then infected individuals would not be able to contaminate locations with a large amount of susceptible persons so easily.

In [**15**] **Bettencourt et al.** study the development of urban environments, especially with respect to their structural and social evolution, trying to identify emerging patterns between all of them. Further, they analytically investigate the consequences of the population size on the pace of life and vice versa. The paper is motivated by the ongoing urbanization of the available landscape all around us. In the year 2000 more than 70% of the population of developed countries lived in cities and around 40% of developing countries. For the year 2030 the estimations are significantly higher. The authors state that we need to understand this "urban expansion" and what it entails. For example, how can we keep improving our health care system and maintain the nutritional supply despite a quickly growing population and how may this impact our economy? These are just a few questions that need to be answered.

The authors present empirical evidence "showing that important demographic, socioeconomic, and behavioral urban indicators are [...] scaling functions of city size that are quantitatively consistent across different nations and times". Said nations include the U.S., China, and a few European countries like Germany. Bettencourt et al. speculated that cities of different size may be only the scaled version of an idealized city, a blueprint so to speak, thus enabling predictions of many kinds (e.g., in terms of growth, flows of resources, traffic management, economical development, and criminal activities).

This hypothesis, with respect to some scaling laws defined by the authors, was then empirically confirmed. These scaling laws include, amongst many others, the amount of new patents, total wages, household electrical consumption, the amount of gasoline stations, and the total road surface. In the analytical section the authors deduce equations describing the urban growth of an artificial city and study its dynamics. The equations show evidence that the pace of urban life increases with city size. In terms of growth (of a city or social organization), there is evidence that said growth will slow down over time and eventually stop as long as it is mainly driven by economies of scale. On the other hand, if the driving force is innovation and wealth creation, then one can sustain continued growth by providing major innovations at an accelerated rate. In other words "open-ended wealth and knowledge creation require the pace of life to increase with organization size and for individuals and institutions to adapt at a continually accelerating rate to avoid stagnation or potential crises".

2.1.2.2 Epidemic Spreading

There is plenty of work considering epidemiological processes in different scenarios and on various networks. One of the simplest model of mathematical disease spreading is the so-called SIR model (see [53, 66], for example). In this model the population is divided into three categories: susceptible (S), meaning all individuals who do not have the disease yet but can become infected; infective (I), meaning all individuals who have the disease and can infect others; and recovered (R), meaning all individuals who recovered and have permanent immunity (or have been removed from the system). Most papers model the spread of epidemics by using a differential equation based on the assumption that any susceptible individual has uniform probability β to become infected from any infective individual. Furthermore, any infected individual recovers at some stochastically constant rate γ.

In this section we mainly present papers considering epidemic spreading in the context of simulations. However, note that the topic of rumor spreading is closely related to the one we study here. Therefore, many of the results cited in Section 2.1.2.1 and Section 2.1.1 can be transfered to the context of epidemic spreading.

In reality the spread of an epidemic is influenced by many factors. One of these factors is the awareness of the affected individuals about the epidemic. In **[45] Funk et al.** focused on this type of (natural) countermeasure and analyzed the spread of awareness on epidemic outbreaks. That is, in addition to the disease, the information about the epidemic is also spread in the network and has its own dynamics. Their analysis contains results considering a fully mixed model (mean field analysis) as well as an analysis from the point of view of a node with common properties (such as average node degree, for example). The authors describe the spreading scenario by the following model. Each individual has a level of awareness, which depends on the number of hops the information has passed before arriving to said individual. This is combined with the traditional SIR model. The authors show that the spread of awareness can result in a slower outbreak in a well-mixed population, however, it will not change the threshold of the epidemic. Nevertheless, if the spread of awareness about a disease is considered in certain (geometric) networks as a local effect in the proximity of an outbreak, then such an awareness process can completely stop the epidemic. Moreover, the impact of spreading awareness is even intensified if the social network of infections and informations overlap.

In **[55] Iozzi et al.** present an approach to compute a variety of synthetic age-specific contact matrices using individual-based simulations. Hereby an entry $[i, j]$ of a contact matrix represents the average number of contacts between individuals of age group i and j per unit of time. Behavioral patterns, i.e., activities performed in a specific place at a specific time, were provided by the Italian Time Use survey that was carried out by the Italian National Statistical Agency in 2002 and 2003. This survey sampled over 55 thousand individuals in more than 21 thousand households. Additionally,

socio-demographic data was used to improve their model with respect to: 1.) family size and composition, 2.) company size with respect to the number of employees, and 3.) school class size.

Utilizing this data, the authors created an artificial society with properties similar to the real one. However, a few assumptions were made. For example, only individuals with data collected over an entire (average) work day were considered, weekend days were ignored and the environment was set to a grid of size 150×150. Households, schools, the university, and workplaces then were randomly allocated on this grid structure. Despite the very detailed (anonymized) data that was available to construct the artificial society, information regarding the specific places occupied by specific individuals at a certain time were (mostly) still lacking. Consequently, the agents' household, working place, and so on were partially determined at random.

The authors deduced different contact matrices representing the average time of a contact, the average number of repetition of contacts, and the average number of different persons contacted by simulating the aforementioned artificial society. It was observed that these contact matrices did not fit the serological data for Varicella as well as they did with the serological data for ParvoVirus.

Carrat et al. [**18**] developed a model to simulate the spread of the Influenza virus in a large population or community. Their model contains two main components: 1.) individual parameters, i.e., parameters like the risk of infection, treatment, and vaccination status, and 2.) community parameters represented by a (suitable) random graph modeling contacts between individuals. All data regarding the infection and health-care use was mostly deduced from previously conducted experimental studies or set according to the authors' best judgment.

For example, the most symptomatic subjects would seek medical advice at a 90% rate, which is a higher rate compared to other studies due to a potentially higher public awareness during a pandemic situation.

The community model, on the other hand, incorporates a demographic profile based on the French national census data. Consequently, the artificial society contains households, workplaces, and other places of occupation, which in turn were assigned to specific districts. Contacts in households were modeled by a complete graph, whereas contacts in schools and all other places were modeled by a Preferential Attachment random graph. The final calibration of the system was performed using serological data of 1.) the Asian Influenza during 1957, 2.) the Spanish flu of 1918, and 3.) the Hong Kong flu of 1968.

Each scenario was simulated 200 times with 10,000 agents in each simulation run. The results show that the combination of measures focusing on reducing the meeting frequency and virus transmissibility have the potential to mitigate the effects of a pandemic similar to the ones experienced in the 20th century. It is very interesting that especially non-pharmaceutical countermeasures like school and workplace closures had

a significant impact, while the effect of pharmaceutical countermeasures such as the treatment with antivirals was almost negligible.

Borgs et al. [**17**] focused on how to distribute antidotes to control epidemics. The authors analyzed a variant of the contact process in the *susceptible-infected-susceptible model (SIS model)* on a finite graph in which the cure rate is allowed to vary among the vertices. That is, the rate ρ_v at which an infected node v becomes healthy is proportional to the amount of antidote it receives, given a fixed amount of antidote $R := \sum_{x \in V} \rho_x$ for the whole network.

The authors obtain three main results: 1.) contact tracing, meaning the identification and diagnosis of individuals that had contact with an infected individual, as a strategy to deploy an antidote requires a total amount of said antidote which is super-linear in the number of vertices, and therefore prove to be inefficient, 2.) on a general graph with bounded average degree, a simple strategy distributing the antidote proportional to the degree results in an epidemic threshold[3] bounded below by a strictly positive number, and 3.) for expander graphs, meaning graphs with strong connectivity properties, the total amount of antidote needed by the strategy of 2) cannot be reduced by more than a constant factor.

Valler et al. [**78**] examined the epidemic threshold[3] for a mobile ad-hoc network in a 2D field. They showed that if the connections between devices are defined by a sequence of matrices $A_1, \dots A_T$, then no epidemic outbreak occurs for $\lambda_S < 1$, where λ_S is the first eigenvalue of $\Pi_{i=1}^{T}(1-\delta)I + \beta A_i$, with high probability. Here β and δ are the virus transmission probability and the virus death probability, respectively. They also approximated the epidemic threshold for different mobility models in a predefined 2D area.

Hatchett et al. [**51**] tested a hypothesis that the early activation of multiple interventions could be associated with reduced disease transmission. The authors used detailed timing records for about 19 classes of non-pharmaceutical interventions of 17 U.S. cities during the pandemic of the Spanish flu in 1918 for this purpose. These interventions, amongst others, included: 1.) making Influenza a notifiable disease, 2.) isolation policies, 3.) school closures, 4.) bans on public gatherings, and 5.) community-wide business closures.

Intuitively, the early implementation of aggressive interventions was significantly associated with a lower peak of excess mortality. Especially early closures of schools, churches, theaters and the ban on public gatherings had a significantly positive impact. However, the obtained data also suggests that the viral spread may be renewed upon relaxation of such measures in the current or even the subsequent waves. To be more specific, the authors found a statistically significant inverse correlation of the height of the first and second peak. Consequently, cities with lower peaks during the first wave were at greater risk of being impacted by a large second wave. Additionally, cities with

lower peak mortality rates during the first wave showed a tendency of experiencing a second wave after a shorter time.

In short, the data suggests that cities applying the most effective interventions for a longer time were not affected by subsequent waves and suffered a significantly lower peak of excess mortality. On the other hand, applying interventions only for a short time resulted in a lower peak of excess mortality for a short time whilst provoking a higher peak of excess mortality in the second wave.

The mitigation of an influenza pandemic on a national scope is the main focus of **Ferguson et al.** [**41**]. They examined the effect of different interventions, including national and international travel restrictions, within the U.S. and Great Britain on an influenza pandemic via simulations. Their individual-based model was created using high-resolution density data (landscan global population data provided by the Oakridge National Laboratory), data on travel patterns, and seeding of infection via international travel. However, national air travel was explicitly modeled for the United States only and a great deal of their parameters were adjusted according to (more or less good) estimations or assumptions. For example, 30% of the epidemic transmissions occur within the household and 70% outside, whereas the authors assume that 33% of the outside transmissions occur in the general community and 37% in schools and workplaces. Another example is that 50% of the infected cases are assumed to be ill enough to be classified as clinical cases. Finally, the system was calibrated using serological data of the Spanish flu of 1918.

The authors made one major assumption with respect to the epidemic: it must be imported via an international flight. Therefore, estimations of the expected number of infected individuals arriving in the corresponding nation on a daily basis were used for the seeding process. Consequently, the epidemic could only spread within the corresponding nation if and only if at least one infected individual was granted access to said nation.

Since both countries have a significantly different size, the spreading patterns were different as well. The infection did not seem to follow a strict pattern in Great Britain. However, a pattern emerged in the case of the U.S. There the epidemic first focused on the area around the seed infections, followed by an almost homogeneously distribution across the rest of the country. As a direct consequence of the seeding process, the reduction of inbound travelers could only delay the spread of the epidemic but was not able to stop it. Even a 99.9% reduction of inbound travelers could only delay the epidemic for at most 6 weeks.

Additionally, the data suggests that treatment and isolation on an individual basis may help significantly only if these actions are performed rapidly, while school or workplace closures always reduced the peak of attack rates substantially. On the other hand, socially targeted prophylaxis targeting classmates or close work colleagues showed a great impact. In terms of vaccination, the authors found that a staged vaccination

program was most promising if children were vaccinated first, while stockpiling vaccines had a substantial impact only if a very high amount was available (for at least 20% of the population).

Longini et al. [62] examine antiviral prophylaxis in the rural southeast Asia. They consider two strategies: 1.) targeted antiviral prophylaxis (TAP), meaning that identified index cases (first symptomatic illness cases in a mixing group) are treated and only close contacts (e.g., households and neighborhood clusters) to these index cases are offered a prophylaxis, and 2.) geographically targeted antiviral prophylaxis (GTAP), meaning a specific percentage of the population in the entire locality of the index case, varied in a sensitivity analysis, are given one course of antiviral medicine (the authors used efficacy estimates of Oseltamivir).

The model in use represents a population of 500,000 persons in 36 geographic localities across 5,625 km^2. Initial data with respect to the age and household size distributions of the population is based on the Thai 2000 census. The social study of the Nang Rong District in rural Thailand is used to determine mixing group sizes and distributions. Social interactions, and therefore the spread of the epidemic, are modeled by contact group mixing. Close contact groups include households, preschool groups and workplaces, for example, whereas casual contact groups consist of markets, temples, and a hospital. Each individual may mix in its household or within clusters of households. Additionally, children may mix in (different) schools, and regular adults in their workplaces. The authors used a system calibration with parameters in between the ones observed in the Asian Influenza during 1957 and the Hong Kong flu of 1968, targeting for an overall illness attack rate of 33% (similar to the one observed during the Hong Kong flu).

The results indicate that targeted antiviral prophylaxis, if applied fast enough, is highly effective for lower basic reproduction numbers[7] ($R_0 \leq 1.4$), whereas higher values of the basic reproduction number make the combination with quarantines or pre-vaccination necessary, for example. On the other hand, not much of a difference could be found between the two strategies (TAP and GTAP) themselves. The real differences were due to the combination of TAP and other countermeasure approaches.

Del Valle et al. [77] present a method to estimate transmission matrices based on a synthetic population. The authors generate and simulate a synthetic population of Portland and analyze the resulting social network with special interest in mixing patterns as a case study. Census data was used to generate a synthetic population with the correct demographics (e.g., age distribution or population density). The city itself was divided into 181,230 locations (e.g., households or workplaces). For each household within the synthetic population an appropriate activity pattern was selected according to the surveys. Such a pattern specifies the type of activity performed by each individual in a household including the location of said activity. The authors defined nine different types of activities: home, work, shopping, visiting, social recreation, passenger server,

[7]The basic reproduction number defines the amount of cases in an otherwise fully susceptible population that one case generates on average over the course of its infectious period.

school, college, and other. Business locations were included according to official business directory data and activity schedules were generated based on the National Household Transportation Survey data.

However, individuals at the same location are not necessarily participating in the same activity or are even close enough to spread an infection. Therefore, each location was further subdivided into smaller units called rooms. The various activities then take place in these rooms and the disease can only be spread within the same room.

The results indicate two main blocks of mixing: 1.) young individuals, meaning individuals of an age less than 20 years, and 2.) adults, meaning individuals at least 20 years of age. However, children showed a weak coupling between themselves and middle-aged adults, which is most likely due to parenthood and school attendances. Consequently, school children turn out to be more susceptible than the rest of the population.

2.2 Different Types of Simulations

In addition to the papers described above, different approaches were presented focusing on how to model the spread of epidemics and the ability to simulate more efficiently. Two of the most popular directions are the so-called agent-based (also called individual-based) and structured meta-population-based approach, respectively (cf. [3, 56], for example). Both approaches have their advantages and weaknesses. The main idea of the meta-population-based approach is to model entire regions, e.g., georeferenced census areas around airport hubs [10], and connect them by a mobility network. Then the spread of an epidemic is analyzed within these regions by using the well-known mean field theory. The agent-based approach models individuals with agents in order to simulate their behavior. In this context, the agents may be defined very precisely, including, e.g., race, gender, educational level, etc. [60, 61], and thus provide a huge amount of detailed data conditioned on the agents' setting. Furthermore, this kind of model is also able to integrate different locations like schools, theaters, and so on. Thus, an agent may or may not be infected depending on his own choices and the ones made by agents in his vicinity. The main issue of the agent-based approach is the huge amount of computational capacity that is required to simulate large cities, continents or even the world itself [3]. This limitation can be attenuated by reducing the number of agents, which then entails a decreasing accuracy of the simulation. In the meta-population approach the simulation costs are lower, but one does have to sacrifice accuracy and some kind of noncollectable data to achieve this.

To combine the advantages of both systems, hybrid environments were implemented (e.g., [16]). The main idea of such systems is to use an agent-based approach at the beginning of the simulation up to some point where a suitable amount of agents is infected. Then the system switches to a meta-population-based approach. Certainly, such a system

combines the high accuracy of the agent-based simulations at the beginning of the procedure with the faster simulation speed of the meta-population-based approach at stages in which both systems may provide similar predictions. In [16] the authors defined the situation to switch between both approaches (in both directions) by a threshold T describing a specific amount of infected agents. They compared average epidemic trajectories produced by both approaches and determined which threshold value (if any) results in equivalent average trajectories. For certain epidemics, especially unknown ones, this adjustment may be difficult or even not feasible at all.

Parts of this chapter have been published in:

[35, 36, 37]

Chapter 3

Our Contribution

Our results contribute to almost all of the fields we referred to in Chapter 2 in at least one way. In the following we describe these results briefly.

In **Section 5.1** we consider a modification of the Random Phone-Call model on random power law graphs with n nodes. Our graphs are chosen uniformly at random from the space of all (simple) graphs with power law degree distribution, where the smallest degree is $\delta \geq \gamma(\log\log n)^2$ with γ being a constant. In each round, every node is allowed to call on ρ different neighbors chosen uniformly at random, and to establish communication channels to these nodes. The channels can be used for bi-directional communication, i.e., each node is allowed to send messages over the channels incident to it. At the end of each round, the channels are closed.

We show that there is an algorithm in this communication model that completes broadcasting in time $O(\log n)$ by using $O(n \log\log n)$ many transmissions of the message with high probability. That is, by allowing each node to call on a constant number of different neighbors instead of one, the average number of message transmissions decreases significantly compared to the Random Phone-Call model (cf. [34]). In the case of random power law graphs our result matches the bounds found in [13] for hypercubes and random graphs, but our algorithm is not as efficient as the ones (using more complex methods and/or assumptions) presented in [14, 9], for example.

In **Section 5.2** we show that an infection disseminated by a malicious dissemination process $\mathcal{V}$ does not infect more than a $o(1)$-fraction of a network with power law degree distribution with high probability before it is eliminated by a counteracting dissemination process Φ under the assumption of a constant delay d and a minimum degree of $\Omega(\log\log n)$, where n denotes the amount of overall nodes in the network. Hereby, $\mathcal{V}$ spreads over a constant amount of randomly chosen edges for each infected node, whereas Φ performs its actions for each infected node independently after some delay of d rounds. W.l.o.g. let v be such a node and let i be the round Φ is activated on v. Then v is immunized and notifies all its neighbors about $\mathcal{V}$ in round i, i.e., in the same round. Consequently, all of v's yet uninformed neighbors become immunized as well. However, all v's already infected neighbors remain infected. Our result implies

that an unbounded but relatively small degree is sufficient to prevent almost the entire population of getting infected in this model.

Secondly, in the case where the minimum degree is limited by a constant, we show that a more complex counteracting process called Λ eliminates a malicious piece of information spread by $\mathcal{V}$. Similarly to process Φ, Λ is activated on an infected node after a constant delay d. Then each newly activated node forms a new notification tree with root node v. In short, a notification tree is a tree-like structure, starting with the root node, representing a "who-notified-whom-relationship" of the contained nodes. The maximum path length of each notification tree from its root to its leaves is limited by $\beta \left\lceil \log^2(\deg_v) \right\rceil$, where $\deg_v$ denotes the degree of the root node v and $\beta \in \Theta(1)$. We show that process $\mathcal{V}$ does not infect more than $o(n)$ many nodes while being counteracted by Λ with high probability. This can be interpreted as follows. Due to the social structure, messages from individuals with significant influence are treated with a high priority. Hence, they initiate a tree structure in which each level represents the number of hops a message traverses from the source of the message to the corresponding nodes.

The main idea behind our models is similar to the one of [45] and [76], especially the part considering the spread of awareness. However, we explicitly consider social network-like graphs with power law degree distribution instead of a fully mixed model (mean field analysis) or a common-node-analysis, i.e., an analysis of the situation an average node (with common properties like the average degree, for example) would face.

In **Section 6.1** we consider an epidemic process operating in a dynamically changing environment. This environment represents an urban topology inhabited by n mobile susceptible individuals. Further, we integrate the results of [39] and assume that in each round every individual chooses a location independently at random according to a power law distribution. This is the first analytical result on an epidemic process in a dynamic scenario where the impact of the power law distribution describing the attractiveness of the locations is considered. We combine the standard SIR-model (cf. Section 2.1.2.2) with the following two different society models:

1. the regular model (REG), where the behavior of the susceptible entities is not influenced by an epidemic outbreak in any way, and

2. the restricted model ($REST$), where an epidemic outbreak does influence the behavioral pattern of each susceptible entity.

Using these models we obtain two main results.

First, we show that a polynomial fraction of the individuals in the regular society model remains uninfected even if they are allowed to transmit the infection for $o(\log n)$ many rounds. That is, it is very unlikely for a (deadly) epidemic to wipe out the entire population. This holds due to the decreasing number of survivors and infected individuals over time. That is, the remaining population size is decreasing over time

while the available space for each individual is increasing. Consequently, the probability for a healthy individual to avoid the infected ones increases over time.

Secondly, we show that only a polylogarithmic amount of individuals in the restricted society model will be infected until the epidemic process is stopped. This is achieved within $\mathcal{O}((\log \log n)^4)$ rounds with probability $1 - o(1)$. Here the individuals are allowed to transmit the infection only for a constant number of rounds. Additionally, the power law exponent describing the attractiveness of all target places increases up to a large constant.

In **Section 6.2** we analyze the model of Section 6.1 on a large scale using simulations. We show that the accuracy of the obtained results heavily depends on the number of agents in use and conjecture that this is a problem in general. Note that, depending on the epidemic, different calibrations may lead to different results or may not be feasible at all.

The second main result of this section is that one can approximate the effect of some non-pharmaceutical countermeasures, which are usually adopted if an epidemic outbreak occurs, by setting the parameters properly. This observation is supported by the empirical study of [64]. Interestingly, the right choices of parameters in our experiments seem to be in line with previous observations in the real world (e.g., the right power law exponent seems to be in the range of 2.6-2.9, cf. [39]). To analyze the effect of the countermeasures mentioned above, we integrate the corresponding mechanisms on a smaller scale, and subsequently verify their impact on a larger scale too.

Chapter 4

Preliminary Remarks and Basic Definitions

In this section we provide some commonly used notation, principles, and methods needed for the upcoming analyses. References are given for all statements that are not proven here.

We start with the Configuration Model, which is heavily used in the upcoming analyses. This model basically enables us to create a random graph with respect to a given degree distribution. The Principle of Deferred Decision extends this feature by letting the random graph evolve during the formal analysis.

Definition 4.1 (Configuration Model)**.** *Let $G = (V, E)$, with $V = \{v_1, ..., v_n\}$, be a random graph with a given degree distribution* $\deg = \{\deg_{v_1}, ..., \deg_{v_n}\}$*, where the degree of a node v is denoted by deg_v. Further, each node $v \in V$ provides exactly* $\deg_v$ *many* stubs*. Edges are continuously created in so-called rounds by connecting two stubs in the following manner. We start with an empty graph, i.e., with $E = \emptyset$. A new edge is created by choosing two stubs (i.e., the connectors of the two nodes incident to the corresponding edge) uniformly at random from all yet free ones. Hereby we say that a stub of a node v is free if said stub was not chosen before (i.e., it is not connected to an already existing edge). The above procedure is repeated until all edges are created.*

More formally, one can define the procedure as follows. Let E_i be the set of edges present in round i. Further, let the stubs of a node u that were already chosen up to the corresponding point in time in round i be defined as $|\Gamma_i(u)|$. Then, in each round $i \geq 0$, a new edge $e = \{v, w\}$ is added to E_i by choosing the endpoints v, w of e with probability proportional to $\frac{\deg_v - |\Gamma_i(v)|}{2m - \sum_{\forall v \in V} |\Gamma_i(v)|}$ and $\frac{\deg_w - |\Gamma_i(w)|}{2m - \sum_{\forall v \in V} |\Gamma_i(v)|}$, respectively. Further, it holds that the amount of edges is given by $m = 1/2 \sum_{\forall v \in V} \deg_v$. Consequently, the graph is completely constructed after exactly m rounds.

In the context of this thesis the following principle is used with respect to random graphs only. Therefore, the definition is tailored to the use case of a random graph.

Definition 4.2 (Principle of Deferred Decision)**.** *Let $G = (V, E)$ be a graph whose edges are created by an arbitrary random process P. Then, instead of fixing the entire graph in advance, we let the random choices of P unfold during the analysis.*

The next two lemmas state two standard Chernoff bounds, one lower and one upper bound. The definition and corresponding proof can be found in [65].

Lemma 4.1 (Chernoff Bounds - Upper Bound)**.** *Let $X_1, X_2, ..., X_n$ be independent Poisson trials such that, for $1 \leq i \leq n$, $\Pr[X_i = 1] = p_i$, where $0 < p_i < 1$. Then, for $X = \sum_{i=1}^{n} X_i, \mu = \mathbb{E}[X] = \sum_{i=1}^{n} p_i$, and any $\delta > 0$,*

$$\Pr[X > (1+\delta)\mu] < \left[\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}}\right]^{\mu}.$$

Lemma 4.2 (Chernoff Bounds - Lower Bound)**.** *Let $X_1, X_2, ..., X_n$ be independent Poisson trials such that, for $1 \leq i \leq n$, $\Pr[X_i = 1] = p_i$, where $0 < p_i < 1$. Then, for $X = \sum_{i=1}^{n} X_i, \mu = \mathbb{E}[X] = \sum_{i=1}^{n} p_i$, and any $0 < \delta \leq 1$,*

$$\Pr[X < (1-\delta)\mu] < e^{-\mu\delta^2/2}.$$

Occasionally, we make use of the following slightly different formulation.

Lemma 4.3. *Let $X = \sum_{i=1}^{n} X_i$ where $X_1, ..., X_n$ are independent Bernoulli random variables and let be $0 < \epsilon < 1$. If $0 < \mu \leq \mathbb{E}[X]$, then it holds*

$$\Pr(X \leq (1-\epsilon)\mu) \leq e^{-\frac{\epsilon^2}{2}\cdot\mu} \text{ and } \Pr(X \geq (1+\epsilon)\mu) \leq e^{-\frac{\epsilon^2}{3}\cdot\mu}.$$

Although Chernoff Bounds are an important and mighty tool, they are often inferior to Martingale techniques whenever the independence of the random events is not guaranteed. We often use the Vertex Exposure Martingale combined with the Azuma Hoeffding Inequality in such a case.

Definition 4.3 (Vertex Exposure Martingale)**.** *Let G be a random graph with vertex set $V = \{1, ..., n\}$ and let f be any real valued function defined over the space of all graphs. Let E_i be the subgraph of G induced by the first i vertices. Further, let $Y_i = \mathbb{E}[f(G) \mid E_1, ..., E_i]$, for $1 \leq i \leq n$, and $Y_0 = \mathbb{E}[f(G)]$. Then the sequence $Y_0, ..., Y_n$ forms a martingale.*

The proof of the following lemma can be found in [65].

Lemma 4.4 (Azuma (Hoeffding) Inequality)**.** *Let $X_0, X_1, ...$ be a martingale sequence such that for each k, $|X_k - X_{k-1}| \leq c_k$, where c_k may depend on k. Then, for all $t \geq 0$ and any $\lambda > 0$,*

$$\Pr[|X_t - X_0| \geq \lambda] \leq 2\exp\left(-\frac{\lambda^2}{2\sum_{k=1}^{t} c_k^2}\right).$$

The following Lemma states a basic insight with respect to a set of (not necessarily independent) random events.

Lemma 4.5 (Union Bound)**.** *Let $A_1, ..., A_n$ be an arbitrary countable set of events. Then it holds that*

$$\Pr\left(\bigcup_{i=1}^{n} A_i\right) \leq \sum_{i=1}^{n} \Pr(A_i).$$

Proof. Using an inductive argument, the claim follows directly from the two following facts:

$$\Pr\left(\bigcup_{i=1}^{n} A_i\right) = \Pr\left(\bigcup_{i=1}^{n-1} A_i\right) + \Pr(A_n) - \Pr\left(\bigcup_{i=1}^{n-1} (A_i \cap A_n)\right)$$

and

$$\Pr\left(\bigcup_{i=1}^{n-1} (A_i \cap A_n)\right) \geq 0.$$

□

Parts of this chapter have been published in:

[35]

Chapter 5 The Static Case

Throughout this chapter we focus on information dissemination in *non-dynamic networks*. By definition, these networks do not evolve over time. However, we slightly relax this constraint by including networks that are subject to only minor changes over time. Consequently, the variety of considerable networks increases drastically and more interesting scenarios and use cases emerge. In the following we first present some examples of the aforementioned networks. Afterwards, we give a short outline of the upcoming sections of this chapter.

To present some examples of networks that fit in the context of this chapter, we follow the same approach that we used in Chapter 2. That is, we categorize said examples into:

- Technical Context, and
- Social Context.

We begin with the technical context. Recall, said context primarily covers all types of networks and (relevant) methods from a technological point of view. Since the main interest of this chapter is to examine information dissemination protocols and processes in (mostly) static networks, almost all hard wired networks such as telecommunication networks and the Internet belong here. However, if we consider fixed access points/antennas only, then also wireless networks are of interest to us. Due to their nature, such networks mostly do not change that much over time. A significant change on a daily basis would imply huge expenses. We therefore assume that the change experienced by all of the considered networks is very limited over large time intervals.

The second kind of networks that we are interested in are the ones incorporating a social component. Such networks often experience some kind of change over time. Acquaintance-based networks such as Facebook or Twitter are natural examples. However, since we do not consider dynamic networks in this chapter, we assume that these networks are subject to only a negligible change over a large period of time. For instance, in Twitter or Facebook one can often observe slight (local) changes only within each interval of time (e.g., a work day). This implies that the changes on a daily basis become more subtle as soon as a participant achieves a certain amount of contacts.

A similar property holds with respect to the contacts in our phone book, for instance. Even if we add or remove contacts, the overall change will be negligible provided our contact list is already large enough. This is especially interesting in the context of epidemics. Acquaintance-based networks are of great importance here, since they can be used to model real-world (regular) contacts and to analyze their implications. Consider a group of persons who meet together on a regular basis on the football field in order to compete. Clearly, such habits may be exploited by an epidemic. On the other hand, also countermeasures may benefit from an acquaintance-based network. In order to see this, consider an ordinary family with one child attending the kindergarten. Let us suppose that there is an epidemic outbreak (like the chicken pox, for example) in said kindergarten. Most likely the child's parents will be worried enough to spread the news about the outbreak, and therefore a warning thereof, by phone. This simple example highlights the importance of acquaintance-based networks.

The rest of this chapter is structured as follows. In Section 5.1 we consider the spread of a piece of information in a static network with power law degree distribution. We propose a distributed broadcasting protocol and show that the propagation speed is not worse than the one achieved in d-regular graphs. This implies that the nodes with large degree in such a network, which are a minority, do not automatically absorb the majority of all message transmissions, thus leaving the rest of the nodes compulsorily without a chance. On the contrary, our results indicate that these nodes act as information sources that uninformed nodes (with low degree) can communicate with to speed up the dissemination.

In Section 5.2 then we examine different types of counteracting dissemination processes to neutralize a malicious rumor in the same type of network as before. Depending on the network's properties, one of two different counteracting dissemination processes is used to counteract the malicious one. We observe that simple counteracting strategies can be very powerful, given the average degree of the graph is large enough. For graphs with only a (very low) constant average degree, on the other hand, far more complicated strategies are needed. This is surprising, since it implies that a counteracting strategy focusing on the hubs only may not be the best or the fastest solution.

5.1 Propagation Speed in Static Power Law Graphs

The main purpose of this section is to estimate the propagation speed of a piece of information M in a static network with n nodes of the type we introduced in the preamble of this chapter. We model the communication within these types of networks using a simple modification of the Random Phone-Call model. Our graphs are chosen uniformly at random from the space of all graphs with power law degree distribution, where the smallest degree is in $\Theta((\log\log n)^2)$. Formally, if $\gamma(\log\log n)^2$ denotes the smallest degree, with γ being a constant, then the fraction of nodes with degree $\deg > \gamma(\log\log n)^2$ is proportional to $(\deg -\gamma(\log\log n)^2)^{-\alpha}$, where $\alpha > 3$ is the so-called power law exponent. Additionally, we assume that the largest degree in the network is $n^{1/\alpha} + \gamma(\log\log n)^2$. That is, every node that is originally assigned some degree larger than $n^{1/\alpha} + \gamma(\log\log n)^2$ (according to the distribution above) will have degree $n^{1/\alpha} + \gamma(\log\log n)^2$.

Each node in the graph executes a specific dissemination protocol in a synchronous time model. However, only nodes connected by an edge may communicate with each other. To be more specific, every node calls on ρ different neighbors chosen uniformly at random, and establishes communication channels to these nodes. These channels then can be used for bi-directional communication, i.e., each node is allowed to send messages over the channels incident to it. At the end of each round, the channels are closed. Note that this procedure is similar to the one presented in [12, 38].

In the following we present an algorithm in this communication model that broadcasts a message M to all n nodes in time $O(\log n)$ by using $O(n\log\log n)$ many transmissions with high probability.

5.1.1 Theoretical Model and Algorithmic Framework

The outline of this section is the following. First, we describe our model in more detail and give some basic definitions. Then, we state some properties entailed by said model. Finally, we introduce our dissemination/broadcasting algorithm.

The Network We model the network used by the message dissemination process as an undirected random graph $G = (V, E)$ with n nodes and m edges. The degrees of the nodes are distributed according to a power law. To be more specific, let $\deg_v$ denote the degree of a node $v \in V$. Then v has degree $\deg_v > \gamma(\log\log n)^2$, where γ is some large constant, with probability proportional to $(\deg_v -\gamma(\log\log n)^2)^{-\alpha}$. Hereby the constant $\alpha > 3$ is the so-called power law exponent. In the case that a node is originally assigned some degree $\deg > n^{1/\alpha} + \gamma(\log\log n)^2$ according to the distribution above, the degree of this node is set to $n^{1/\alpha} + \gamma(\log\log n)^2$. By the above arguments it follows

that $m = 1/2 \sum_{\forall v \in V} \deg_v = \Theta(n \cdot \delta)$, where $\delta \in \Theta((\log \log n)^2)$ denotes the minimum degree in G.

Edges in G are randomly created according to the *Configuration Model* (cf. Definition 4.1) and the *Principle of Deferred Decision* (cf. Definition 4.2). Note that the application of the Configuration Model may generate self-loops, disconnected components, and multiple edges. However, since $\delta \in \Theta((\log \log n)^2)$ and $\alpha > 3$, G will be connected with high probability with at most a negligible amount of self loops and multiple edges. Additionally, all nodes have access to a global clock and work synchronously. Further, we assume that every node knows n.

The nodes represent entities executing the broadcasting algorithm described below. To keep track of the dissemination of each message M, we need some auxiliary constructs. According to their state of knowledge, we distinguish between two sets of nodes:

- $\mathcal{U}_M$ (uninformed/susceptible), meaning the set of nodes that do not know M yet, and
- $\mathcal{I}_M$ (informed/infected), meaning the set of nodes that already received message M.

The corresponding sets in a specific round i are denoted by $\mathcal{U}_M(i)$ and $\mathcal{I}_M(i)$, respectively. Let $\mathcal{I}_M^+(i) := \mathcal{I}_M(i) \setminus \mathcal{I}_M(i-1)$ be the set of nodes that became informed in round $i-1$ for the first time. Let G_j be the set (also called group) containing all nodes with degree $2^j\delta$ up to $2^{j+1}\delta - 1$ and let $|G_j|$ denote the size of G_j. Further, let $\mathcal{I}_{M,G_k}(i)$ be the set of informed nodes in G_k in round i with respect to M and let $\mathcal{U}_{M,G_k}(i)$ be the corresponding set of uninformed nodes. Let $\Gamma(v) := \{u \in V \mid \{v,u\} \in E\}$ be v's neighborhood. Then $\mathcal{U}_{M,l}(i) := \{v \in \mathcal{U}_M(i) \mid |\Gamma(v) \cap \mathcal{U}_M(i)| \geq l\}$ denotes the set of uninformed nodes with at least l neighbors in $\mathcal{U}_M(i)$. Since our analysis focuses on one specific message, we will omit the index M in the upcoming proofs. If it is clear from the context, we simply use $\mathcal{U}$ and $\mathcal{I}$, respectively.

The Broadcasting Algorithm Each node in the network represents an entity executing a communication protocol with respect to currently available data/messages on this network. In the following we present the communication model used by the nodes. Our algorithm is based on the Random Phone-Call model introduced by Karp et al. [59]. In the Random Phone-Call model, in every round each node calls on a neighbor chosen uniformly at random, and opens a communication channel to this node. These channels then can be used for bidirectional communication in the corresponding round. We distinguish between two types of transmissions depending on the direction of the communication: 1.) *push*, and 2.) *pull*.
Let node $v \in V$ be the calling node and $w \in V$ the callee one. Then v actively sends a message by itself to w in the case of a push transmission. In contrast, v actively requests a (so far unknown) message from w in the case of a pull transmission. Node w

responds, if possible, by sending a message back to v. In other words, the calling node is the sending one in a push transmission and the receiver in a pull transmission.

We slightly modify this procedure for our protocol (Algorithm 5.1). In each round every node $v \in V$ calls on ρ *different* neighbors $\{w_1, ..., w_\rho\}$ instead of one. Then v establishes a communication channel to each of these nodes. As soon as a channel is established between a pair of nodes, both of them are allowed to send data over this channel. Hereby, we differentiate between vital messages, meaning actual push or pull transmissions with respect to genuine data, and auxiliary messages, meaning messages needed to establish, maintain, and close communication channels (such as messages during the handshake phase, for example). Further, we assume that the message size is unbounded.

In order to estimate the cost of our broadcasting algorithm, communication transmissions are counted. However, only vital messages are considered for the communication overhead. In other words, transmissions needed to establish, maintain, or close a communication channel are not considered as a part of the total number of message transmissions. This is motivated by the following reason. Suppose we operate a large network where all nodes frequently generate new data (i.e., messages to be transmitted). To disseminate this newly generated data to each individual node in the network, communication channels need to be created anyway. Therefore, the costs induced by communication channels amortize with respect to the cost of transmitting messages [59].

Our algorithm is *distributed*, i.e., the nodes use only local knowledge to make the decision whether or not to send a message over an open channel. This local knowledge might be the age and number of broadcasted messages so far, the times at which earlier messages arrived, or the node's identifier, for example. Our algorithm is a so-called address-oblivious algorithm. An algorithm is called address-oblivious if the decisions of each node v do not depend on the IDs of the nodes that v was connected to in any previous round via an open channel. In other words, nodes are not allowed to remember or keep track of the nodes they have already communicated with in the past (cf. [59]).

Table 5.1 states some procedures that any node $v \in V$ may perform in a single round. Clearly, our goal is to broadcast a (vital) message M. In order to reduce the amount of vital messages, each node combines all its currently stored data and transmits it in one message only. Further, we assume that each transmitted message contains a timestamp $M.t$ indicating its payload's (data) original creation time, i.e., the round M was originally created in. In the case that a message M contains multiple payloads $\{M_1, ..., M_k\}$, we assume that it also provides the set of the corresponding timestamps $\{M.t_1, ..., M.t_k\}$. If not stated otherwise, we use the terms message and (the corresponding) payload synonymously.

Our broadcasting algorithm is executed individually for each message at each node in the network at any given round. Let v be a node that receives message M for the

Procedures	
open	v chooses ρ different neighbors $\{w_1, ..., w_\rho\} \in \Gamma(v)$ uniformly at random and establishes *outgoing* communication channels to them. Additionally, v establishes *incoming* communication channels to all nodes $\{w_1, ..., w_j\} \in \Gamma(v)$ that call on v.
push(M)	v sends message M over all outgoing channels.
pull(M)	v sends message M over all incoming channels.
receive	v receives and stores all incoming messages over all open channels.
close	v closes all open channels.

Table 5.1: *General procedures that any node $v \in V$ may perform in a single round.*

first time in round i. Then v calculates $i - M.t$ to determine M's age and executes the corresponding step of Algorithm 5.1. The same holds true for all messages already stored by v. Recall that each node always combines all payloads that are scheduled to be transmitted via push (pull) and forwards the result over all open outgoing (incoming) channels as one message only. In the following we state the algorithm w.l.o.g. for one fixed message M. Further, we assume that the message is created in round 0. Note that β is a suitable large constant.

Algorithm 5.1 (i : age of M, v : node executing the algorithm)

```
1 open
2 if i ≤ ⌈β log n⌉ then {Phase 1}
3    if the message was created by v in round i or received by v
     for the first time in the previous round, then push(M) over
     all (ρ) outgoing channels
4 if ⌈β log n⌉ + 1 ≤ i ≤ ⌈β(log n + log log n)⌉ then {Phase 2}
5    if v is informed (knows M), then push(M) over all (ρ)
     outgoing channels
6 if ⌈β(log n + log log n) + 1⌉ ≤ i ≤ ⌈β log n + 2β log log n⌉ then {Phase 3}
7    if v is uninformed, then pull(M) over all incoming channels
8 receive
9 close
```

Figure 5.1: *Each node $v \in V$ executes the above algorithm for each message M individually. The age i of a message M is obtained by subtracting the round $M.t$ in which M was originally created in from the current round j, i.e., $i = j - M.t$.*

Algorithm 5.1 consists of three different phases of specific durations. In the first phase each previously informed node v transmits message M via a push transmission exactly ρ times in a single round for one round only. Note that a node v executing Phase 1 may become informed by another node w via a push transmission or by creating the message M itself. The goal in Phase 1 is to rapidly disseminate M while keeping the costs low. Clearly, this can be achieved whenever the majority of the network is still unaware of M. We show that $\Theta(n)$ many nodes will be informed in no more than $\mathcal{O}(\log n)$

many rounds causing at most $\mathcal{O}(n)$ many messages in total (Lemma 5.4). During the second phase then, all previously informed nodes perform push transmissions over all ρ outgoing channels. That is, every informed node is actively disseminating M. In Phase 2 the amount of uninformed nodes shrinks to $n/\log^{\Theta(1)} n$ (Lemma 5.5). Recall that the sender actively calls and transmits M to the receiver in the case of a push transmission. Consequently, at the end of Phase 2, an uninformed node has a higher chance to obtain M by a pull transmission than by a push transmission. Phase 3 copes with this situation.

5.1.1.1 The Network Structure

The following three lemmas state some basic properties of the network.

Let G_j be the set (also called group) containing all nodes with degree $2^j\delta$ up to $2^{j+1}\delta - 1$ and let deg_{G_j} be the highest degree of group G_j. Further, let $|G_j|$ denote the size of G_j. We begin by examining the properties of such a grouping. Due to the power law degree distribution, the amount of nodes within each group is naturally upper bounded.

Lemma 5.1. *The amount of nodes within group j, for all $j \geq 1$, is upper bounded by*

$$|G_j| = \mathcal{O}\left(\frac{n}{((2^j-1)\delta-1)^{\alpha-1}} + \log n\right),$$

with high probability.

Proof. The expected group size for G_j is given by

$$cn \cdot \sum_{i=2^j\delta}^{2^{j+1}\delta-1} \frac{1}{(i-\delta)^\alpha} \leq cn \cdot \sum_{i=(2^j-1)\delta}^{(2^{j+1}-1)\delta-1} \frac{1}{i^\alpha} = \mathcal{O}\left(\frac{n}{((2^j-1)\delta-1)^{\alpha-1}}\right),$$

where c is a normalizing constant and $n/\left((2^j-1)\delta-1\right)^{\alpha-1} \geq \log n$. Standard Chernoff bounds (cf. Lemma 4.2) imply the lemma. □

Lemma 5.1 implies that the amount of low degree nodes dominates the amount of high degree nodes.

Lemma 5.2. *The amount of nodes with degree smaller than 2δ is $n(1-o(1))$, with high probability.*

Proof. The expected amount of nodes with degree at least 2δ is bounded by

$$cn \cdot \sum_{i=2^1\delta}^{\infty} \frac{1}{(i-\delta)^\alpha} = cn \cdot \sum_{i=\delta}^{\infty} \frac{1}{i^\alpha} = \Theta\left(\frac{n}{\delta^{\alpha-1}}\right),$$

where c is a normalizing constant. Since $\delta \in \Theta((\log\log n)^2)$ and $\frac{n}{\delta^{\alpha-1}} \in \omega(\log n)$, the claim follows by applying standard Chernoff bounds (cf. Lemma 4.2). □

The previous lemma implies that the vast majority of all nodes is contained in the first group, i.e., in the group with degree between δ and $2\delta - 1$. Consequently, most nodes have degree $\Theta(\delta)$. Now we state a lemma which gives us an estimation on the average node degree of all nodes with degree at least 2δ.

Lemma 5.3. *The average node degree among the nodes with degree at least 2δ is upper bounded by $\mathcal{O}(\delta)$, with high probability.*

Proof. Let G^{log} be the set of groups of at least logarithmic size and degree at least 2δ and let G_{log} be the set of groups of less than logarithmic size and degree at least 2δ. More formally:

$$G^{log} := \left\{G_i \mid |G_i| \geq \log n, 1 \leq i \leq \log\left({}^{(n^{1/\alpha}+\delta)}\!/_{\delta}\right)\right\}$$
$$G_{log} := \left\{G_i \mid |G_i| < \log n, 1 \leq i \leq \log\left({}^{(n^{1/\alpha}+\delta)}\!/_{\delta}\right)\right\}$$

Note that G^{log} and G_{log} only include nodes of groups up to a degree of $n^{1/\alpha} + \delta$. Let further $\hat{G}$ be the group of nodes which would have been assigned a degree larger than $n^{1/\alpha} + \delta$. Note that these nodes have degree exactly $n^{1/\alpha} + \delta$ due to our model. Due to the power law degree distribution it holds that

$$|\hat{G}| < \sum_{i=n^{1/\alpha}+\delta}^{\infty} \frac{cn}{(i-\delta)^{\alpha}} = \sum_{i=n^{1/\alpha}}^{\infty} \frac{cn}{i^{\alpha}} = \mathcal{O}(n^{1/\alpha})$$

with high probability, where c is a normalizing constant. Further, according to Lemma 5.1, it holds that for any $j \geq 1 : |G_j| = \mathcal{O}(n/((2^j - 1)\delta - 1)^{\alpha-1} + \log n)$ with high probability.

Since our grouping implies that we can have no more than logarithmic many groups G_i, it follows that G_{log} may contain at most $\mathcal{O}(\log n)$ many groups of size at most $\mathcal{O}(\log n)$. Further, recall that $\alpha > 3$. Additionally, according to Lemma 5.2, we know that the amount of nodes X with degree at least 2δ is bounded by $\frac{c_1 \cdot n}{\delta^{\alpha-1}} \geq X \geq \frac{c_2 \cdot n}{\delta^{\alpha-1}}$, for some positive constants c_1, c_2. It follows that the average node degree for all nodes with degree at least 2δ is upper bounded by:

$$\begin{aligned}
&\frac{\delta^{\alpha-1}}{nc_2}\left(\sum_{G_k \in G_{\log}} deg_{G_k}|G_k| + \sum_{G_k \in G^{\log}} deg_{G_k}|G_k| + \mathcal{O}(|\hat{G}|n^{1/\alpha})\right) \\
&\quad \leq \frac{\delta^{\alpha-1}}{nc_2}\left(\sum_{G_k \in G^{\log}} 2^{k+1}\delta\frac{3^{\alpha-1} \cdot cn}{(2^k\delta)^{\alpha-1}} + \sum_{G_k \in G_{\log}} deg_{G_k}\mathcal{O}(\log n) + \mathcal{O}(n^{2/\alpha})\right) \\
&\quad \leq \frac{\delta^{\alpha-1}}{nc_2}\left(3^{\alpha}\sum_{G_k \in G^{\log}} 2^k\delta\frac{cn}{(2^k\delta)^{\alpha-1}} + \mathcal{O}(\log^2 n \sqrt[\alpha]{n}) + \mathcal{O}(n^{2/\alpha})\right) \\
&\quad \leq \frac{\delta^{\alpha-1}}{nc_2}\left(\mathcal{O}\left(\log^2 n \sqrt[\alpha]{n} + n^{2/\alpha}\right) + \frac{3^{\alpha} \cdot cn}{\delta^{\alpha-2}}\sum_{1 \leq k \leq \log n}\left(\frac{1}{2^{\alpha-2}}\right)^k\right).
\end{aligned}$$

Further, it holds that $\sum_{1\leq k\leq \log n}\left(\frac{1}{2^{\alpha-2}}\right)^k \leq \sum_{1\leq k\leq \log n}\left(\frac{1}{2}\right)^k \leq 2\,(1-o(1)) \leq 2$. Hence, we obtain that

$$\begin{aligned}
&\frac{\delta^{\alpha-1}}{nc_2}\left(\sum_{G_k\in G_{\log}} deg_{G_k}|G_k| + \sum_{G_k\in G^{\log}} deg_{G_k}|G_k| + \mathcal{O}(|\hat{G}|n^{1/\alpha})\right)\\
&\leq \frac{\delta^{\alpha-1}}{nc_2}\left(\frac{3^{\alpha+1}\cdot cn}{\delta^{\alpha-2}} + \mathcal{O}(\log^2 n\,\sqrt[\alpha]{n} + n^{2/\alpha})\right)\\
&\leq \frac{\delta^{\alpha-1}}{nc_2}\frac{3^{\alpha+1}\cdot cn}{\delta^{\alpha-2}}\left(1+\mathcal{O}\left(\frac{\log^2 n\cdot\delta^{\alpha-2}}{n^{1-1/\alpha}}\right)+\mathcal{O}\left(\frac{\delta^{\alpha-2}}{n^{1-2/\alpha}}\right)\right)\\
&= \mathcal{O}(\delta)
\end{aligned}$$

with high probability. □

So far we stated some basic insights concerning the given network structure. However, in our analysis we consider the so-called Configuration Model to construct a random graph with a certain degree distribution [66]. Thus, our results basically depend on random pairing decisions according to that model. These decisions may lead to a multigraph. Our goal, however, is to provide an efficient broadcasting algorithm for realistic networks, which are usually simple. The following theorem states the probability for a graph generated by the Configuration Model to remain simple [33, Theorem 3.1.2].

Theorem 5.1. *Let G be a random graph constructed according to the Configuration Model and let the degrees be distributed according to a power law distribution with exponent $\alpha > 3$. Furthermore, let n denote the size of the graph and let δ be the smallest degree. Then G is a simple graph with probability $\frac{1}{e^{\mathcal{O}(\delta^2)}}$.*

Note that our results hold with high probability, i.e., with probability at least $1-n^{-c_2}$, for some positive constant c_2. On the other hand, a simple graph is constructed by the Configuration Model with probability at least $e^{-c_1\delta^2} = n^{-\Theta((\log\log n)^2/\log n)} \gg n^{-c_2}$, for some positive constant c_1. Therefore, we conclude that our algorithm does not fail for a randomly created simple graph according to the Configuration Model with probability at least

$$e^{-c_1\delta^2}\cdot(1-o(1))\,.$$

5.1.2 Analysis

In this section we analyze the behavior and properties of the nodes executing Algorithm 5.1 for a specific message M. As we mentioned already, we use the Principle of Deferred Decision and the Configuration Model in the analysis (cf. Section 4). In the following we first give a more refined description of our model before presenting the corresponding

analysis. Note that, for the reader's convenience, the most important definitions are listed in Section 5.1.3 in a more condensed form.

Let the set (also called group) G_j contain all nodes with degree $2^j\delta$ up to $2^{j+1}\delta - 1$, where $j \geq 0$, and let the highest degree of said group be denoted by $\deg_{G_j}$. Further, let $|G_j|$ denote the size of G_j. In the following we assume that $n^{1/\alpha} + \delta$ can be written as $2^k\delta$ for some k. Further, we assume that the last group, i.e., the group $G_{k'}$ with $k' = \log(^{(n^{1/\alpha}+\delta)}/_{\delta})$, contains nodes of degree $n^{1/\alpha} + \delta$ only. Therefore, we obtain less than $\log n$ different groups in total.

Creating the Network Recall that each node executing Algorithm 5.1 uses ρ neighbors in each round to call on. Since we use the Principle of Deferred Decision and the Configuration Model in the analysis, we need to specify a procedure which ensures a proper pairing. Let $s(v) := \{s_1^v, ..., s_{\deg_v}^v\}$ be the set of all *stubs* of a node $v \in V$. Note that the meaning of the term *stub* is exactly the same as in the standard Configuration Model. Further, let $s(W) := \bigcup_{v \in W} s(v)$, where $W \subseteq V$. Let the set $s_f(v)$ contain all free stubs of v, i.e., all stubs that are not already connected (also called paired) to existing edges at a certain time, while the set $\overline{s_f(v)}$ contains all already connected ones. Then $S_f(G) := \bigcup_{\forall v \in V} s_f(v)$.

W.l.o.g. let $v \in V$ be an arbitrary node executing Algorithm 5.1 with respect to an arbitrary message M. In the first step of the broadcasting algorithm v chooses ρ different stubs $S := \{s_1^v, ..., s_\rho^v\} \in s(v)$. Then all free stubs among these are paired with their randomly chosen counterparts of $S_f(G) \setminus S$ according to the Configuration Model. To obtain a proper pairing, we divide this current pairing round into ρ sub-rounds. That is, each stub $s_j^v \in S$ is being considered separately in an individual sub-round. In each sub-round $j \leq \rho$ we choose a yet unpaired stub $s_j^v \in S$ and pair it with some stub chosen uniformly at random from the set $S_f(G) \setminus S$. Afterwards, v executes the corresponding step, determined by the age of M, of Algorithm 5.1 for all edges incident to the previously chosen stubs in S.

To generalize the idea let $\tilde{V}(i) := \{v_1, \ldots, v_k\}$ be the set of nodes allowed to open communication channels in round i. Then round i is subdivided into $\rho \cdot k$ many sub-rounds. Similar as before, a node $v_{\lceil j/\rho \rceil}$ chooses in each sub-round $1 \leq j \leq \rho \cdot k$ a stub s' uniformly at random among its own stubs that were not considered in this round so far. If s' is not paired yet, then $v_{\lceil j/\rho \rceil}$ pairs s' with an unpaired stub of $S_f(G)$, and subsequently executes the corresponding step, determined by the age of M, of Algorithm 5.1 for this newly created edge. Otherwise, since s' is already paired, $v_{\lceil j/\rho \rceil}$ simply executes the corresponding step, determined by the age of M, of Algorithm 5.1 for the edge that s' is incident to. This implies that in every round each node may generate up to ρ new edges.

Formal Analysis Finally, we start the analysis at the point where one node generates message M. The first phase covers the situation where the majority of the nodes are uninformed. We take advantage of this fact in the proof of the next lemma. In the following we show that at least $\Omega(n)$ many nodes will be informed in no more than $\mathcal{O}(\log n)$ many rounds causing at most $\mathcal{O}(n)$ many transmissions of M in total. The resulting situation then is the starting point for Phase 2 of the algorithm.

Lemma 5.4. *(Phase 1.) Let M be a newly generated message and let G be a random network as described in Section 5.1.1. Further, let $\delta = \gamma(\log\log n)^2$. Then, by using $\mathcal{O}(n)$ many message transmissions, Algorithm 5.1 informs at least $\frac{n}{2}$ many nodes about M in at most $\mathcal{O}(\log n)$ many rounds with high probability.*

Proof. The lemma is proven by induction in the following manner. We start with a source node u generating message M, whose propagation we are going to analyze. Therefore, node u is considered for the base case. We show that u informs ρ distinct nodes in $\mathcal{U}$ in the very first round with high probability. The induction step on the other hand is more complex. We show that two major properties hold for each round $i \geq 0$:

1. each node $v \in \mathcal{I}^+(i)$ provides at least $|s_f(v)| \geq \deg_v /2$ many free stubs, and
2. each node $v \in \mathcal{I}^+(i)$ chooses at least ρ/c' many distinct nodes in $\mathcal{U}(i)$ on average, where $c' \leq 3 \cdot (\alpha - 1) \cdot 2^\alpha$ is the constant defined by Equation 5.7.

Hereby *choose* refers to the pairing process described earlier. Note that only push transmissions are used during Phase 1 according to Algorithm 5.1. The combination of the above arguments implies that the lemma follows by setting $\rho \geq \max\{100, (c')^2\}$.

Let $\overline{X_j^i(v)}$ describe the event that v does not inform a previously uninformed node in sub-round j of round i. Let $v \rightsquigarrow \mathcal{I}(i)$ indicate the amount of already informed nodes chosen by v in round i and let $v \rightsquigarrow \mathcal{U}(i)$ indicate the amount of uniformed nodes chosen by v in round i.

Case $i = 0$: W.l.o.g. let u be the node generating message M. In the first round u chooses exactly ρ stubs of the set $s_f(u)$ uniformly at random. If our network does not contain any self-loops or double edges, it follows directly that u will inform ρ different neighbors. On the other hand, if we consider a network with double edges and self loops, then we obtain the following statement. For any sub-round j in this round it holds that

$$\Pr\left(\overline{X_j^0(u)}\right) < \frac{\rho \cdot \sqrt{n}}{n} = \frac{\rho}{\sqrt{n}}. \tag{5.1}$$

Note that the nodes chosen in sub-rounds $1 \dots j-1$ may have a degree of at most $n^{1/\alpha} + \delta$ each. Hence, at most $n^{1/\alpha} + \delta < \sqrt{n}$ many stubs may be available at these

nodes. Thus, after the first round (i.e., after the first ρ sub-rounds) the expected amount of newly informed nodes is bounded by

$$\mathbb{E}\left[u \rightsquigarrow \mathcal{U}(0)\right] \overset{Eq.\ 5.1}{\geq} \rho\left(1 - \frac{\rho}{\sqrt{n}}\right) = \rho(1 - o(1)). \tag{5.2}$$

Equation 5.1 implies that u contacts ρ distinct uninformed nodes over all ρ sub-rounds with probability at least

$$\left(1 - \frac{\rho}{\sqrt{n}}\right)^{\rho} \geq 1 - \mathcal{O}\left(\frac{1}{\sqrt{n}}\right). \tag{5.3}$$

Thus, by combining Equation 5.2 and 5.3, we conclude that ρ nodes are informed in the very first round with high probability.

Case $i > 0$**:** In the following we perform a case analysis for the induction step. To be more specific, we distinguish between the case $|\mathcal{I}^+(i)| < \log n$ and $|\mathcal{I}^+(i)| \geq \log n$ for each considered $i > 0$.

Subcase $|\mathcal{I}^+(i)| < \log n$**:** In this subcase we focus on already informed individual nodes rather than properties of already informed sets. First, we show that each informed node chooses only uninformed nodes with high probability. Using this property we then obtain that $|\mathcal{I}^+(i+1)| \geq |\mathcal{I}^+(i)| \cdot \rho/2$.

We begin by bounding the probability for a node $v' \in \mathcal{U}(i)$ to be chosen multiple times by at least one node $v \in \mathcal{I}^+(i)$. Due to $|\mathcal{I}^+(i)| < \log n$, the amount of total pairings initiated by the nodes in $\mathcal{I}^+(i)$ is limited by $\rho \log n$. Further, it holds that $|s_f(v')| < 2n^{1/\alpha}$, since $\deg_{v'} \leq \sqrt[\alpha]{n} + \delta$. Let $v_1^i, ..., v_k^i \in \mathcal{U}(i)$ be a sequence of the uninformed nodes in round i sorted in descending order by the degree. Note that $k < n$. Let the indicator variable x_j^i be equal to 1 if v_j^i is chosen at least twice by the nodes in $\mathcal{I}^+(i)$ and 0 otherwise. Further, let $X = \sum_{1 \leq j \leq k} x_j^i$. Thus, the probability for v_1^i to be chosen multiple times is upper bounded by

$$\Pr\left(x_1^i\right) \leq \binom{\rho \log n}{2} \cdot \left(\frac{n^{1/\alpha}}{\Theta(n\delta)}\right)^2 = \left(\mathcal{O}\left(\frac{\log n}{n^{1-1/\alpha}\delta}\right)\right)^2 = n^{-(1+\Omega(1))}. \tag{5.4}$$

Note that these events are negatively correlated, since each connection established to some v_j^i diminishes the amount of connections available for all yet unconsidered nodes of the sequence. Therefore, $\Pr(x_1^i) \geq \Pr(x_2^i) \geq ... \geq \Pr(x_k^i)$. Consequently,

$$\mathbb{E}[X] \leq n \cdot \left(\mathcal{O}\left(\frac{\log n}{n^{1-1/\alpha}\delta}\right)\right)^2 = n^{-\Omega(1)},$$

since $3 < \alpha$. Utilizing standard Chernoff bounds (cf. Lemma 4.1), we conclude that none of the nodes $v_1^i, ..., v_k^i$ is connected to more than one node in $\mathcal{I}^+(i)$ with probability $1 - n^{-\Omega(1)}$.

In the following we assume an exponential growth of $\mathcal{I}$ up to the current round due to the induction. In order to compute $\mathcal{I}^+(i+1)$, i.e., the amount of newly informed nodes by nodes in $\mathcal{I}^+(i)$, we need the following insight. By considering the unpaired stubs of an arbitrary node $v \in \mathcal{I}^+(i)$ only, the probability for v to choose another node $v' \in \mathcal{I}(i)$ is upper bounded by

$$\mathcal{O}\left(\frac{\sqrt{n}\log n}{n\delta}\right) = \mathcal{O}\left(\frac{\log n}{\sqrt{n}\delta}\right). \tag{5.5}$$

Further, Equation 5.4 states that none of the nodes in $\mathcal{I}^+(i)$ was chosen more than once with high probability. Recall that there can be at most $\log n$ many (newly) informed nodes in round i. Additionally, the probability that no node $u \in \mathcal{U}$ is chosen multiple times is at least $(1 - 1/n^{\Omega(1)})$. Thus, the amount of uninformed nodes chosen by each $v \in \mathcal{I}^+(i)$ is lower bounded by

$$\rho - 1$$

with probability at least

$$\left(1 - \mathcal{O}\left(\frac{\log n}{\sqrt{n}\delta}\right)\right)^{\rho} \cdot \left(1 - \frac{1}{n^{\Omega(1)}}\right) \geq 1 - n^{-\Omega(1)}.$$

We obtain the same result for all nodes in $\mathcal{I}^+(i)$ combined by utilizing the Union Bound (cf. Lemma 4.5).

Subcase $|\mathcal{I}^+(i)| \geq \log n$: In this subcase we consider the behavior of the entire group $\mathcal{I}^+(i)$ instead of individual nodes. Similar to the previous subcase, we first estimate the amount of nodes $v \in \mathcal{I}^+(i)$ that were chosen multiple times in the previous round. The probability of choosing a node $v \in \mathcal{I}^+(i)$ at least $\deg_v/2$ many times is upper bounded by

$$\binom{\rho\frac{n}{2}}{\deg_v/2} \cdot \left(\Theta\left(\frac{\deg_v}{n\delta}\right)\right)^{\frac{\deg_v}{2}} = \left(\mathcal{O}\left(\frac{n}{n\delta}\right)\right)^{\frac{\delta}{2}} \leq \left(\frac{c}{\delta}\right)^{\frac{\delta}{2}} = o\left(\log^{-\sqrt{\delta}} n\right), \tag{5.6}$$

for some constant c. Let X_j be a random variable with $X_j = 1$ if the jth node v_j in $\mathcal{I}^+(i)$ was chosen less than $deg_{v_j}/2$ many times and 0 otherwise. Furthermore, let $X = \sum_{j=1}^{|\mathcal{I}^+(i)|} X_j$. Then it holds that

$$\mathbb{E}[X] \overset{Eq.\ 5.6}{\geq} |\mathcal{I}^+(i)| \cdot \left(1 - \log^{-\sqrt{\delta}} n\right) = |\mathcal{I}^+(i)| \cdot (1 - o(1)).$$

Since the random variables X_j are negatively correlated and $|\mathcal{I}^+(i)| \geq \log n$, we obtain the same result with high probability using standard Chernoff bounds (cf. Lemma 4.2) up to a constant factor of 2/3. Consequently, each remaining node $v \in \mathcal{I}^+(i)$ with $|s_f(v)| \geq {}^{\deg_v}/_2$ chooses an already informed node with probability at most 1/2. On the other hand, v chooses an uninformed node with probability at least $1/(2c)$, provided v decided not to pick an already informed one. Here c is a normalizing constant. The

latter claim follows directly from the power law distribution and the argument that an arbitrary node of $\mathcal{U}(i)$ is chosen with probability at least

$$\frac{n\delta/2}{cn\delta} = \frac{1}{2c},$$

whenever $|\mathcal{U}(i)| \geq n/2$.

Thus, we obtain that

$$\mathbb{E}[|\mathcal{I}^+(i+1)|] \geq |\mathcal{I}^+(i)| \cdot \frac{2}{3} \cdot \left(\rho \cdot \left(1 - \frac{1}{2}\right) \cdot \frac{1}{2c}\right) := \mu. \tag{5.7}$$

Note that we can model the ρ choices of all the nodes within $\mathcal{I}^+(i)$ by a Vertex Exposure Martingale $X_0, X_1, ...$ (cf. Definition 4.3) of length $t = |\mathcal{I}^+(i)|$. Further, note that this Martingale is ρ-Lipschitz, since each node may inform at most ρ many nodes of $\mathcal{U}(i)$. Utilizing the Azuma Inequality (cf. Lemma 4.4), we obtain that $|\mathcal{I}^+(i+1)| \geq c' \cdot |\mathcal{I}^+(i)|$, for some constant c', with probability at least $1 - 1/n^{\Omega(1)}$, since $\mu = \Omega(|\mathcal{I}^+(i)|)$, $|\mathcal{I}^+(i)| \geq \log n$, and

$$\Pr[|X_t - X_0| \geq (1-c')\mu] \leq 2\exp\left(-\frac{((1-c')\mu)^2}{2\rho^2|\mathcal{I}^+(i)|}\right) = \frac{1}{n^{\Omega(1)}},$$

for some arbitrarily small constant c'. Therefore, the lemma follows whenever ρ is large enough. □

At this point we have a substantial amount of informed nodes in the network. We exploit this property in the proof of the next lemma. It basically states that at least a constant fraction of the edges of at least a constant fraction of all uninformed nodes point to already informed ones. More formally,

$$|\{u \in \mathcal{U} \mid (|\{v \in \mathcal{I} \mid v \in \Gamma(u)\}|) = \Omega(\deg_u)\}| = \Omega(|\mathcal{U}|).$$

Lemma 5.5. *(Phase 2.) Let G be a random network as described in Section 5.1.1 and let $\delta = \gamma(\log\log n)^2$. Then no more than $\frac{n}{\log^{9+\Omega(1)} n}$ uninformed nodes are left after $\mathcal{O}(\log\log n)$ additional rounds and $\mathcal{O}(n\log\log n)$ message transmissions, with high probability.*

Proof. In the following we consider a random graph constructed by the Configuration Model (with possible loops and double edges) to show our claim. To be more specific, we show that for each considered round i at least $\Omega(\delta)$ many stubs of each node of at least a constant fraction of $\mathcal{U}(i)$ are (randomly) connected to nodes in $\mathcal{I}(i)$, with high probability. Note that the pairing choices in this random graph depend on the pairing choices of the previous rounds. However, since we only used push-transmissions so far, the stubs of the uninformed nodes together with the free stubs of the informed nodes still form a random graph.

Due to Algorithm 5.1, we know that every informed node has actively used at most $\mathcal{O}(\log\log n)$ of its stubs during the previous rounds $1,\dots,i-1$. Thus,

$$|\mathcal{I}(i)|^{-1}\sum_{v\in\mathcal{I}(i)}|s_f(v)| = \delta - \Theta(\log\log n). \tag{5.8}$$

Our analysis exploits this property to construct a residual network.

Let $\mathcal{I}'(i) := \{v \in \mathcal{I}(i) \mid \deg_v \leq 2\delta\}$ denote the set of nodes of $\mathcal{I}(i)$ with degree at most 2δ. Further, let $S_u \subseteq s_f(u)$, with $|S_u| = \delta$, contain exactly δ randomly chosen stubs of each corresponding node $u \in \mathcal{U}(i)$ and let $\overline{S}_u = s_f(u) \setminus S_u$ be its complement. Note that all stubs of each node in $\mathcal{U}$ are still free. Therefore, the stubs of $\mathcal{I}'(i)$ together with the ones of $\bigcup_{u\in\mathcal{U}(i)} S_u$ form a residual network, i.e., a random graph that we use in the upcoming analysis (cf. Figure 5.2). Obviously, this network contains $\Theta(n\delta)$ many stubs, since the amount of nodes with degree at most 2δ is $\Theta(n)$ according to Lemma 5.2. Thus, an unused stub of $\mathcal{I}'(i)$ is paired with a stub in the residual network with some constant probability p. Recall that at least $\delta(1-o(1))$ many stubs are free for each node in $\mathcal{I}(i)$ on average (cf. Equation 5.8). Further, note that $|\mathcal{I}(i)| \geq n/2 \geq |\mathcal{U}(i)|$. It follows that the expected amount of edges between $\mathcal{I}'(i)$ and $\bigcup_{u\in\mathcal{U}(i)} S_u$ within the residual network is at least

$$\begin{aligned} &p\cdot\frac{|\mathcal{I}(i)|}{2}\cdot\overbrace{(\delta-\Theta(\log\log n))}^{\text{avg. amount of unused stubs per node}}\cdot\overbrace{\frac{|\mathcal{U}(i)|\delta}{2\delta|\mathcal{I}(i)|+\delta|\mathcal{U}(i)|}}^{\text{prob. to choose a node in } \mathcal{U}(i) \text{ within the residual network}} \\ &\geq p\frac{\delta(1-o(1))}{2}\frac{|\mathcal{U}(i)|}{3} = \frac{|\mathcal{U}(i)|\delta(1-o(1))}{c}, \end{aligned} \tag{5.9}$$

where c is a suitable constant. Further, it holds that $|\mathcal{I}'(i)| \geq |\mathcal{I}(i)|/2$, since $n(1-o(1))$ many nodes have degree at most 2δ and $|\mathcal{I}(i)| \geq n/2$.

Inequality 5.9 implies that the expected amount of edges between $\mathcal{I}'(i)$ and the rest of the residual network is at least $\frac{|\mathcal{U}(i)|\delta(1-o(1))}{c}$. Note that the above procedure can be defined as a Vertex Exposure Martingale $X_0, X_1, \dots$ (cf. Definition 4.3) with length $t = |\mathcal{I}'(i)|$. Due to the construction, the Martingale is 2δ-Lipschitz. Further, $|\mathcal{U}(i)| \geq n/\log^{\mathcal{O}(1)} n$ and $|\mathcal{I}'(i)| \leq n$. Thus, by applying the Azuma Hoeffding Inequality (cf. Lemma 4.4), it follows that

$$\Pr[|X_t - X_0| \geq |\mathcal{U}(i)|\delta/c] \leq \exp\left(-\frac{\left(|\mathcal{U}(i)|\delta/c\right)^2}{8\delta^2|\mathcal{I}'(i)|}\right) \leq \exp\left(-\sqrt{n}\right).$$

In the following we prove by contradiction that at least $\frac{|\mathcal{U}(i)|}{4c}$ many nodes of $\mathcal{U}(i)$ must have at least $\frac{\delta}{4c}$ many connections to $\mathcal{I}'(i)$ in a random pairing. Therefore, we assume that less than $|\mathcal{U}(i)|/4c$ nodes of $\mathcal{U}(i)$ are incident to more than $\delta/4c$ edges leading to nodes of $\mathcal{I}'(i)$. Consequently, more than $(|\mathcal{U}(i)| - |\mathcal{U}(i)|/4c)$ many nodes of $\mathcal{U}(i)$ must be

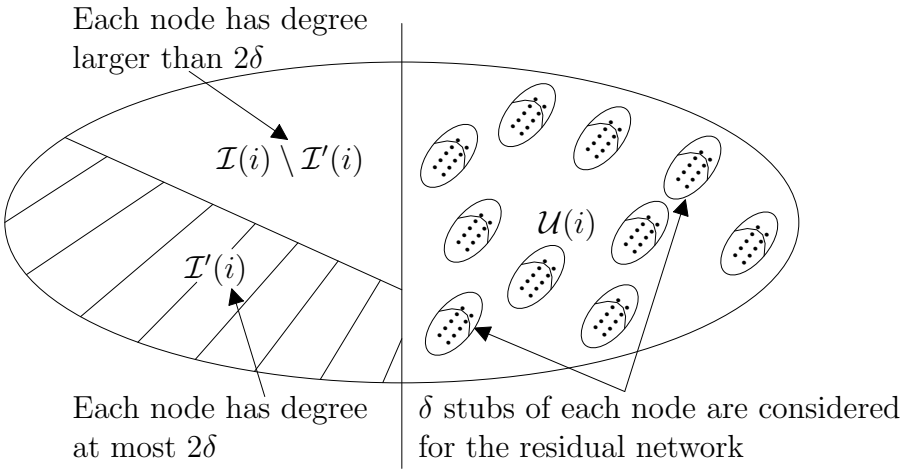


Figure 5.2: *The above figure depicts the situation of the network G during Phase 2. The residual network is a subgraph of G formed by the random pairings of $s(\mathcal{I}'(i)) \cup \bigcup_{u \in \mathcal{U}(i)} S_u$, where $\mathcal{I}'(i) := \{v \in \mathcal{I}(i) \mid \deg_v \leq 2\delta\}$ and $S_u \subseteq s_f(u)$ with $|S_u| = \delta$. It is used to estimate the amount of edges between $\mathcal{I}'$ and $\mathcal{U}$ during Phase 2 of the broadcasting algorithm (cf. Lemma 5.5).*

incident to less than $\delta/4c$ of such edges. Recall that $|S_u| = \delta$ for every $u \in \mathcal{U}(i)$ in the residual network. Therefore, the total amount of edges in the residual network between $\mathcal{I}'(i)$ and $\mathcal{U}(i)$ is limited by

$$\frac{|\mathcal{U}(i)|\delta}{4c} + \frac{3|\mathcal{U}(i)|}{4c}\frac{\delta}{4c} = \frac{|\mathcal{U}(i)|\delta}{4c} + \frac{3|\mathcal{U}(i)|\delta}{16c^2} < \frac{|\mathcal{U}(i)|\delta}{2c},$$

which is a contradiction to Inequality 5.9. Thus, at least $\frac{|\mathcal{U}(i)|}{4c}$ many nodes of $\mathcal{U}(i)$ must have at least $\frac{\delta}{4c}$ many connections to $\mathcal{I}'(i)$ in a random pairing. In the following we examine these nodes and their edges leading to nodes in $\mathcal{I}'(i)$.

W.l.o.g. let $u \in \mathcal{U}(i)$ have $m' \geq \frac{\delta}{4c}$ many connections to nodes of $\mathcal{I}'(i)$ and let the set $L \subset \mathcal{I}'(i)$ contain these nodes. In the worst case each connection leads to a different informed node, i.e., it holds that $|L| = m'$. It follows that the probability for u not to be chosen by any informed neighbor in the residual network is at most

$$q := \left(1 - \frac{1}{2\delta}\right)^{\frac{\delta}{4c}} = \Theta(1),$$

since we use only at most 2δ many stubs of each node $v \in \mathcal{I}'(i)$ to construct the residual network.

Thus, the expected amount of nodes in $\mathcal{U}(i)$ that are informed by some node of $\mathcal{I}'(i)$ is at least $(1-q)\frac{|\mathcal{U}(i)|}{4c}$. Utilizing the Azuma Hoeffding inequality as we did above, we obtain that at least a constant fraction of the uninformed nodes becomes informed in round i, with high probability. □

During Phase 1 and Phase 2 only push transmissions were utilized to disseminate M. This implies that none of the stubs of a yet uninformed node was paired so far. At this point we are almost ready to prove Theorem 5.2. Before we begin, however, we first need to prove the following auxiliary lemmas.

Lemma 5.6. *Let G be a random network as described in Section 5.1.1 and let s be the first round right after Phase 2. Further, let $\delta = \gamma(\log\log n)^2$ and let the number of uninformed nodes before round s be $\frac{n}{\log^c n}$, where $c = 9 + \Theta(1)$. Then it holds that*

$$\forall u \in \mathcal{U}(s) : \deg_u \leq c' \log n \cdot \log\log n,$$

for some positive constant c' with probability $1 - 1/n$.

Proof. First note that round s defines the transition point from Phase 2 to Phase 3. Further, note that at least $\Omega(n)$ many nodes execute Phase 2 over $\Theta(\log\log n)$ many rounds due to Algorithm 5.1 and Lemma 5.4. Recall that each informed node transmits only once over ρ connections in Phase 1 and $\mathcal{O}(\log\log n)$ many times in Phase 2. Therefore, each informed node has paired at most $\mathcal{O}(\log\log n)$ many of its stubs due to push transmissions initiated by itself. This implies that at most $\mathcal{O}(n \log\log n)$ many message transmissions occurred in total, which in turn means that the network provides at least $\Theta(n\delta)$ many free stubs.

In the following we consider the first group G_0 only, i.e., nodes with degree up to $2\delta - 1$. Note that there are at least $n(1 - o(1))$ many of these in G (cf. Lemma 5.2). Further, at least a constant fraction of them become informed in Phase 1, since $|\mathcal{I}| \geq n/2$ after Phase 1. It follows that at most $\rho \cdot n$ many stubs are chosen for communication in each of the rounds in Phase 2 and at most $\mathcal{O}(n)$ many in total during Phase 1. Consequently, the amount of (incoming) connections established by other nodes to a newly informed node $v \in G_0$ during Phase 1 or Phase 2 is limited by $c''\sqrt{\log\log n}$ with probability at least

$$p := 1 - \binom{\mathcal{O}(n)}{c''\sqrt{\log\log n}} \cdot \left(\frac{\deg_v}{\Theta(n\delta)}\right)^{c''\sqrt{\log\log n}} \geq 1 - \left(\mathcal{O}\left(1/\delta^{1/4}\right)\right)^{c''\sqrt{\log\log n}}.$$

Let p' be the probability that an informed node $v \in G_0$ does not provide more than $\mathcal{O}((\log\log n)^{3/2})$ many inner connections after Phase 2 has ended. It follows that

$$p' \geq p^{\Theta(\log\log n)} \geq \left(1 - \frac{\delta^{\Theta\left(\frac{\sqrt{\log\log n}}{\log(\delta)}\right)}}{\delta^{(1/4)\cdot c''\sqrt{\log\log n}}}\right)^{\Theta(\log\log n)} \geq 1 - \frac{1}{\delta^{(1/8)\cdot c''\sqrt{\log\log n}}} = 1 - o(1).$$

Therefore, using Chernoff bounds, we obtain that the majority of the nodes in G_0 that transmit during Phase 2 have no more than $\mathcal{O}((\log\log n)^{3/2})$ many used stubs, with high probability. Further, each such node chooses at least $\rho/2$ free stubs during each push operation with probability at least

$$\left(1 - \frac{\mathcal{O}((\log\log n)^{3/2})}{\delta}\right)^{\rho/2} \geq 1 - \mathcal{O}\left(\frac{1}{\sqrt{\log\log n}}\right).$$

This implies that, on expectation, almost all of these nodes choose at least $\rho/2$ free stubs for each push operation. Since the above equation is a lower bound on the probability, we apply Chernoff bounds to bound the expectation. Thus, at least a constant fraction of the nodes in G_0 that transmit during Phase 2 and that have no more than $\mathcal{O}((\log\log n)^{3/2})$ many used stubs pair at least $\rho/2$ free stubs during each push operation. Consequently, at least $\Omega(n)$ many stubs are paired in each of these rounds due to the push operation. Further, since none of the stubs of an uninformed node u is paired, u is informed with probability proportional to its degree. Thus, whenever c' is a suitably large constant, we obtain that the probability for each uninformed node with degree at least $c' \log n \cdot \log\log n$ to remain uninformed up to round s is at most

$$\left(1 - \frac{\deg_u}{\Theta(n\delta)}\right)^{\Omega(n\cdot\log\log n)} \leq \exp\left(-\Omega(\deg_u) \cdot \frac{\log\log n}{\delta}\right) \leq \frac{1}{n^2}.$$

Since there are at most $n/2$ many uninformed nodes, applying the Union Bound gives the lemma. □

The following lemma is similar to the one presented in [12] and examines the structure within the uninformed nodes in round s, i.e., in the first round right after Phase 2.

Lemma 5.7. *Let G be a random network as described in Section 5.1.1 and let $\mathcal{U}_l(i)$ be the set of uninformed nodes in round i with at least l neighbors in $\mathcal{U}(i)$. Further, let s be the first round right after Phase 2. Additionally, let $\delta = \gamma(\log\log n)^2$ and let the number of uninformed nodes in round s be $\frac{n}{\log^c n}$, where $c = 9 + \Theta(1)$. Then it holds that*

$$\Omega\left(\frac{|\mathcal{U}(s)|\delta}{\log^c n}\right) \ni |\mathcal{U}_1(s)| \in \mathcal{O}\left(\frac{|\mathcal{U}(s)|}{\log^{q-1} n}\right)$$

and

$$|\mathcal{U}(s+1)| = \mathcal{O}\left(|\mathcal{U}(s)|\left(\frac{1}{\log^{c-2} n}\right)^{\rho/2}\right),$$

where $2q < c$, with high probability.

Proof. In the following we state some structural properties of the network in round s. That is, we mainly focus on *revealing* (some) connections of each node in $\mathcal{U}(s)$ to obtain an upper bound on the amount of nodes contained in $\mathcal{U}(s+1)$. To prove our claim, we utilize an additional grouping scheme of the nodes in $\mathcal{U}$. This scheme incorporates three groups.

Let q be a constant with $2q < c$. Then the first group is formed by nodes of degree between δ and $\log^q n$. The second group contains the nodes of degree at least $\log^q n$ up to $\log^{c+1} n - 1$. Let v be such a node, then $\frac{\log n}{\deg_v} \geq \frac{|\mathcal{U}(s)|}{n}$. The last group is the set of nodes with degree at least $\log^{c+1} n$. Let v' be such a node, then $\frac{\log n}{\deg_{v'}} \leq \frac{|\mathcal{U}(s)|}{n}$. The groups are called $\mathcal{U}^{(1)}(s)$, $\mathcal{U}^{(2)}(s)$, and $\mathcal{U}^{(3)}(s)$, respectively. Further, an edge is called

inner edge if both of its endpoints lie in $\mathcal{U}(s)$. Recall that, on average, none of the nodes in $\mathcal{I}$ used more than a $o(1)$-fraction of their stubs so far (cf. proof of Lemma 5.5). Further, according to Lemma 5.6, it holds that $\forall u \in \mathcal{U}(s) : \deg_u \leq c' \log n \cdot \log\log n$, for some large constant c'. Consequently, we obtain the following claims.

Claim 5.1. *Each node $v \in \mathcal{U}^{(1)}(s)$ has at most $\gamma' \log n$ many inner edges for some suitable large constant γ', with high probability.*

Proof. By definition the largest degree of a node in $\mathcal{U}^{(1)}(s)$ is $\log^q n$. Consequently, the probability for an arbitrary node $v \in \mathcal{U}^{(1)}(s)$ to have at least $\gamma' \log n$ many inner edges is upper bounded by

$$\begin{aligned}
&\binom{\deg_v}{\gamma' \log n} \left(\frac{|\mathcal{U}(s)| \cdot (c' \log n \cdot \log\log n)}{n\delta(1-o(1))} \right)^{\gamma' \log n} \\
&\qquad < \left(\frac{e \log^q n}{\gamma' \log n} \cdot \frac{{}^{n}\!/_{\log^c n} \cdot c' \log^2 n}{n\delta(1-o(1))} \right)^{\gamma' \log n} \\
&\qquad \leq \left(\frac{e \cdot c'}{\gamma' \log^{c-q-1} n \cdot \delta(1-o(1))} \right)^{\gamma' \log n} = o\left(\frac{1}{n^2} \right),
\end{aligned}$$

with γ' large enough and $|\mathcal{U}(s)| = {}^{n}\!/_{\log^c n}$. □

Claim 5.2. *Each node $v \in \mathcal{U}^{(2)}(s)$ has no inner edges, with high probability.*

Proof. According to Lemma 5.6 it holds that $\mathcal{U}^{(2)}(s) = \emptyset$ with high probability. □

Claim 5.3. *Each node $v \in \mathcal{U}^{(3)}(s)$ has no inner edges, with high probability.*

Proof. According to Lemma 5.6 it holds that $\mathcal{U}^{(3)}(s) = \emptyset$ with high probability. □

By definition it holds that $\mathcal{U}(s) := \mathcal{U}^{(1)}(s) \cup \mathcal{U}^{(2)}(s) \cup \mathcal{U}^{(3)}(s)$. To obtain a proper bound on $|\mathcal{U}_1(s)|$, we consider each of the groups $\mathcal{U}^{(1)}(s)$, $\mathcal{U}^{(2)}(s)$, and $\mathcal{U}^{(3)}(s)$ separately. By definition we obtain that $\sum_{\forall v \in \mathcal{U}(s)} \deg_v \leq |\mathcal{U}(s)| \cdot c' \log n \cdot \log\log n$. Consequently, each node in $\mathcal{U}^{(1)}(s)$ is incident to at least one inner edge with probability at most

$$1 - \left(1 - \frac{|\mathcal{U}(s)| \cdot c' \log n \cdot \log\log n}{n\delta(1-o(1))} \right)^{\log^q n}.$$

Thus, it holds that

$$\begin{aligned}
|\mathcal{U}_1(s)| &\leq \mathcal{O}\left(|\mathcal{U}(s)|\right) \left[1 - \left(1 - \frac{|\mathcal{U}(s)| \cdot c' \log n \cdot \log\log n}{n\delta(1-o(1))} \right)^{\log^q n} \right] + |\mathcal{U}^{(2)}(s)| + |\mathcal{U}^{(3)}(s)| \\
&\leq \mathcal{O}\left(|\mathcal{U}(s)|\right) \left[1 - e^{-1/\left(\log^{c-1-q} n \cdot \Omega(\sqrt{\delta})\right)} \right] + |\mathcal{U}^{(2)}(s)| + |\mathcal{U}^{(3)}(s)| \\
&\leq \mathcal{O}\left(\frac{|\mathcal{U}(s)|}{\log^{c-1-q} n} \right) + |\mathcal{U}^{(2)}(s)| + |\mathcal{U}^{(3)}(s)|. \qquad (5.10)
\end{aligned}$$

It remains to estimate $|\mathcal{U}^{(2)}(s)| + |\mathcal{U}^{(3)}(s)|$. According to Lemma 5.6 we obtain that

$$|\mathcal{U}^{(2)}(s)| + |\mathcal{U}^{(3)}(s)| = 0$$

holds with high probability. Further, note that $\frac{|\mathcal{U}(s)|}{\log^{c-1-q} n} = \omega(\log n)$. Thus, for a properly chosen constant q with $2q < c$, the desired upper bound on $\mathcal{U}_1(s)$ follows by applying standard Chernoff bounds.

On the other hand, since the minimum degree is given by δ, the lower bound is given by

$$\begin{aligned}
|\mathcal{U}_1(s)| &\geq \Omega\left(|\mathcal{U}(s)|\right)\left[1 - \left(1 - \frac{|\mathcal{U}(s)|\delta}{n\delta(1-o(1))}\right)^{\delta}\right] \\
&\geq \Omega\left(|\mathcal{U}(s)|\right)\left[1 - e^{\delta/\log^c n \cdot (1-o(1))}\right] \\
&\geq \Omega\left(\frac{|\mathcal{U}(s)|\delta}{\log^c n \cdot (1-o(1))}\right).
\end{aligned}$$

Now we show the second part of our claim, i.e., the upper bound on $\mathcal{U}(s+1)$. According to Claim 5.3, each node $v \in \mathcal{U}^{(3)}(s)$ is incident to 0 inner edges. Likewise, according to Claim 5.2, the nodes in $\mathcal{U}^{(2)}(s)$ are incident to 0 inner edges. Let $\mathcal{U}_1^{(1)}(s), \mathcal{U}_1^{(2)}(s), \mathcal{U}_1^{(3)}(s)$ be the above grouping taking into account nodes in $\mathcal{U}_1(s)$ only. Then, according to Claim 5.1, Claim 5.2, and Claim 5.3, the amount of inner edges in $\mathcal{U}_1(s)$ is upper bounded by

$$|\mathcal{U}_1^{(1)}(s)|\gamma' \log n \leq |\mathcal{U}_1(s)|\gamma' \log n. \tag{5.11}$$

Note that Equation 5.11 holds with high probability. Since we know the upper bound on the amount of inner edges within the set of uninformed nodes and the total amount of yet unrevealed edges, it follows that a node in $\mathcal{U}^{(1)}(s)$ is incident to ρ inner edges with probability at most $\binom{\log^q n}{\rho}\left(\mathcal{O}\left(\frac{|\mathcal{U}_1(s)|\cdot\log n}{n\delta(1-o(1))}\right)\right)^{\rho}$.

Thus, the expected amount of nodes in $\mathcal{U}(s+1)$ is upper bounded by

$$|\mathcal{U}^{(1)}(s)|\binom{\log^q n}{\rho}\left(\mathcal{O}\left(\frac{|\mathcal{U}_1(s)|\cdot\log n}{n\delta(1-o(1))}\right)\right)^{\rho}.$$

We obtain our second statement by using a Vertex Exposure Martingale (cf. Definition 4.3) and the Azuma Hoeffding Inequality (cf. Lemma 4.4) as we did before. □

Finally, we are able to examine M's dissemination throughout the entire broadcasting algorithm.

Theorem 5.2. *Let G be a random network as described in Section 5.1.1. Further, let $\delta = \gamma(\log\log n)^2$, where γ is some (large) constant. Then, with high probability, Algorithm 5.1 informs all nodes of G within $\mathcal{O}(\log n)$ many rounds by using $\mathcal{O}(n \log\log n)$ many message transmissions.*

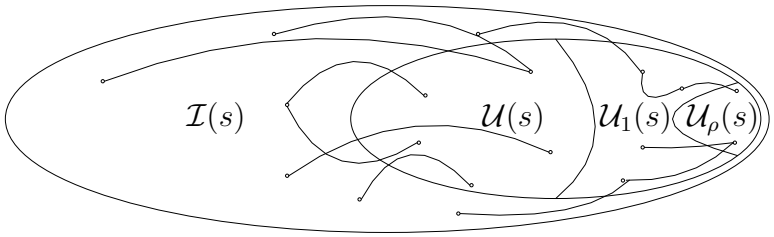


Figure 5.3: *The structure of the network after Phase 2 has ended. The vast majority of nodes in $\mathcal{U}$ is connected to nodes in $\mathcal{I}$ only.*

Proof. According to Lemma 5.4, Algorithm 5.1 informs at least $\frac{n}{2}$ different nodes in $\mathcal{O}(\log n)$ many rounds using $\mathcal{O}(n)$ many message transmissions. After a few additional rounds and message transmissions we obtain that $|\mathcal{U}(s)| \leq \frac{n}{\log^{9+\Omega(1)} n}$ (cf. Lemma 5.5), where s is the first round after Phase 2 of Algorithm 5.1.
Due to Lemma 5.7, we know that a $(1-o(1))$-fraction of the nodes in $\mathcal{U}(s)$ is connected to nodes in $\mathcal{I}(s)$ only. This situation is depicted by Figure 5.3. That is, only few nodes in $\mathcal{U}(s)$ possess at least one inner connection, i.e., a connection to some other node in $\mathcal{U}(s)$. On the other hand, a node $v \in \mathcal{U}$ may remain uninformed only if v has at least ρ inner connections. Otherwise, due to Algorithm 5.1, v would definitely choose at least one informed node and thus become informed. Keeping this property in mind we continue our analysis, which is similar to the one given in [12] in the context of random d-regular graphs.

The outline of this theorem is as follows. First, we show by induction that

$$\frac{|\mathcal{U}(s+i+2)|}{|\mathcal{U}(s+i+1)|} = o\left(\frac{|\mathcal{U}(s+i+1)|^2}{|\mathcal{U}(s+i)|^2}\right)$$

for $i \geq 0$, unless $|\mathcal{U}(j)| \leq \log^{q'} n$ for a (large) constant q' and some round j. Afterwards, we show that in additional $\mathcal{O}(\log\log n)$ many rounds all remaining nodes will be informed with high probability. In order to start the formal proof, we first need an auxiliary statement.

Claim 5.4. *For each uninformed node $u \in \mathcal{U}(s+i)$, with $i \geq 0$, u is incident to at most $O(\log n \cdot \log\log n)$ many inner edges, with high probability.*

Proof. By Lemma 5.6 we know that each node in $\mathcal{U}(s)$ has degree at most $O(\log n \cdot \log\log n)$. Since $\mathcal{U}(s+i) \subseteq \mathcal{U}(s)$, for all $i \geq 0$, the claim follows. □

From now on Algorithm 5.1 uses pull-transmissions only. Recall that a node may remain uninformed only if said node has at least ρ inner connections. Due to Lemma 5.4, it follows that the total amount of inner connections within $\mathcal{U}(s+i)$ is upper bounded by

$$\mathcal{O}(|\mathcal{U}(s+i)| \cdot \log n \cdot \log\log n).$$

Further, by Lemma 5.7 we know

$$|\mathcal{U}_1(s)| = \Omega\left(\frac{|\mathcal{U}(s)|\delta}{\log^c n}\right) \text{ and } |\mathcal{U}(s+1)| = \mathcal{O}\left(|\mathcal{U}(s)|\left(\log^{-(c-2)(\rho/2)} n\right)\right).$$

Consequently, for the base clause with $i = 0$, we obtain that

$$\mathbb{E}[|\mathcal{U}(s+2)|] \leq \mathcal{O}\left(|\mathcal{U}(s+1)|\left(\frac{|\mathcal{U}(s+1)| \cdot (\log n \cdot \log\log n)^2}{|\mathcal{U}_1(s)|}\right)^{\rho}\right),$$

which equals $o\left(\left(\frac{|\mathcal{U}(s+1)|}{|\mathcal{U}(s)|}\right)^2\right)$ for ρ large enough. Note that we obtain the above bound in the situation where:

1. each node $v \in \mathcal{U}_1(s) \setminus \mathcal{U}(s+1)$ provides exactly one yet unpaired stub incident to an inner edge of v, and
2. none of the stubs incident to the inner edges of the nodes in $\mathcal{U}(s+1)$ are paired yet.

Using similar Martingale techniques as in the proof of Lemma 5.7, we obtain that the upper bound on $\mathbb{E}[|\mathcal{U}(s+2)|]$ holds with high probability.

Now we prove the induction step where $i > 0$ holds. We begin by examining the structure of the network in order to compute a lower bound for $\mathbb{E}[|\mathcal{U}_1(s+i)|]$. To be more specific, we determine whether the inner edge of a node in $\mathcal{U}(s+i)$ is incident to a node in $\mathcal{U}(s+i)$ or to one in $\mathcal{U}_1(s+i-1) \setminus \mathcal{U}(s+i)$. Therefore, we consider one (randomly chosen) inner edge e of each node $v \in \mathcal{U}(s+i)$ individually. Note that at least one such inner edge must exist, since $v \in \mathcal{U}_1(s+i-1)$. Further, note that by considering the inner edges only, we do not have to take available stubs in $\mathcal{I}$ into account. To lower bound $\mathbb{E}[|\mathcal{U}_1(s+i)|]$ we consider the following scenario. Each node in $\mathcal{U}(s+i)$ has only one available stub left (its inner edge), while the nodes in $\mathcal{U}_1(s+i-1)$ have $\mathcal{O}(\log n \cdot \log\log n)$ many stubs available. Recall that each node in $\mathcal{U}_1(s+i-1)$ has at most $\mathcal{O}(\log n \cdot \log\log n)$ many inner edges according to Claim 5.4. Since $|\mathcal{U}(s+i-1)| \geq |\mathcal{U}_1(s+i-1)|$, we obtain that

$$\begin{aligned}\mathbb{E}[|\mathcal{U}_1(s+i)|] &= \Omega\left(|\mathcal{U}(s+i)| \cdot \frac{|\mathcal{U}(s+i)|}{\log n \cdot \log\log n \cdot |\mathcal{U}_1(s+i-1)|}\right) \\ &= \Omega\left(|\mathcal{U}(s+i)| \cdot \frac{|\mathcal{U}(s+i)|}{|\mathcal{U}(s+i-1)| \cdot \log n \cdot \log\log n}\right). \qquad (5.12)\end{aligned}$$

Now we derive a similar upper bound on $\mathbb{E}[|\mathcal{U}_1(s+i)|]$ and obtain that

$$\begin{aligned}\mathbb{E}[|\mathcal{U}_1(s+i)|] &= \mathcal{O}\left(|\mathcal{U}(s+i)| \cdot \frac{|\mathcal{U}(s+i)|(\log n \cdot \log\log n)^2}{|\mathcal{U}_1(s+i-1)|}\right) \\ &= \mathcal{O}\left(|\mathcal{U}(s+i)| \cdot \frac{|\mathcal{U}_\rho(s+i-1)| \log^3 n}{|\mathcal{U}_1(s+i-1)|}\right).\end{aligned}$$

Similarly as before, we obtain the above bound in the scenario where each node in $\mathcal{U}(s+i)$ provides $\Theta(\log n \cdot \log\log n)$ many unpaired (inner) stubs, whereas each node in $\mathcal{U}_1(s+i-1)$ has only one unpaired (inner) stub left.

In the following we estimate $\mathbb{E}[|\mathcal{U}_1(s+i)|]$ by bounding the amount of inner edges in $\mathcal{U}(s+i)$. First, observe that the expected amount of inner edges may differ from the expected amount of nodes with inner edges by at most a polylogarithmic factor. Now consider the following process P': each stub of each node $u \in \mathcal{U}(s+i)$ is considered an inner edge independently with probability

$$p' := \frac{|\mathcal{U}(s+i)| \cdot \deg' - 1}{|\mathcal{U}_1(s+i-1)| \cdot \deg'' - 1},$$

where $\deg'$ and $\deg''$ represent the average amount of stubs available for pairing in $\mathcal{U}(s+i)$ and $\mathcal{U}_1(s+i-1)$, respectively. Note that a node $u \in \mathcal{U}(s+i) \subset \mathcal{U}_1(s+i-1)$ trying to pair a stub s must choose another stub $s' \neq s$ to pair s with. Let $x_{k,j}$ be the indicator variable for the event that the kth stub of node $v_j \in \mathcal{U}(s+i)$ is considered an inner edge and let $X := \sum_{v_j \in \mathcal{U}(s+i)} \sum_{s_k \in s_f(v_j)} x_{k,j}$. Then, with $\deg_u = \mathcal{O}(\log n \cdot \log\log n)$, it follows that

$$\begin{aligned} \mathbb{E}[X] &= \Omega\left(|\mathcal{U}(s+i)| \frac{|\mathcal{U}(s+i)|}{|\mathcal{U}_1(s+i-1)| \cdot \log n \cdot \log\log n}\right), \text{ and} \\ \mathbb{E}[X] &= \mathcal{O}\left(|\mathcal{U}(s+i)| \frac{|\mathcal{U}(s+i)| \cdot (\log n \cdot \log\log n)^2}{|\mathcal{U}_1(s+i-1)|}\right). \end{aligned} \tag{5.13}$$

If $\mathbb{E}[X] \geq \log^{q'-2} n$, we use the analysis below to state that $X \leq (p' \cdot |s(\mathcal{U}(s+i))|)(1+o(1))$ with high probability. Otherwise, we skip this part and go directly to the one where the case $|\mathcal{U}(j)| \leq \log^{q'} n$ is considered. Note that in this case $\mathcal{O}(X \cdot \log n \cdot \log\log n) \leq \log^{q'} n$, which implies that $|\mathcal{U}| \leq \log^{q'} n$.

Suppose that $\mathbb{E}[X] \geq \log^{q'-2} n$ and let P be the original process where $s_1, ..., s_k$ are the stubs considered for pairing. Further, let $s'_1, ..., s'_k$ be the stubs of process P'. In the following we consider process P and P' up to stub $s_{\lfloor k/4 \rfloor}$ and $s'_{\lfloor k/4 \rfloor}$, respectively. Consequently, the probability for each stub s_j, with $j \leq \lfloor k/4 \rfloor$, of P to be part of an inner edge is not less than half the probability of s_1. Now we show that the amount of inner edges of P' up to s'_k is at most $4(1+o(1))$ times the amount of inner edges of P' up to $s'_{\lfloor k/4 \rfloor}$. Observe that $k \geq \log^{q'-2} n$ must hold, since $\mathbb{E}[X] \geq \log^{q'-2} n$. Further, by definition, the choices of each stub of P' are independent. Therefore, by applying Chernoff bounds, we obtain that at least $(p'k/4)(1-o(1))$ many stubs of $s'_1, ..., s'_{\lfloor k/4 \rfloor}$ are considered inner edges with high probability. More formally, $\Pr(X \leq (1-\delta') \cdot \mathbb{E}[X]/4) \leq 1/n^2$, where X is the mentioned event and $\delta' = 1/\log n$. Similarly, we obtain that at most $(p'k)(1+o(1))$ many stubs of $s'_1, ..., s'_k$ are considered inner edges.

Therefore, it follows that $\mathbb{E}[X]/16 = \mathcal{O}\left(|\mathcal{U}_1(s+i)| \cdot \log n \cdot \log\log n\right)$. Hence,

$$\Omega\left(X \cdot (\log n \cdot \log\log n)^{-1}\right) \ni |\mathcal{U}_1(s+i)| \in \mathcal{O}\left(X \cdot \log n \cdot \log\log n\right). \tag{5.14}$$

Note that the second inequality is derived from the situation where all inner edges belong to distinct nodes only.

Consequently, whenever $\rho \geq 20$, it holds that

$$\begin{aligned}
\mathbb{E}[|\mathcal{U}(s+i+2)|] &\leq |\mathcal{U}(s+i+1)| \left(\frac{|\mathcal{U}(s+i+1)|\mathcal{O}((\log n \cdot \log\log n)^2)}{|\mathcal{U}_1(s+i)|} \right)^{\rho} \\
&\leq |\mathcal{U}(s+i+1)| \left(\frac{|\mathcal{U}(s+i+1)|}{|\mathcal{U}(s+i)|} \cdot \frac{|\mathcal{U}(s+i-1)|}{|\mathcal{U}(s+i)|} \cdot \mathcal{O}(\log^6 n) \right)^{\rho} \\
&\overset{(IH)}{\leq} |\mathcal{U}(s+i+1)| \left(\sqrt{\frac{|\mathcal{U}(s+i+1)|}{|\mathcal{U}(s+i)|}} \cdot \mathcal{O}(\log^6 n) \right)^{\rho} \\
&\leq |\mathcal{U}(s+i+1)| \left(\frac{|\mathcal{U}(s+i+1)|}{|\mathcal{U}(s+i)|} \right)^{2} \cdot \mathcal{O}\left(\frac{|\mathcal{U}(s+i+1)|}{|\mathcal{U}(s+i)|} \right)^{\rho/2-6}.
\end{aligned}$$

The second inequality is obtained due to Equation 5.12, the Estimation 5.14, and the fact that $\log^{c'} n \cdot (\log\log n)^{\mathcal{O}(1)} \leq \log^{c'+1} n$ holds for any positive constant c'. The third equation follows directly from our induction hypothesis. Further, note that $\mathcal{O}\left(\frac{|\mathcal{U}(s+i+1)|}{|\mathcal{U}(s+i)|}\right)^4 = o\left(1/\log^6 n\right)$, since

$$\frac{|\mathcal{U}(s+i+1)|}{|\mathcal{U}(s+i)|} \leq \frac{|\mathcal{U}(s+1)|}{|\mathcal{U}(s)|} \leq \frac{1}{\log^{\rho} n}.$$

Similar as before, we conclude that $|\mathcal{U}_\rho(s+i+1)|$ is upper bounded by

$$\mathcal{O}\left(\mathbb{E}[|\mathcal{U}_\rho(s+i+1)|] \cdot \log n \cdot \log\log n\right)$$

with high probability. Consequently, the invariant follows.

It remains to show that the yet uninformed nodes in $\mathcal{U}(j)$, with $|\mathcal{U}(j)| \leq \log^{q'} n$ for some round j, will be informed within $\mathcal{O}(\log\log n)$ additional rounds, with high probability. Note that until now each node in $\mathcal{U}(j)$ used pull transmissions only. Further, note that the amount of remaining inner edges of each node in $\mathcal{U}(j)$ determines its probability to remain uninformed. This holds true for all rounds $i \geq j$. Additionally, we know that each node in $\mathcal{U}(j)$ has degree at most $n^{1/\alpha} + \delta < 2 \cdot n^{1/\alpha}$, thus $|s_f(\mathcal{U}(j))| \leq 2 \cdot \log^{q'} n \cdot n^{1/\alpha}$.

Let $S \subset V$ be a random set of nodes of size $|\mathcal{U}(j)|$. It follows that each node in S has degree at most $2 \cdot n^{1/\alpha}$. In the following we focus on pairs of nodes within S. Note that there are $|S|^2/2$ many different pairs overall. Let $x_{v,w}$ be the event that $\{v, w\} \in E$ for the nodes $v, w \in S$. The probability for $x_{v,w}$, i.e., the probability for the two nodes to be connected, is at most

$$\frac{4 \cdot n^{1/\alpha} \cdot n^{1/\alpha}}{n} = 4 \cdot n^{-(1-2/\alpha)}.$$

Recall that v and w have degree at most $2 \cdot n^{1/\alpha}$ and the network provides $\Omega(n\delta)$ many free stubs. Further, note that the above events are negatively correlated, since each

connection between a pair consumes two stubs of S. Consequently, on expectation, there are at most

$$\mu := \frac{|S|^2}{2} \cdot \frac{4}{n^{1-2/\alpha}}$$

many pairs connected in S. With $\delta' := \frac{\rho \cdot n^{1-2/\alpha}}{8 \cdot |S|} - 1$, standard Chernoff bounds imply that more than $|S|\rho/4$ many inner edges exist in S with probability at most

$$\left[\frac{e^{\delta'}}{(1+\delta')^{1+\delta'}}\right]^{\mu} \leq \left[\frac{e^{\delta'}}{\left(\frac{\rho \cdot n^{1-2/\alpha}}{8 \cdot |S|}\right)^{1+\delta'}}\right]^{\mu} \leq \left(\frac{8e \cdot |S|}{\rho \cdot n^{1-2/\alpha}}\right)^{|S| \cdot \rho/4}.$$

Therefore, w.h.p., at most $|S|/4$ many nodes can have at least ρ inner connections, and thus may remain uninformed. Since there are at most $\binom{n}{|S|}$ many different sets of size $|S|$, we obtain that

$$\binom{n}{|S|} \cdot \left(\frac{8e \cdot |S|}{\rho \cdot n^{1-2/\alpha}}\right)^{|S| \cdot \rho/4} \leq \left[\mathcal{O}\left(\frac{|S|^{\frac{\rho}{4}-1}}{n^{(1-2/\alpha)\frac{\rho}{4}-1}}\right)\right]^{|S|} = \frac{1}{n^{1+\Omega(1)}},$$

whenever ρ is large enough. □

5.1.3 Glossary

At this point we list some of the most important definitions we used in the above analysis.

Definitions concerning the node states

$\mathcal{U}(i)$	Set of uninformed nodes in round i.
$\mathcal{U}_l(i)$	Set of uninformed nodes with at least l neighbors in $\mathcal{U}(i)$.
$\mathcal{U}_{G_k}(i)$	Set of uninformed nodes in G_k in round i.
$\mathcal{I}^+(i)$	Set of nodes that became informed in round $i-1$ for the first time.
$\mathcal{I}_{G_k}(i)$	Set of nodes in G_k that are informed in round i.

Definitions concerning the network

G	G is the undirected random graph, also called network, all messages are broadcasted in.
n	Number of nodes in G.
m	Number of edges in G.
$\deg_v$	Degree of node $v \in V$.
$\Gamma(v)$	Neighborhood of node v.
δ	Minimum degree of G. It holds that $\delta \geq \gamma(\log\log n)^2$.
γ	A large constant used to define δ.
α	The power law exponent used to determine G's degree distribution. Throughout our analysis we assume that $\alpha > 3$.
G_j	G_j is the set (also called group) containing all nodes with degree $2^j\delta$ up to $2^{j+1}\delta - 1$.
deg_{G_j}	Highest degree of group G_j.
$s(v), s_i^v$	$s(v)$ is the set of all *stubs* of a node $v \in V$, with s_i^v being the ith stub in an ordered sequence.
$s_f(v), \overline{s_f(v)}$	The set $s(v)$ is divided into a set of free stubs $s_f(v)$, and a set of already paired/used stubs $\overline{s_f(v)}$. Hereby free stubs are available for pairing, whereas used stubs are already connected to edges in G.
$S_f(G)$	Sum of all free stubs in the corresponding round. Since said round is always clear from the context, we omit to specify it explicitly.

5.2 Conquer an Epidemic Type Process using an Epidemic Type Process

In the following we consider a static network with power law degree distribution similar to the one presented in Section 5.1. We examine different types of counteracting dissemination processes in order to neutralize a malicious rumor. Since we already know that broadcasting is pretty efficient in this kind of networks, one can wonder how powerful a counteracting dissemination process must be in order to neutralize a malicious one. Clearly, the effectiveness of such a counteracting dissemination process does not only rely on the behavior of the malicious one, but also on the network's properties. It turns out that the nodes with large degree, although only a minority, play a crucial role in this endeavor.

To be more specific, we consider two models, which share some properties. In the first model we assume a constant delay of d rounds between the two processes, i.e., the counteracting process Φ performs its actions in round i on each node that was originally infected by the malicious process $\mathcal{V}$ in round $i - d$, and a minimum node degree of $\Omega(\log \log n)$ (see Section 5.2.1). Under these assumptions we show that an infection disseminated by the malicious process $\mathcal{V}$ does not infect more than a $o(1)$-fraction of the network before it is eliminated by the counteracting process Φ, with high probability. Note that $\mathcal{V}$ spreads over a constant amount of randomly chosen edges of each infected node. On the other hand, each node that Φ is activated on for the very first time sends a warning to all its neighbors for one time only. This warning then immunizes the recipients, excluding the currently infected ones, permanently.

In the second model the network's minimum degree is constant. This time the counteracting process Λ is more complicated. Similarly to process Φ, process Λ is activated on an infected node after a constant delay d. Then each newly activated node v forms a new notification tree with root v. In short, a notification tree is a tree-like structure, starting with the root node, representing a "who-notified-whom-relationship" of the contained nodes (see Section 5.2.1 for a definition). The maximum path length of each notification tree from its root v to its leaves is limited by $\beta \left\lceil \log^2(\deg_v) \right\rceil$. We show that process $\mathcal{V}$ does not infect more than $o(n)$ many nodes while being counteracted by Λ with high probability.

5.2.1 Theoretical Models

In the following we present some basic concepts, models, and notation needed for the understanding of the upcoming formal analysis. We start by describing the network (also called environment), which is an undirected graph used by the dissemination processes we examine in more detail. Afterwards, said dissemination processes are defined precisely.

The Network We model the network used by the message dissemination processes as an undirected random graph $G = (V, E)$ with n nodes and m edges. The degrees are randomly assigned according to a power law distribution. Let $\deg_v$ denote the degree of a node $v \in V$. Then v has degree $\deg_v = \deg$, where $\deg \geq \delta$, with probability proportional to $\deg^{-\alpha}$, where the constant $2 < \alpha < 3$ is the so-called power law exponent and $\delta \in \mathcal{O}(\log \log n)$ denotes the minimum degree in G. It follows that $m = 1/2 \sum_{\forall v \in V} \deg_v = \Theta(n \cdot \delta)$. Further, let $s(v)$ be the set of all *stubs* of a node $v \in V$, with s_i^v being the ith stub in an ordered sequence, and let $s(v)$ be divided into a set of already paired/used stubs $\overline{s_f(v)}$ and a set of free stubs $s_f(v)$. Hereby, free stubs are available for the pairing process, whereas used stubs are already connected to edges in G. Additionally, we define the maximum degree to be proportional to $n^{1/\alpha}$.

Edges in G are randomly created using the so-called *Configuration Model* (cf. Definition 4.1) and the *Principle of Deferred Decision* (cf. Definition 4.2). Note that the application of the Configuration Model may generate self-loops, disconnected components, and multiple edges. However, since $\delta \in \Omega(1)$ large enough and $2 < \alpha < 3$, G will be connected with high probability with at most a negligible amount of self loops and double edges. Additionally, all nodes have access to a global clock and work synchronously.

Each node has a status indicating its state of knowledge with respect to the dissemination processes. This status partitions the set of nodes into three subsets:

1. $\mathcal{U}$, meaning the set of all (uninformed/susceptible) nodes that are not notified yet in any way,

2. $\mathcal{I}$, meaning the set of all (infective/contagious) nodes that were already notified (also called infected) exclusively by the dissemination process $\mathcal{V}$ and did not recover so far, and

3. $\mathcal{R}$, meaning the set of all (recovered) nodes v formerly part of $\mathcal{I}$ that were healed and all nodes v' formerly part of $\mathcal{U}$ that were already notified/immunized exclusively by the dissemination process Φ or Λ.

Hereby a node $v \in \mathcal{I}$ recovers (or heals), i.e., v is moved to $\mathcal{R}$, automatically after a specific delay d, i.e., d rounds after the one v was initially infected by $\mathcal{V}$. The corresponding sets with respect to a specific round i are denoted by $\mathcal{U}(i), \mathcal{I}(i)$ and $\mathcal{R}(i)$, respectively. If it is clear from the context, we simply write $\mathcal{U}, \mathcal{I}$ and $\mathcal{R}$.

The Dissemination Processes All nodes in the network represent entities the message dissemination processes try to notify. We consider two main types of dissemination processes:

1. harmful processes, i.e., dissemination processes aiming to inflict damage to the nodes, and

2. remedying processes, i.e., dissemination processes counteracting the effects of the previously mentioned harmful processes.

In the following we consider three competing dissemination processes in total. A *harmful process* $\mathcal{V}$ acting with the goal to infect as many nodes with a malicious piece of information as possible, and two *remedying processes* Φ *and* Λ anxious to stop $\mathcal{V}$. Note that only one of the remedying processes, depending on the model, is used to counteract the harmful one.

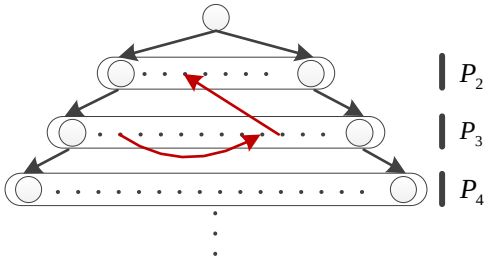


Figure 5.4: *An arbitrary dissemination process P in G. P_i depicts the set of nodes of said process that are allowed to send in round i. Possible conflicts are indicated in red. For example, newly informed nodes in round i may pair stubs with nodes that were informed in round $j \leq i$. Obviously, contacting an already informed node again does not aid the dissemination of P.*

Process $\mathcal{V}$ The harmful dissemination process is called $\mathcal{V}$. Its goal is to infect as many nodes of the network as possible and thus inflict damage and/or costs. In the following we often use the terms *notify* and *infect* with respect to $\mathcal{V}$ synonymously. In short, in each round i process $\mathcal{V}$ proceeds as follows. Each newly infected node, i.e., each node that was infected in round $i-1$, *tries to infect* $\gamma(\mathcal{V})$ many randomly chosen neighbors in round i. Recall that only thus far uninfected nodes may be infected. Now we are ready to describe $\mathcal{V}$ more formally.

Let $u \in V$ be the starting point for $\mathcal{V}$, i.e., the first node infected by $\mathcal{V}$. In order to simplify the formal analysis we assume that u is chosen uniformly at random from the set of all nodes. In the first round $\mathcal{V}$ is only able to notify neighbors of u, i.e., $\mathcal{V}$ may notify up to $\gamma(\mathcal{V})$ many different neighbors $\{v_1, ..., v_{\gamma(\mathcal{V})}\} \subseteq \Gamma(u)$ of u, where $3 \leq \gamma(\mathcal{V}) < \delta$ and $\Gamma(u) := \{v \in V \mid \{v, u\} \in E\}$. Let $\mathcal{V}_i := \mathcal{I}(i) \setminus \mathcal{I}(i-1)$ be the set of nodes $\{v_1, ...v_j\}$ of G that were notified by process $\mathcal{V}$ in round $i-1$. Then each node $v \in \mathcal{V}_i$ chooses $\gamma(\mathcal{V})$ many of its neighbors $\{v_1, ..., v_{\gamma(\mathcal{V})}\} \subseteq \Gamma(v)$ uniformly at random and tries to infect them.

Consequently, $\mathcal{V}_0 = \{u\}$ while $\mathcal{V}_1$ is a randomly chosen subset of $\Gamma(u)$ of size $\gamma(\mathcal{V})$. Generally speaking it holds that

$$\mathcal{V}_{i+1} := \left(\mathcal{U}(i) \cap \left(\bigcup_{u \in \mathcal{V}_i} \xi(u)\right)\right) \setminus \Psi_i,$$

where $\xi(u)$ is a randomly chosen subset of $\Gamma(u)$ of size $\gamma(\mathcal{V})$ and Ψ_i is the set of uninformed nodes contacted by Φ or Λ, i.e., the corresponding remedying process, in round i. That is, whenever $\mathcal{V}$ and Φ or Λ simultaneously notify the same uninformed node, said node is considered immunized. Figure 5.4 illustrates the behavior of an arbitrary dissemination process P in G.

Process Φ Process Φ is one of the two remedying processes used to counteract $\mathcal{V}$. By definition its goal is to counteract $\mathcal{V}$ by reactively curing all infected nodes and proactively immunizing possible contact points of $\mathcal{V}$. The extinction of $\mathcal{V}$ is the ultimate goal. In the following we often use the terms *notify* and *immunize* with respect to Φ synonymously. In short, process Φ is *activated*, i.e., Φ performs its actions, on each infected node after some delay of d rounds. W.l.o.g. let v be such a node and let i be the round in which Φ is activated on v. Then v notifies all its neighbors, i.e., v immunizes all its neighbors, in round i. All of v's yet uninformed neighbors become immunized and v recovers. However, all of v's currently infected neighbors remain infected. Note that Φ is only activated on previously infected nodes, i.e., uninformed nodes that become immunized do not actively participate in Φ themselves in any way.

Let $\Phi_i := \mathcal{V}_{i-d}$ be the set of nodes $\{v_1, ...v_j\}$ of G at which Φ is activated on in round i. It follows that process Φ is activated on each node v in round i if and only if $v \in \mathcal{I}(i-d) \setminus \mathcal{I}(i-d-1)$. Further, once Φ is activated on a node v, said node notifies all its yet uninformed neighbors and remains silent for the rest of the dissemination process. More formally,

$$\begin{aligned}
\forall i \geq d \ : \Phi_i &:= \mathcal{V}_{i-d}, \\
\mathcal{I}(i) &:= \bigcup_{i-d<j\leq i} \mathcal{V}_j, \\
\mathcal{U}(i) &:= V \setminus \left(\mathcal{I}(i) \cup \bigcup_{j<i} \Gamma(\Phi_j)\right), \\
\mathcal{R}(i) &:= V \setminus (\mathcal{U}(i) \cup \mathcal{I}(i)).
\end{aligned}$$

Process Λ Process Λ is our second remedying process. By definition its goal is identical to the one of Φ. However, its behavior is far more complex. In the following we often use the terms *notify* and *immunize* with respect to Λ synonymously.

Before we are able to describe Λ, we need to define the term *notification tree*. In short, a notification tree is a tree-like structure, starting with the root node, representing a

"who-notified-whom-relationship" of the contained nodes. More formally, a *notification tree* Υ_v with root $v \in V$ in round i, where $\Upsilon_{v,i}$ refers to the nodes of said tree that are participating in Λ in the ith round, and depth j[1] is a tree with the following property: each node in $\Upsilon_{v,i}$ was contacted by another node of $\Upsilon_{v,i-1}$ provided that $j \geq 1$.

In the following let

$$\Lambda_i := \bigcup_{\forall v \in \mathcal{R}(i)} \Upsilon_{v,i}$$

be the set of nodes that participate in Λ in round i. That is, Λ_i contains all nodes forming root nodes of new notification trees in round i as well as all nodes that were notified by Λ in round $i-1$.

At this point we can finally define $\Upsilon_{v,i}$. Let $t(v)$ be the round in which v was initially infected by $\mathcal{V}$. Additionally, in the case v was not infected by $\mathcal{V}$ by round i, we define $t(v) = -\infty$ for all rounds $j \leq i$. Further, let $\Gamma(V') := \bigcup_{v' \in V'} \Gamma(v')$ for some subset $V' \subseteq V$. Then

$$\Upsilon_{v,i} := \begin{cases} \{\emptyset\} & i < t(v) + d \\ \{v\} & t(v) + d \leq i \leq t(v) + d + 1 \\ \Gamma(\Upsilon_{v,i-1}) \setminus \mathcal{I}(i-1) & t(v) + d + 1 < i \leq t(v) + d + \beta \left\lceil \log^2(\deg_v) \right\rceil + 1 \\ \{\emptyset\} & i > t(v) + d + \beta \left\lceil \log^2(\deg_v) \right\rceil + 1 \end{cases}$$

Note that infected nodes do not become part of a notification tree. Additionally, a root node v of a newly created notification tree in round $t(v) + d$ generates and transmits a preventive message in round $t(v) + d + 1$ for the first time. Further, note that the maximum path length of a notification tree from its root v to its leaves is limited by $\beta \left\lceil \log^2(\deg_v) \right\rceil$. In summary, we obtain that

$$\begin{aligned} \forall i \geq d \ : \mathcal{I}(i) &:= \bigcup_{i-d<j\leq i} \mathcal{V}_j, \\ \mathcal{R}(i) &:= \bigcup_{\forall v \in V} \bigcup_{j \leq i} \Upsilon_{v,j} \\ \mathcal{U}(i) &:= V \setminus (\mathcal{R}(i) \cup \mathcal{I}(i)) \,. \end{aligned}$$

As stated above, all nodes work synchronously. For the analysis we divide each round into three consecutive phases:

1. In the first phase all nodes that are scheduled to create new notification trees in this round are moved to $\mathcal{R}$.

2. In the second phase all other status changes scheduled for the current round are executed. That is, the sets $\mathcal{V}, \Lambda, \mathcal{I}, \mathcal{U}$, and $\mathcal{R}$ are updated according to the transmissions received by the end of the previous round.

[1] Let $t(v)$ be the round v was initially infected. Then the *depth* j of Υ_v in round i is defined as $j = i - t(v) - d - 1$ for all $i \geq t(v) + d + 1$ and 0 otherwise.

3. In the third phase then all nodes perform their scheduled transmissions for the current round.

At this point the description of Λ is quite intuitive. In order to give process $\mathcal{V}$ a head start, we define Λ to be inactive up to round $i' = c_1 \log\log n$, for some large constant c_1. That is, $\mathcal{V}$ is allowed to spread in the network up to a point where it might be a threat. Therefore, in all rounds $j' < i'$ it holds that $\Lambda_{j'} = \emptyset$. In round i' then we immunize all nodes that were infected until round $i' - d - 1$, i.e., these nodes are moved to $\mathcal{R}$. Additionally, similar to process Φ, process Λ is activated on each node v in an arbitrary round $i \geq i'$ if and only if $v \in \mathcal{I}(i-d) \setminus \mathcal{I}(i-d-1)$. Each such newly activated node v then forms a new notification tree.

Models We consider two fixed combinations of the harmful and the remedying processes. The combination of $\mathcal{V}$ and Φ is called *Model 1* and the one of $\mathcal{V}$ and Λ *Model 2*. In Model 1 it holds that $\delta = \Omega(\log\log n)$, whereas $\delta = \Theta(1)$ holds in Model 2.

5.2.2 Analysis

In the following we present a formal analysis for Model 1 and Model 2. As we mentioned earlier, we use the Principle of Deferred Decision and the Configuration Model during these analyses (cf. Section 4). Note that, for the reader's convenience, the most important definitions are listed in Section 5.2.3 in a more concentrated form. Throughout this section let $s(V') := \bigcup_{v' \in V'} s(v')$ be defined as the set of stubs for some subset $V' \subseteq V$. Further, let $s_f(V') := \bigcup_{v' \in V'} s_f(v')$ be defined as the set of free stubs for some subset $V' \subseteq V$. Additionally, we need an auxiliary structure. Let G_k be defined as the set (also called group) containing all nodes with degree k. Likewise, let $G^{\mathcal{U}}_{k,i} := \{u \in \mathcal{U}(i) \mid \deg_u = k\}$ be the group of all yet uninformed nodes in round i with degree exactly k.

5.2.2.1 Model 1 - $\mathcal{V}$ competing with Φ

In Model 1 process $\mathcal{V}$ competes with Φ. Recall that Φ is activated on each newly infected node v after a specific delay d. Then each such node notifies its (non-infected) neighborhood and remains silent afterwards (cf. Section 5.2.1). Obviously, the dissemination of Φ depends on the dissemination of $\mathcal{V}$. In the following we primarily focus on the amount of edges between the infected nodes and the uninfected ones. We show that the set of infected nodes is connected to a sufficient amount of uninfected ones, whenever the set of infected nodes is large enough, i.e., whenever $|\mathcal{I}| = \Theta(n/\delta^{\epsilon})$ holds for some suitable constant ϵ. We argue that $|\Gamma(\mathcal{I}) \cap \mathcal{U}| = |\mathcal{U}|(1 - o(1))$, where $\Gamma(V') := \bigcup_{v' \in V'} \Gamma(v')$, holds under this circumstance. Consequently, immunizing the neighborhood of each infected node implies that $\mathcal{V}$ cannot infect more than $o(n)$ many nodes overall.

Our general approach We consider the following (slightly different) process in order to model the dissemination of $\mathcal{V}$ and Φ. In this process we assume that only process $\mathcal{V}$ is applied to the nodes of the network G as long as $|\mathcal{I}(t)| < n/\delta^\epsilon$ for some small constant ϵ to be determined later. Here $\mathcal{I}(t)$ denotes the set of nodes notified by process $\mathcal{V}$ that are still infected *after round* t. Note that infected nodes do not get automatically healed after d many rounds (see below).

At this point we are ready to describe the dissemination process we analyze below in more detail. Let $\mathcal{I}^+(t)$ represent the set of nodes that become infected in round t. Then we first check for each $v \in \mathcal{I}^+(t)$ whether $|\Gamma(v) \cap \mathcal{I}(t-d-1)| \geq 1$ at the beginning of round $t+1$. That is, for each newly infected node we check if said node is connected to a neighbor that Φ would have already been executed on in the original process. If so, v is moved from $\mathcal{I}(t)$ to $\mathcal{R}(t)$. In other words, we set $\mathcal{R}(t) := \{v \in \mathcal{I}^+(t) \mid |\Gamma(v) \cap \mathcal{I}(t-d-1)| \geq 1\}$ and $\mathcal{I}(t) := \mathcal{I}(t-1) \cup (\mathcal{I}^+(t) \setminus \mathcal{R}(t))$. Afterwards, each node of $\mathcal{I}^+(t) \setminus \mathcal{R}(t)$ chooses $\gamma(\mathcal{V})$ many neighbors uniformly at random and infects these neighbors according to $\mathcal{V}$. These nodes form the set $\mathcal{I}^+(t+1)$. Once $|\mathcal{I}(t)| \geq n/\delta^\epsilon$ for some small constant ϵ and a round t, we apply process Φ once to immunize all nodes of $\Gamma(\mathcal{I}(t)) \cap H(t)$. Here $\Gamma(\mathcal{I}(t))$ represents the neighborhood of $\mathcal{I}(t)$, and the set of all susceptible nodes is given by $H(t) := V \setminus (\mathcal{I}(t) \cup_{i=1}^{t} \mathcal{R}(i))$. We show that, in the end, there will be at most $o(n)$ many nodes that have been infected during the whole process with high probability.

Theorem 5.3. *Process $\mathcal{V}$ infects at most $o(n)$ many nodes in total, with high probability.*

Proof. First, we prove a structural lemma w.r.t. the distribution of the edges between $\mathcal{I}(t)$ and $H(t)$ when $|\mathcal{I}(t)| \geq n/\delta^\epsilon$, where ϵ is some small constant, for the first time.

Recall that $G = (V, E)$ represents the original (random) graph used by all dissemination processes. We define $\mathcal{G}(d_1, \ldots, d_l)$ as the probability space of all graphs with l nodes and degree sequence $(d_1, \ldots, d_l)$ having the following property: all edges have at least one of their incident vertices in $H(t)$, i.e., all stubs in $V \setminus H(t)$ are connected to some stub in $H(t)$. By definition the ith node, for $i \in \{1, \ldots, l\}$, of any graph $G' \in \mathcal{G}(d_1, \ldots, d_l)$ has degree d_i.

Let $G(t) = (V, E(t))$ be the graph defined over the set of nodes $V = H(t) \cup (\mathcal{I}(t) \cup_{i=1}^{t} \mathcal{R}(i))$ for a fixed round t. We define $E(t) = \{\{v, w\} \mid \{v, w\} \in E \text{ and } v \in H(t), w \notin H(t) \text{ or } v, w \in H(t)\}$. Note that $G(t)$ can be regarded as a random variable. For a *fixed* $S \subset V$ we define G_S as the subgraph of G induced by the nodes of S, i.e., $G_S = (S, E \cap (S \times S))$. Accordingly, we define $G_{\bar{S}} = (V, E_{\bar{S}})$ as a subgraph that does not have any edges connecting two nodes in S, i.e., $E_{\bar{S}} = E \setminus (S \times S)$, for a fixed S. Note that G_S is fixed, whereas $G_{\bar{S}}$ is a random variable. Finally, we define $G \setminus G(t)$ as the graph $(V, E \setminus E(t))$.

Lemma 5.8. *Let*

$$A(t, S, G_S) = ((H(t) = \bar{S}) \wedge ((G \setminus G_{\bar{S}}) = G_S)).$$

Then it holds that

$$\Pr\left[G(t) = G(d_1, d_2, \ldots, d_{|V|}) \mid A(t, S, G_S)\right] = \frac{1}{|\mathcal{G}(d_1, \ldots, d_{|V|})|}$$

for any fixed graph $G'(d_1, d_2, \ldots, d_{|V|}) \in \mathcal{G}(d_1, \ldots, d_{|V|})$ *whenever*

- $\Pr[A(t, S, G_S)] \neq 0$, *and*
- $A(t, S, G_S)$ *results in a degree sequence* $(d_1, \ldots, d_{|V|})$ *for* $G(t)$.

Proof. We assume w.l.o.g. that the vertices $v_1, \ldots, v_n \in V$ are ordered in such a way that $v_1, \ldots, v_{|H(t)|} \in \bar{S}$. For $1 \leq i, r \leq n$ we call an edge $\{v_i, v_r\} \in E$ the jth edge of v_i if there are exactly $j-1$ edges $\{v_i, v_k\} \in E$ with $k < r$. For $1 \leq i \leq n$ we define the event

$$B(v_i, j, \ell) = \{\text{node } v_i \text{ chooses its } j\text{th neighbor in round } \ell \leq t\}.$$

We define $U \subset V \times \{1, \ldots, n\} \times \{1, \ldots, t\}$ such that for any $v_i \in V$ and some round $\ell \leq t$ it holds that $|U \cap \{(v_i, j, \ell) \mid 1 \leq j \leq \deg_{v_i}\}| = \gamma(\mathcal{V})$ and $|U \cap \{(v_i, j, \ell') \mid 1 \leq j \leq \deg_{v_i}\}| = 0$ for all rounds $\ell' \neq \ell$. That is, every node communicates with $\gamma(\mathcal{V})$ many neighbors in one round and one round only. Clearly, if a node is not infected or if it belongs to some $\mathcal{R}(j)$, then no infection is transmitted to these $\gamma(\mathcal{V})$ many neighbors. Now we define the event

$$B(t) = \bigwedge_{(v_i, j, \ell) \in U} B(v_i, j, \ell).$$

Our goal is to show that

$$\Pr[G(t) = G' \mid (H(t) = \bar{S}) \wedge ((G \setminus G_{\bar{S}}) = G_S) \wedge B(t)]$$

is the same for any fixed G', as long as

$$\Pr[H(t) = \bar{S} \wedge (G \setminus G_{\bar{S}}) = G_S \wedge B(t)] \neq 0.$$

We know that

$$\begin{aligned} P_t &= \Pr[G(t) = G' \mid (H(t) = \bar{S}) \wedge ((G \setminus G_{\bar{S}}) = G_S) \wedge B(t)] \\ &= \frac{\Pr[(G(t) = G') \wedge (H(t) = \bar{S}) \wedge ((G \setminus G_{\bar{S}}) = G_S) \wedge B(t)]}{\Pr[(H(t) = \bar{S}) \wedge ((G \setminus G_{\bar{S}}) = G_S) \wedge B(t)]}. \end{aligned}$$

Suppose that G' and G'' are two arbitrary graphs with $G', G'' \in \mathcal{G}(d_1, \ldots, d_{|V|})$. Now we show that if

$$((G \setminus G_{\bar{S}}) = G_S) \wedge (G_{\bar{S}} = G') \wedge B(t)$$

results in $H(t) = \bar{S}$, then the same holds if we replace G' by G''.

We prove by induction on i that $H(i)$ is the same in both $G' \cup G_S$ and $G'' \cup G_S$ for any $i \leq t$. For $i = 0$ the assumption is trivially fulfilled. Now assume that the claim holds for $i-1$. If a node v in the graph $G = G' \cup G_S$ becomes informed in round i, then there must be some event $B(u, j, i)$ such that the jth edge of $u \in \mathcal{I}(i-1)$ is adjacent to v. The corresponding event implies that $v \in G'' \cup G_S$ becomes informed as well, since this edge (jth edge of u) is contained in G_S. Furthermore, for the same reasons, v will belong to the same set, $\mathcal{I}(i)$ or $\mathcal{R}(i)$, irrespective of G' or G''.

On the other hand, if a node v of $G' \cup G_S$ is in $H(i)$, then for all events $B(u, j, i)$, for which the jth edge of u is adjacent to v, it holds that $u \in H(i-1) \cup_{j=1}^{i-1} \mathcal{R}(j)$. This finishes our induction.

Now assume that $((G \setminus G_{\bar{S}}) = G_S) \wedge (G_{\bar{S}} = G') \wedge B(t)$ leads to $H(t) = \bar{S}$. Then, according to our inductive proof above, $((G \setminus G_{\bar{S}}) = G_S) \wedge B(t)$ leads to $H(t) = \bar{S}$ as well, irrespective of the structure of $G_{\bar{S}}$. It follows that

$$\begin{aligned}
P_t &= \frac{\Pr[(G(t) = G') \wedge (H(t) = \bar{S}) \wedge ((G \setminus G_{\bar{S}}) = G_S) \wedge B(t)]}{\Pr[(H(t) = \bar{S}) \wedge ((G \setminus G_{\bar{S}}) = G_S) \wedge B(t)]} \\
&= \frac{\Pr[(G_{\bar{S}} = G') \wedge ((G \setminus G_{\bar{S}}) = G_S)] \cdot \Pr[B(t)]}{\Pr[((G \setminus G_{\bar{S}}) = G_S)] \cdot \Pr[B(t)]} \\
&= \frac{\Pr[(G_{\bar{S}} = G') \wedge ((G \setminus G_{\bar{S}}) = G_S)]}{\Pr[((G \setminus G_{\bar{S}}) = G_S)]} \\
&= \Pr[(G_{\bar{S}} = G') \mid ((G \setminus G_{\bar{S}}) = G_S)] \\
&= \frac{1}{|\mathcal{G}(d_1, \ldots, d_{|V|})|}.
\end{aligned}$$

The second equality is obtained by the two following facts: 1.) $((G \setminus G_{\bar{S}}) = G_S) \wedge B(t)$ implies that $H(t) = \bar{S}$, and 2.) the event $B(t)$ is independent of the structure of $G \setminus G_{\bar{S}}$ or $G_{\bar{S}}$. The fifth equality holds, since, by definition, the distribution of the edges among the nodes of $G \setminus G_{\bar{S}}$ is independent of the distribution of the edges in $G_{\bar{S}}$ given the degree sequence $(d_1, \ldots, d_{|V|})$ for $G_{\bar{S}}$. Note that the degree sequence $(d_1, \ldots, d_{|V|})$ is obtained from the fixed graph $(G \setminus G_{\bar{S}}) = G_S$. This completes the proof. □

Next, we show that we have at least $\delta|S||\bar{S}|/n \cdot (1-\epsilon)$ many edges between S and $\bar{S}$ with probability at least $1 - n^{-2}$, where $\epsilon > 0$ is some small constant, for any set $S \subset V$ with $|S| = n/\delta^{1/4}$. This lemma is similar to (but stronger than) Lemma 5 of [42]. A similar result has been derived for random d-regular graphs in [26].

Lemma 5.9. *Let X and Y be two disjoint subsets of vertices in a random n-vertex power law graph $G' = (V', E')$ with minimum degree δ and $|X| \leq |Y|$, where $|X|$ and $|Y|$ refer to the size of X and Y, respectively. Further, let $E(X, Y)$ be the set of edges between the nodes in X and Y. Then for each pair of a disjoint subset of vertices X and Y of sizes $|X|$ and $|Y|$ it holds that*

$$\left| E(X,Y) - \frac{d_X|X|d_Y|Y|}{d_V n} \right| \leq \sqrt{d_X|X||Y|} \cdot \sqrt{32\log(ne/|Y|)} \tag{5.15}$$

with probability at least $1 - 2e^{-|Y|}$*, where* $d_R|R|$ *denotes the number of stubs in subset* R *(*d_R *is the average degree in* R*).*

Proof. Let X, Y be two fixed disjoint vertex sets with $|X| \leq |Y|$, and let $Z(S)$ be the number of edges between X and Y in a configuration S. We view S as a sequence of numbers $(t_1, t_2, \dots\ t_q)$, where q is the total number of edges. Each number t_i defines the ith pair of stubs (i.e., the ith edge) in S and is chosen uniformly at random between 1 and $2q - (2i - 1)$. That is, it holds that $1 \leq t_i \leq 2q - (2i - 1)$ for each t_i. Further, let the stubs in G be ordered in such a way that the first $d_X|X|$ stubs correspond to the vertices in X.

In the following we describe the pairing process used to construct the random graph. Let L_i denote the *sequence* of the remaining unpaired stubs after the first $(i-1)$ pairs have been selected. Then the first entry in L_i is chosen as the first stub of the ith edge, whereas the second one is given by the one at the position $1 + t_i$ in L_i. Recall that the value of t_i is randomly chosen and corresponds to the range between 1 and the amount of unpaired stubs in L_i after the first one of the ith edge was selected. Note that we ensure that $Z(S)$ is determined by $(t_1, t_2, \dots, t_{d_X|X|})$ by fixing the points corresponding to the vertices in X first.

For $i = 0, 1, \dots, q$ and a configuration $S = (t_1, t_2, \dots\ t_q)$, let

$$Z_i(S) \ \equiv \ Z_i(t_1, t_2, \dots, t_q) \ = \ \mathbb{E}_{\tau_{i+1},\dots,\tau_q}(Z(t_1, \dots, t_i, \tau_{i+1}, \dots, \tau_q)) \ \equiv \ Z(t_1, t_2, \dots, t_i).$$

That is, $Z_i(S)$ is the expected number of edges between the stubs of X and Y in a random configuration that agrees with the configuration S on the first i pairs.

We have $Z_0(S) = \mathbb{E}(Z(S)) = d_X|X|d_Y|Y|/(d_V n)$ and $Z_{d_X|X|}(S) = Z(S)$. The sequence of random variables Z_i, where $Z_i \equiv Z(t_1, ..., t_i)$ for any $i = 0, 1, \dots, q$, is a martingale because $\mathbb{E}(Z_{i+1}|Z_i) = Z_i$:

$$\begin{aligned}
\mathbb{E}(Z_{i+1}|Z_i = z) &= \mathbb{E}(Z(t_1, \dots, t_{i+1})|Z(t_1, \dots, t_i) = z) \\
&= \sum_{t_1,\dots,t_{i+1}:Z(t_1,\dots,t_i)=z} \frac{\Pr(t_1, \dots, t_{i+1})}{\Pr(Z_i = z)}\ Z(t_1, \dots, t_{i+1}) \\
&= \frac{1}{\Pr(Z_i = z)} \sum_{t_1,\dots,t_i:Z(t_1,\dots,t_i)=z} \Pr(t_1, \dots, t_i) \sum_{t_{i+1}} \Pr(t_{i+1}) Z(t_1, \dots, t_{i+1}) \\
&= \frac{1}{\Pr(Z_i = z)} \sum_{t_1,\dots,t_i:Z(t_1,\dots,t_i)=z} \Pr(t_1, \dots, t_i)\ Z(t_1, \dots, t_i) \ = \ z.
\end{aligned}$$

Let $F_i^X(S)$ and $F_i^Y(S)$ denote the number of free stubs in X and Y, respectively, after the first i pairs in S have been matched. If $F_i^X(S) = 0$, then $Z_j(S) = Z(S)$ for all $j \geq i$. If $F_i^X(S) \geq 1$, then, dropping S from the notation for simplicity,

$$Z_i \ = \ (Y - F_i^Y) + \frac{F_i^X F_i^Y}{nd_V - (2i+1)},$$

where $Y - F_i^Y$ is the number of edges between the stubs of X and Y given by the first i pairs in S, and the last term is the expected number of edges between the stubs of X and Y contributed by a random pairing of the remaining stubs. When the next $(i+1)$-th pair is matched, then (F_{i+1}^X, F_{i+1}^Y) is either $(F_i^X - 1, F_i^Y - 1)$, or $(F_i^X - 2, F_i^Y)$, or $(F_i^X - 1, F_i^Y)$, and it can be checked that in all three cases

$$|Z_{i+1} - Z_i| \;\leq\; 2. \tag{5.16}$$

Applying the Azuma-Hoeffding inequality to our martingale $(Z_0, Z_1, \ldots)$, we obtain that

$$\begin{aligned}\Pr\left(\left|E(X,Y) - \frac{d_X|X|d_Y|Y|}{d_V n}\right| \geq \rho\right) &= \Pr(|Z_{d_X|X|}(S) - Z_0(S)| \geq \rho) \\ &\leq 2\exp\left(-\frac{\rho^2}{8d_X|X|}\right).\end{aligned}$$

The number of pairs of disjoint sets of sizes $|X| \leq |Y|$ is at most

$$\binom{n}{|X|}\binom{n-|X|}{|Y|} \;\leq\; \left(\frac{ne}{|X|}\right)^{|X|}\left(\frac{ne}{|Y|}\right)^{|Y|} \;\leq\; \left(\frac{ne}{|Y|}\right)^{2|Y|}.$$

The last inequality holds because $(ne/z)^z$ is monotonically increasing for $0 \leq z \leq n$. Let A be the event that *there are disjoint vertex sets $X, Y \subseteq V$ of sizes $|X|$ and $|Y|$ such that $|E(X,Y) - d_X|X|d_Y|Y|/(d_V n)| \geq \rho$*. Then we obtain that

$$\Pr(A) \leq 2\exp\left(-\frac{\rho^2}{8d_X|X|} + 2|Y|\log(ne/|Y|)\right)$$

holds for given set sizes $|X| \leq |Y|$ and a random n-vertex power law graph $G' = (V', E')$ by using the Union Bound. The above bound is at most $2e^{-|Y|}$ if

$$\frac{\rho^2}{8d_X|X|} - 2|Y|\log(ne/|Y|) \;\geq\; |Y|,$$

which is fulfilled for

$$\rho \geq \sqrt{d_X|X||Y|}\cdot\sqrt{32\log(ne/|Y|)}.$$

This completes the proof. □

Next, we show that there is a good chance for the edges between $\mathcal{I}(t)$ and $H(t)$ to be "evenly" distributed regardless of the sequence of random choices $B(t)$ (defined similarly as in the proof of Lemma 5.8).

Lemma 5.10. *Let G' be some graph chosen uniformly at random from the space of all power law graphs with the given degree distribution, and let $B(t)$ be the sequence of random choices of the nodes in G'. That is, for $1 \leq i \leq t$ we define the event*

$$B(v_i, j, l) = \{v_i \text{ chooses its } j\text{th neighbor in round } l \leq t\}, \text{ and}$$

$$B(t) = \wedge_{(v_i,j,l)\in U} B(v_i, j, l).$$

Further, we assume that there is some $U \subset V \times \{1, \ldots, n\} \times \{1, \ldots, t\}$ *such that for any* $v_i \in V$ *and some round* $\ell \leq t$ *it holds that* $|\{(v_i, j, \ell) \in U \mid 1 \leq j \leq \deg_{v_i}\}| = \gamma(\mathcal{V})$ *and* $|\{(v_i, j, \ell') \in U \mid 1 \leq j \leq \deg_{v_i}\}| = 0$ *for all rounds* $\ell' \neq \ell$.

Let $X, Y \in V$ *be two distinct subsets of nodes and let* $E(X, Y)$ *be the set of edges between the nodes in* X *and* Y. *Then it follows that*

$$\Pr[(|E(H(t), \mathcal{I}(t))| \geq \delta|\mathcal{I}(t)|/2 \text{ or } |E(H(t), \mathcal{I}(t))| = 0) \mid B(t)] \geq 1 - n^{-2}, \text{ and}$$

$$\Pr[|\Gamma(\mathcal{I}(t)) \cap H(t)| = |H(t)|(1 - o(1)) \mid |E(H(t), \mathcal{I}(t))| \geq \delta|\mathcal{I}(t)|/2] \geq 1 - n^{-2}.$$

Proof. We start with our first claim. Let G' be some graph chosen uniformly at random from the space of all power law graphs with the given degree distribution. According to Lemma 5.9, no subset with the assumptions of Lemma 5.9 violates the result of that lemma with probability at least $1 - 2e^{-n/\delta^{1/4}}$. That is, for all these graphs $B(t)$ results in a subset $H(t)$ with $|E(\mathcal{I}(t), H(t))| = 0$ or

$$|E(\mathcal{I}(t), H(t))| \geq \frac{d_{\mathcal{I}(t)}|\mathcal{I}(t)|d_{H(t)}|H(t)|}{d_V n}(1 - o(1)),$$

where $d_X|X|$ denotes the number of stubs in subset X and d_X represents the average degree in X.

According to Lemma 5.8, we know that the edges are randomly distributed among the corresponding stubs of $\mathcal{I}(t)$ and $H(t)$. Since G' is a power law graph with exponent at least 2 and $|H(t)| = n(1-o(1))$, it follows that $d_V = \Theta(\delta)$ and $2/3 \cdot d_V < d_{H(t)} < d_V \cdot 3/2$. Then $|E(\mathcal{I}(t), H(t))| \geq \delta|\mathcal{I}(t)|/2$ or $|E(H(t), \mathcal{I}(t))| = 0$, with probability $1 - 2e^{-n/\delta^{1/4}}$.

Now we consider the case when $|E(\mathcal{I}(t), H(t))| \geq \delta|\mathcal{I}(t)|/2$. We know that these edges between $\mathcal{I}(t)$ and $H(t)$ are randomly distributed. In the following we compute the number of nodes in $H(t)$ that have at least one neighbor in $\mathcal{I}(t)$. Since each node in $H(t)$ is incident to at least $\delta = \Omega(\log \log n)$ many edges, a node v is connected to an edge between $\mathcal{I}(t)$ and $H(t)$ with probability at least

$$1 - \left(1 - \frac{\delta|\mathcal{I}(t)|}{2d_V n}\right)^{\delta} = 1 - o(1).$$

Thus, the expected number of nodes connected to edges between $\mathcal{I}(t)$ and $H(t)$ is $H(t)(1 - o(1))$.

Now we have to show that the number of nodes in $H(t)$ connected to edges between $\mathcal{I}(t)$ and $H(t)$ is $H(t)(1 - o(1))$, with high probability. Consider for each node in $H(t)$ exactly δ stubs. We define a Martingale sequence over these stubs, where the δ stubs corresponding to a vertex are joined together to a collection of these δ stubs. Now let Z define the random variable which describes the number of collections of stubs having

at least one stub connected to a stub of $\mathcal{I}(t)$. $Z_0 = \mathbb{E}(Z)$, and Z_i is the expectation of Z when i collections of stubs have been revealed. Clearly, $(Z_i)_{i\geq 0}$ is a Martingale sequence, where $|Z_{i+1} - Z_i| \leq \delta$. We obtain our second claim of the lemma by applying the Azuma-Hoeffding inequality. □

The combination of the above statements implies the theorem. □

5.2.2.2 Model 2 - $\mathcal{V}$ competing with Λ

In the second model $\mathcal{V}$ competes with Λ instead of Φ. Recall that Λ is activated on each infected node v after a specific delay d, provided that $c_1 \log\log n$ many rounds have passed. Then each such node establishes a notification tree (cf. Section 5.2.1). Note, however, that we analyze a more restrictive form of Λ here. Let v be a sending node within a notification tree $\Upsilon_{v'}$ in round i that was notified by a node $w \in \Upsilon_{v',i-1}$. Then, as a part of $\Upsilon_{v',i}$, v is allowed to notify all nodes of $\Gamma(v) \setminus \{\mathcal{I}(i) \cup \{w\}\}$ in round i. In other words, each node forwards a preventive message only to the neighbors it did not obtain said message from in the previous round. Obviously, the dissemination of Λ may be much faster than the one of Φ. On the other hand, the minimum degree δ is a constant in this model. Since the degrees of the nodes in G are chosen according to a power law, it follows that only $o(n)$ many nodes will have a degree of $\Omega(\log\log n)$, which was the minimum degree in Model 1. The question is whether or not Λ is able to achieve an embankment of $\mathcal{V}$ regardless of these circumstances.

The main statement of this section is the following:

Theorem 5.4. *Let G be a random power law graph as defined in Section 5.2.1 with minimum degree $\delta \in \Theta(1)$. Further, let Model 2 be in place. Then there is a choice for $c_1, \beta \in \Theta(1)$, such that for any $d \in \Theta(1)$ and $3 \leq \gamma(\mathcal{V}) < \delta$, process $\mathcal{V}$ does not infect more than $o(n)$ many nodes in total with probability $1 - o(1)$.*

Our general approach We prove Theorem 5.4 in several phases. In the first phase, i.e., in the first $c_1 \log\log n$ rounds where $|\mathcal{V}| \in \mathcal{O}(\log^c n)$, for some positive constant c, we show that $\mathcal{V}$ grows by an exponential rate. Note that this follows directly from our model. Hence, $\mathcal{V}$ has infected at most a polylogarithmic amount of nodes at the end of Phase 1. In the second phase then we use the expected distances of the nodes in G to argue that an additional amount of $\mathcal{O}((\log\log n)^2)$ many rounds is sufficient for Λ to finally immunize almost the entire network.

By definition only nodes in Λ, i.e., immunized nodes, or in $\mathcal{V}$, i.e., infective ones, may initiate transmissions aiming to immunize and infect, respectively. Further, recall that we use the Principle of Deferred Decision and the Configuration Model in the upcoming analysis. Consequently, pairings occur either in Λ or $\mathcal{V}$, or between nodes in $\Lambda/\mathcal{V}$ and $\mathcal{U}$. Clearly, the latter ones are the most interesting to us. Note that we omit pairings

between nodes in $\mathcal{V}$ and Λ at this point and refer to them when needed. The following definitions simplify the presentation.

Definition 5.1 (ϖ_i)**.** *Let $i > 0$ be an arbitrary round and let $\xi(v)$ be the subset of connections of an arbitrary node $v \in V$ that process $\mathcal{V}$ uses to spread over in round i. Hence, $\xi(v)$ is a randomly chosen subset of $s(v)$ of size $\gamma(\mathcal{V})$. Further, let $\xi_f(v) := \{s \in \xi(v) \mid s \in s_f(v)\}$ be the subset of free stubs among the ones of $\xi(v)$. Then*

$$\varpi_i := \sum_{v \in \Lambda_i} |s_f(v)| + \sum_{v \in \mathcal{V}_i} |\xi_f(v)|$$

is the amount of free stubs of nodes in $\Lambda_i \cup \mathcal{V}_i$. We say that ϖ_i is constrained on Λ_i and $\mathcal{V}_i$ whenever we consider ϖ_i to be equal to $\sum_{v \in \Lambda_i} |s_f(v)|$ and $\sum_{v \in \mathcal{V}_i} |\xi_f(v)|$, respectively.

Definition 5.2 (Type A notification tree)**.** *Let q be some proper constant and let Υ_w be a notification tree with root $w \in V$. Then Υ_w is of Type A if and only if $\deg_w \leq \log^{q/\sqrt{\beta}} n$. In this case the maximum path length of each Type A notification tree from its root to its leaves is limited by $(q \cdot \log\log n)^2$.*

Definition 5.3 (Type B notification tree)**.** *Each notification tree not of Type A is defined as a notification tree of Type B.*

Definition 5.4 (Λ_i^-)**.** *Let i' be defined as in Section 5.2.1.*

Then the set

$$\Lambda_i^- := \left\{ v \in \Lambda_i \mid v \in \bigcup_{\forall w \in \Lambda_{i'} : \deg_w \leq \log^{q/\sqrt{\beta}} n} \Upsilon_{w,i} \right\},$$

with $q \geq \sqrt{\beta}$ being some proper constant and $i \geq i'$, contains all newly immunized nodes in round i of each Type A notification tree created in round i'. In other words, each node $v \in \Lambda_i^-$ is contained in the $(i - i')$-th level of at least one notification tree Υ_w created in round i', in which the root node w has degree at most $\log^{q/\sqrt{\beta}} n$ (cf. Figure 5.5).

Definition 5.5 (Λ_i^+)**.** *Let i' be defined as in Section 5.2.1.*

Then the set

$$\Lambda_i^+ := \left\{ v \in \Lambda_i \mid v \in \bigcup_{\forall w \in \Lambda_{i'} : \deg_w > \log^{q/\sqrt{\beta}} n} \Upsilon_{w,i} \right\},$$

with $q \geq \sqrt{\beta}$ being some proper constant and $i \geq i'$, contains all newly immunized nodes in round i of each Type B notification tree created in round i'. In other words, each node $v \in \Lambda_i^+$ is contained in the $(i - i')$-th level of at least one notification tree Υ_w created in round i', in which the root node w has degree greater than $\log^{q/\sqrt{\beta}} n$ (cf. Figure 5.5).

Definition 5.6 (Blocked node)**.** *Let i be the round in which some node $v \in \Lambda$ transmits a preventive message. By definition, v notifies its entire non-infected neighborhood in round i. We say that v is a* blocked node *in round i if $|\Gamma(v) \cap \mathcal{I}(i)| > 0$, i.e., if v has at least one infected neighbor in the round v transmits a preventive message.*

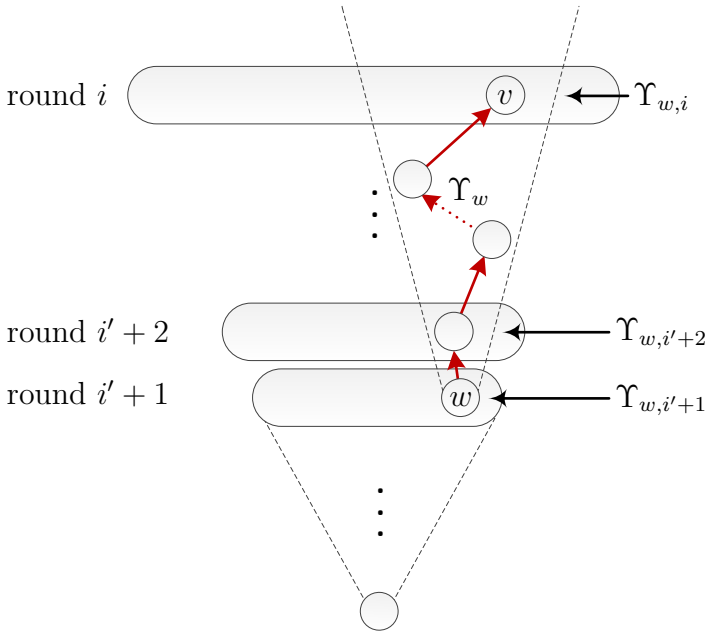


Figure 5.5: *This figure shows the dissemination of Λ up to a node $v \in \Upsilon_{w,i}$ via a notification tree Υ_w with root node w. The corresponding level of Υ_w in round j is denoted by $\Upsilon_{w,j}$. Υ_w is of Type A if $\deg_w \leq \log^{q/\sqrt{\beta}} n$ and of Type B otherwise.*

Finally, we are ready to prove Theorem 5.4.

Proof of Theorem 5.4.

To prove our claim we analyze both processes $\mathcal{V}$ and Λ throughout $\mathcal{O}((\log\log n)^2)$ many rounds. We prove the theorem in two phases.

In the first phase we show that $|\mathcal{V}_{i+1}| \geq |\mathcal{V}_i| \cdot (1+c)$, for some constant $c > 0$, during the first $c_1 \log\log n$ rounds. Note that it holds $\gamma(\mathcal{V}) \cdot |\mathcal{V}_i| \geq |\mathcal{V}_{i+1}|$ due to our model (cf. Section 5.2.1). This implies that process $\mathcal{V}$ grows by at least and at most an exponential rate and cannot infect more than a polylogarithmic amount of nodes overall within the first $c_1 \log\log n$ rounds.

The main idea of Phase 2 is as follows. In every round i of this phase process $\mathcal{V}$ may induce a subgraph/forest $G^{\mathcal{V}}$ in G of size at most $\gamma(\mathcal{V})^{\mathcal{O}((\log\log n)^2)}$. Note that the shape of $G^{\mathcal{V}}$ may change over time but its size will still be limited. On the other hand, process Λ induces a subgraph G^{Λ} in G on its own. Note that G^{Λ} forms a connected component, since every immunized node stays immunized (cf. Figure 5.6). By definition, only infected nodes may *block* the dissemination of preventive messages (cf. Section 5.2.1). However, we argue that the transmission of a preventive message can only be *delayed*

for at most $q \cdot \log \log n$ many rounds, where $q \in \Theta(1)$ is computed below. This is due to the following observations:

1. each infected node v recovers just in time, such that v does not *block* any preventive message sent along the path v was originally infected on by $\mathcal{V}$ (cf. Figure 5.7), and

2. each *blocked node* in Λ is part of a Type B notification tree at most $q \cdot \log \log n$ many rounds after it was blocked for the first time.

Further, G is a random power law graph with power law exponent $2 < \alpha < 3$. Using the above arguments we show that $\mathcal{O}((\log \log n)^2)$ many rounds suffice for Λ to immunize at least $n(1 - o(1))$ many nodes with probability $1 - o(1)$.

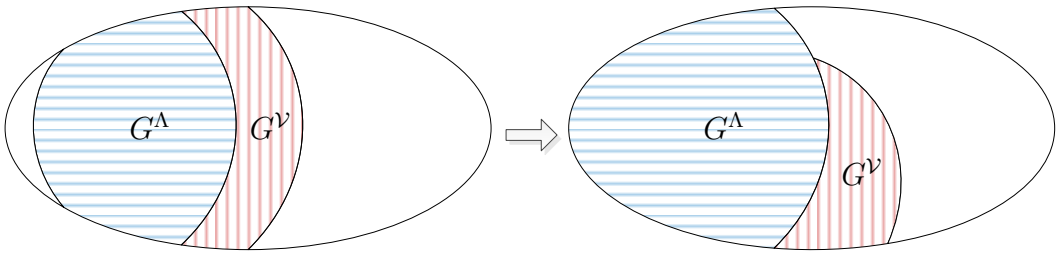


Figure 5.6: *The above figures depict a generic course for the expansion of G^{Λ} and $G^{\mathcal{V}}$, respectively. Note that the size of $G^{\mathcal{V}}$ is limited by $\gamma(\mathcal{V})^{\mathcal{O}((\log \log n)^2)} = \log^{\mathcal{O}(\log \log n)} n$ if the lifespan of $\mathcal{V}$ is limited by $\mathcal{O}((\log \log n)^2)$.*

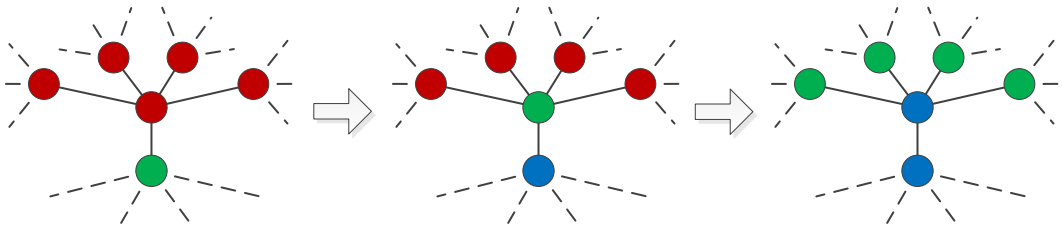


Figure 5.7: *This is an example of* Λ chasing $\mathcal{V}$. *We distinguish between three different types of nodes: 1.) red nodes, i.e., infected nodes, 2.) green nodes, i.e., recovered nodes that have not generated and sent any preventive message yet, and 3.) blue nodes, i.e., immunized nodes that have already generated and sent at least one preventive message. Clearly, since each (transmitting) blue node v notifies its entire non-infected neighborhood, all green nodes w incident to v are notified as well. Consequently, these nodes act as relays to spread the preventive message sent by v. Further, each red node v' turns into a green one just in time, such that none of the preventive messages generated and sent by the green nodes incident to v' are blocked. This implies that process Λ moves along the same path as $\mathcal{V}$ did.*

Phase 1 Let $i' := c_1 \log\log n$ be defined as in Section 5.2.1. Since $\gamma(\mathcal{V})$ is constant, it follows that $\mathcal{V}$ cannot infect more than $\log^{\mathcal{O}(1)} n$ many nodes within the first i' rounds. Further, the lowest and the average degree in the considered graph is constant. Note that it holds $\Lambda_i = \emptyset$ for all rounds $i < i'$ according to our model.

Additionally, note that $\mathcal{V}_0, ..., \mathcal{V}_{i'}$ is the sequence of infected nodes disseminating $\mathcal{V}$ throughout Phase 1. Clearly, $|\mathcal{V}_i| \leq \gamma(\mathcal{V})^i$ holds for each round $i \leq i'$. Consequently, $|s_f(\cup_{i \leq i'} \mathcal{V}_i)| \leq \mathcal{O}(\gamma(\mathcal{V})^{i'} \cdot n^{1/\alpha}) = o(n)$ and $|s(\mathcal{U}(i'))| = \Theta(n)$.

Let $i \leq i'$ be an arbitrary round. Further, let $v_1^i, ..., v_k^i \in \mathcal{U}(i)$ be a sequence of the uninformed nodes in round i sorted in descending order by the degree. Note that $k < n$. Let the indicator variable x_j^i be equal to 1 if v_j^i is chosen at least twice by the nodes in $\mathcal{V}_i$ and 0 otherwise. Further, let $X = \sum_{1 \leq j \leq k} x_j^i$. Then it follows that

$$\Pr\left(x_1^i\right) \leq \binom{\gamma(\mathcal{V}) \cdot |\mathcal{V}_i|}{2} \left(\frac{\deg_{v_1^i}}{\Theta(n)}\right)^2 \leq \Theta\left(\left(\frac{\log^{\log(\gamma(\mathcal{V}))\cdot c_1} n}{n^{1-1/\alpha}}\right)^2\right) = n^{-(1+\Omega(1))},$$

since $\deg_v \leq n^{1/\alpha}$ for all $v \in V$. Note that these events are negatively correlated, since each connection established to some v_j^i diminishes the amount of connections available for all yet unconsidered nodes of the sequence. Therefore, $\Pr(x_1^i) \geq \Pr(x_2^i) \geq ... \geq \Pr(x_k^i)$. Consequently,

$$\mathbb{E}[X] \leq n \cdot \Theta\left(\left(\frac{\log^{\log(\gamma(\mathcal{V}))\cdot c_1} n}{n^{1-1/\alpha}}\right)^2\right) = \Theta\left(\frac{\log^{2\cdot\log(\gamma(\mathcal{V}))\cdot c_1} n}{n^{1-2/\alpha}}\right) = n^{-\Omega(1)},$$

since $2 < \alpha < 3$. Utilizing standard Chernoff bounds (cf. Lemma 4.1), we conclude that none of the nodes $v_1^i, ..., v_k^i$ is connected to more than one node in $\mathcal{V}_i$ with probability $1 - n^{-\Omega(1)}$. In other words, each newly infected node $v \in \mathcal{V}_{i+1}$ does only provide one connection to some infected node $w \in \mathcal{V}_i$ with high probability.

Recall that $|\cup_{i \leq i'} \mathcal{V}_i| = \mathcal{O}(\gamma(\mathcal{V})^{i'}) = \mathcal{O}(\log^{\log(\gamma(\mathcal{V}))\cdot c_1} n) < \log^{\log(\gamma(\mathcal{V}))\cdot c_1 + 1} n$ and $|s_f(V)| = \Theta(n)$ in round i'. Further, note that $|s_f(\cup_{i \leq i'} \mathcal{V}_i)|$ is naturally upper bounded by $c \cdot \log^{\log(\gamma(\mathcal{V}))\cdot c_1} n \cdot n^{1/\alpha}$, for some positive constant c. Let $v_1^{i'}, ..., v_k^{i'} \in (\cup_{i \leq i'} \mathcal{V}_i)$ be an ordered sequence of nodes disseminating $\mathcal{V}$ up to (and including) round i'. Further, let $x_j^{i'}$ be the event that $v_j^{i'}$ chooses at least $\gamma(\mathcal{V}) - 1$ many uninformed nodes among its $\gamma(\mathcal{V})$ many tries. Then

$$\Pr\left(x_k^{i'}\right) \geq \left(1 - \overbrace{\Theta\left(\frac{\log^{\log(\gamma(\mathcal{V}))\cdot c_1} n \cdot n^{1/\alpha}}{n}\right)}^{\text{Probability to choose a node in } \mathcal{I}}\right)^{\gamma(\mathcal{V})}$$

and consequently

$$\Pr\left(x_1^{i'} \cap ... \cap x_k^{i'}\right) \geq \left(1 - \Theta\left(\frac{\log^{\log(\gamma(\mathcal{V}))\cdot c_1} n \cdot n^{1/\alpha}}{n}\right)\right)^{\gamma(\mathcal{V})\cdot|\cup_{i\leq i'}\mathcal{V}_i|}$$
$$\geq 1 - \Theta\left(\frac{\log^{2\cdot(\log(\gamma(\mathcal{V}))\cdot c_1+1)} n}{n^{1-1/\alpha}}\right),$$

since $\Pr\left(x_1^{i'}\right) \geq \Pr\left(x_2^{i'}\right) \geq ... \geq \Pr\left(x_k^{i'}\right)$. Combining the above arguments we conclude that $\mathcal{V}$ remains active until round i' and it holds that

$$\gamma(\mathcal{V}) \cdot |\mathcal{V}_i| \geq |\mathcal{V}_{i+1}| \geq |\mathcal{V}_i| \cdot (1 + c),$$

for some positive constant $c \geq 1$ and all rounds $i \leq i'$, with high probability.

Now we show that we obtain $\Lambda_{i'}^+ \neq \emptyset$ in round i', i.e., that $\mathcal{V}$ has infected at least one node with degree large enough, such that the nodes in the corresponding notification tree are allowed to send at least $c'(\log\log n)^2$ many rounds for some sufficiently large constant c'. Recall that the first nodes contained in $\Lambda_{i'}$ are the newly infected ones at the beginning of round $i' - d$. However, by definition, the first round in which the nodes of Λ are allowed to send is $i' + 1$. Consequently, Λ does not notify any nodes during Phase 1 whatsoever. From the arguments above we already know that $\gamma(\mathcal{V}) \cdot |\mathcal{V}_{i-1}| \geq |\mathcal{V}_i| \geq q' \cdot |\mathcal{V}_{i-1}|$, for all $i \leq i'$ and some constant $q' \geq 2$. Thus, by definition, we obtain that

$$\log^{\log(\gamma(\mathcal{V}))\cdot c_1} n \geq |\mathcal{V}_{i'}| \geq \log^{c_1} n$$

and

$$\frac{\log^{\log(\gamma(\mathcal{V}))\cdot c_1} n}{2^d} \geq |\Lambda_{i'}| = |\mathcal{V}_{i'-d}| \geq \frac{\log^{c_1} n}{\gamma(\mathcal{V})^d}.$$

At this point we have estimated the amount of nodes contained in $\Lambda_{i'}$ and $\mathcal{V}_{i'}$, respectively. It remains to show that $\Lambda_{i'}^+ \neq \emptyset$. In the following we show that there are at least $\Omega(\log n)$ many nodes in $\Lambda_{i'}$ with degree $\log^{c_1-1/\alpha-1} n$.

Recall that $|(\Lambda_{i'} \cup (\cup_{i\leq i'}\mathcal{V}_i))| = \mathcal{O}(\gamma(\mathcal{V})^{i'}) = \log^{\mathcal{O}(1)} n$ holds by definition. Therefore, almost all connections chosen by the nodes in $\mathcal{V}_{i'-d-1}$ lead to nodes in $\mathcal{U}$ (see above), since the amount of nodes in $\Lambda \cup \mathcal{V}$ is sufficiently small. Since $|\mathcal{V}_{i'-d-1}| \geq \frac{\log^{c_1} n}{\gamma(\mathcal{V})^{d+1}}$ and $\gamma(\mathcal{V}) \cdot |\mathcal{V}_{i'-d-1}| \geq \varpi_{i'-d-1} \geq |\mathcal{V}_{i'-d-1}|$, we obtain that

$$\varpi_{i'-d-1} = \Omega(\log^{c_1} n).$$

Let $Y_{\deg_v}^{\varpi_{i'-d-1}}$ be the event that a node $v \in \mathcal{U}(i'-d-1)$ with degree $\deg_v$ is paired with at least one of the $\varpi_{i'-d-1}$ many stubs. Then

$$\begin{aligned}\Pr\left(Y_{\deg_v}^{\varpi_{i'-d-1}}\right) &\geq 1 - \left(1 - \frac{\deg_v}{\Theta(n)}\right)^{\varpi_{i'-d-1}} \\ &= 1 - e^{-\Theta\left(\frac{\deg_v \cdot \varpi_{i'-d-1}}{n}\right)} \\ &= \Theta\left(\frac{\deg_v \cdot \varpi_{i'-d-1}}{n}\right).\end{aligned}$$

Recall that $|\cup_{i\leq i'} \mathcal{V}_i| = \mathcal{O}(\gamma(\mathcal{V})^{i'}) = \mathcal{O}(\log^{\log(\gamma(\mathcal{V}))\cdot c_1} n)$ and $|s(\mathcal{U}(i'))| = \Theta(n)$. Further, note that $|G^{\mathcal{U}}_{k',0}| = \omega(n/\log^{(\alpha/(\alpha-1))\cdot c_1} n)$ for each $k' \leq \log^{c_1-1/\alpha-1} n$ holds due to the power law degree distribution. Therefore,

$$|G^{\mathcal{U}}_{k',i'-d-1}| \geq |G^{\mathcal{U}}_{k',i'}| \geq |G^{\mathcal{U}}_{k',0}| - |\cup_{i\leq i'} \mathcal{V}_i| \geq |G^{\mathcal{U}}_{k',0}| \cdot (1 - o(1)) \geq n/\log^{\mathcal{O}(1)} n$$

must hold for all groups with degree up to $k = \log^{c_1-1/\alpha-1} n$. Consequently, we obtain that the expected amount of newly notified nodes within $G^{\mathcal{U}}_{k,i'-d-1}$ is at least

$$\begin{aligned}|G^{\mathcal{U}}_{k,i'-d-1}| \cdot \Pr\left(Y_k^{\varpi_{i'-d-1}}\right) &\geq c' \cdot \frac{n}{k^\alpha} \cdot (1 - o(1)) \cdot \Theta\left(\frac{k \cdot \varpi_{i'-d-1}}{n}\right) \\ &= \Omega\left(\frac{\varpi_{i'-d-1}}{k^{\alpha-1}}\right) = \Omega(\log n),\end{aligned}$$

where c' is a normalizing constant. Standard Chernoff bounds imply that this lower bound holds with high probability.

Summarized, $\Lambda^+_{i'}$ contains at least a logarithmic amount of nodes with degree $\log^{c_1-1/\alpha-1} n$. Additionally, note that the notification trees established by these nodes may have up to $\mathcal{O}(\beta(c_1-1/\alpha-1 \cdot \log\log n)^2)$ many levels in total.

Phase 2 Let $\mathcal{V}_0 = \{z\}$, i.e., node z is the source of $\mathcal{V}$. From the argumentation above it follows that there is a path $(z, v_1, v_2, ..., v_{i'-d-1}, v)$ of length exactly $(i'-d)$ for each node $v \in \Lambda_{i'}$, where $z, v_k \in \mathcal{R}(i')$ for all $1 \leq k \leq i'-d-1$. Then the following claim holds.

Claim 5.5. *Each node $v \in \Lambda^-_{i'}$ receives an immunizing message from some node $v' \in \Lambda^+_{i'}$ by round $3 \cdot (i'-d) + d + 1$.*

Proof. Due to Phase 1, we know that process Λ is activated in round i'. Then, by definition, each node $v \in \Lambda_{i'}$ is allowed to send for the first time in round $i'+1$. Further, there must be a path from z to v as well as one from z to v' of length exactly $(i'-d)$. Therefore, each node $v \in \Lambda^-_{i'}$ receives an immunizing message from some node $v' \in \Lambda^+_{i'}$

by round

$$\overbrace{(i'+1)}^{\substack{\text{delay until the nodes in } \Lambda^+_{i'} \text{ are allowed to transmit}}} + \overbrace{(i'-d)}^{\substack{\text{path length from } z \text{ to } v}} + \overbrace{(i'-d)}^{\substack{\text{path length from } z \text{ to } v'}} = 3 \cdot (i'-d) + d + 1$$

and the claim follows. Note that there may be shorter paths from z to v or from z to v', or even shorter (simple) paths in $\cup_{i \leq i'} \mathcal{V}_i$ that do not include z. However, by definition, each node in $\mathcal{R}$ forwards an immunizing message to all of its non-infected neighbors excluding the ones he received said message from, which obviously includes the nodes initially infected by $\mathcal{V}$. □

In the rest of this proof we make use of the following definitions.

Definition 5.7 (plus-path)**.** *Let $u \in V$ and $w \in \Lambda^+_{i'}$ be arbitrary nodes. Then we say that the sequence $(w, v_0, ..., v_j)$ is a* plus-path *$p^+(w,u)$ of length $|p^+(w,u)| = j+1$, if the following conditions hold:*

- *v_0 was infected by w,*
- *$v_j = u$, and*
- *for $0 \leq i < j$: v_i infected v_{i+1}.*

Definition 5.8 (minus-path)**.** *Let $u \in V$ and $v \in \Lambda^-_{i'}$ be arbitrary nodes. Then we say that the sequence $(v, v_0, ..., v_j)$ is a* minus-path *$p^-(v,u)$ of length $|p^-(v,u)| = j+1$, if the following conditions hold:*

- *v_0 was infected by v,*
- *$v_j = u$, and*
- *for $0 \leq i < j$: v_i infected v_{i+1}.*

Definition 5.9 (k-blocked)**.** *Let $v \in \Lambda_i$ be an arbitrary transmitting node. Then we say that v is k-blocked in round i, if v has exactly k infected neighbors in round i.*

Definition 5.10 (Λ^+/Λ^- reaches some node $v \in V$)**.** *We say that Λ^+/Λ^- reaches some node v in round i, if at least one preventive message generated by a node in $\Lambda^+_{i'}/\Lambda^-_{i'}$ is received by v in round i.*

Definition 5.11 (i-neighborhood)**.** *We define the i-neighborhood $\Gamma_i(v)$ of a node $v \in V$ as*

$$\Gamma_i(v) := \{w \in V \mid \text{there is some simple path from } v \text{ to } w \text{ in } G \text{ of length at most } i\}.$$

Definition 5.12 ($v \in V$ *resumes* a transmission)**.** *Let $v \in V$ be an arbitrary infected node in round $j \geq i'$ and let $u \in \Lambda_j$ be an arbitrary transmitting neighbor of v. By definition v blocks the transmission T of u in round j. We say that v resumes the transmission T in round i if v is reached by Λ^+ in round $i-1$.*

From now on we primarily focus on the dissemination started by the nodes contained in $\Lambda_{i'}^+$. Recall that the nodes in $\Lambda_{i'}^+$ have a degree large enough to send at least $\beta(\log\log n)^2$ many rounds. Let $z^* \in \Lambda_{i'}^+$ be an arbitrary but fixed node with degree at least $\log^{c_1-1/\alpha-1} n$. From Phase 1 we know that such nodes exist with high probability. In the following we show that the i-neighborhood of z^* is entirely immunized by round

$$\overbrace{(i'+1)}^{\substack{\text{delay until } \Lambda \text{ is allowed to} \\ \text{transmit for the first time}}} + i \cdot \overbrace{(2(i'-d)+d+2).}^{\substack{\text{max. amount of rounds an infected} \\ \text{node may delay a transmission from } \Lambda}}$$

That is, we show that an infected neighbor u of a transmitting node $v \in \Lambda$ can prevent Λ^+ from reaching u for at most $\mathcal{O}(\log\log n)$ many rounds. We prove this claim by induction over the (graph) distance i of each node in Λ to z^*. More formally, we show the following lemma.

Lemma 5.11.

$$\Gamma_i(z^*) \cap \left(\bigcup_{j \leq (2(i'-d)+d+2)\cdot i} \Lambda_{(i'+1)+j} \right) = \Gamma_i(z^*)$$

for any distance $i \geq 1$, where $\Gamma_i(z^*)$ is the i-neighborhood of z^*.

The following observations are a vital part of the upcoming proof:

1. Λ^+ reaches each node $v \in \Lambda_j$, for an arbitrary round $j \geq i'$, for which a minus-path exists by round

$$\overbrace{3(i'-d)+d+1}^{\substack{\text{max. delay until } \Lambda^+ \\ \text{reaches each node in } \Lambda_{i'}^-}} + \overbrace{(j-d)-(i'-d)}^{\substack{\text{length of the minus-path} \\ \text{from some } w \in \Lambda_{i'}^- \text{ to } v}} = 2(i'-d)+j+1 \text{ , and}$$

2. each infected neighbor u of a node $v \in \Lambda_j$ resumes all transmissions of v prevented by u in round $j \geq i'$ by round

$$\overbrace{3(i'-d)+d+1}^{\substack{\text{max. delay until } \Lambda^+ \\ \text{reaches each node in } \Lambda_{i'}^-}} + \overbrace{(j-(i'-d))}^{\substack{\text{remaining max. delay} \\ \text{until } \Lambda^+ \text{ reaches } u}} +1 = 2(i'-d)+d+2+j.$$

Note that we only need to consider the very first blocked transmission by each infected node. Although the same argument would apply for all subsequently blocked transmissions too, the initial one is, in the worst case, resumed first. However, in the case that the initial blocked transmission is not resumed first, the entailed delay by said transmission does still represent an upper bound for the dissemination of Λ.

Proof of Lemma 5.11.

We begin with distance $i = 1$: The first round where the nodes in Λ are allowed to transmit is $i'+1$. By definition the following happens in round i':

1. all infected nodes up to (and including) round $i' - d - 1$ are immunized, i.e., they are moved to $\mathcal{R}$, and

2. process Λ is activated on each node in $\mathcal{V}_{i'-d}$, i.e., these nodes change their status to *immunized* and establish notification trees.

Therefore, $\Lambda_{i'}$ is non-empty. Further, due to Phase 1, we know that there is a path $(z, v_1, v_2, ..., v_{i'-d-1}, v)$ of length exactly $(i' - d)$ for each node $v \in \Lambda_{i'}$, where $z, v_k \in \mathcal{R}(i')$ for each $1 \leq k \leq i' - d - 1$.

In the following we examine the situation for an arbitrary infected neighbor u of an arbitrary node $v \in \Lambda_{i'}$. We first focus on the case where

- there is no minus-path $p^-(w, u)$ for some node $w \in \Lambda^-_{i'}$, and
- the transmission of v to u in round $i' + 1$ is blocked by u.

Then there must be a plus-path $p^+(v', u) = (v', u_0, ..., u_j)$ in round $i' + 1$ of length $j + 1$, where $u_j = u$ and $1 \leq j \leq d$. Further, by assumption, it holds that $v' \in \Lambda^+_{i'}$. Consequently, Λ^+ reaches u after at most $j + 1 \leq d + 1$ additional rounds by traversing $p^+(v', u)$. Therefore, the transmission blocked by u in round $i' + 1$ is resumed after at most

$$d + 2$$

additional rounds in total.

In the case that the transmission of v in round i' is blocked by u under the condition that

- there is no plus-path $p^+(w, u)$ for some node $w \in \Lambda^+_{i'}$,

u is not guaranteed to resume the previously blocked transmission by the round Λ^- reaches u. Notice that in this case there must be a minus-path $p^-(v', u) = (v', u_0, ..., u_j)$ in round $i' + 1$ of length at most $j + 1$, where $u_j = u$ and $1 \leq j \leq d$. However, by Claim 5.5 we know that Λ^+ reaches v' by round $3(i' - d) + d + 1$, i.e., after at most $2(i' - d)$ additional rounds. Since $|p^-(v', u)| = j + 1 \leq d + 1$, it follows that node u resumes the previously blocked transmission after at most

$$2(i' - d) + d + 2$$

additional rounds in total. Consequently, we obtain that

$$\Gamma(z^*) \cap \left(\bigcup_{j \leq (2(i'-d)+d+2)} \Lambda_{(i'+1)+j} \right) = \Gamma(z^*)$$

combining both cases.

Now consider an arbitrary distance $i > 1$: Let $v \in \Lambda_{j'}$, for some round j', be an arbitrary (but fixed) transmitting node at distance exactly i from z^*.

We distinguish two cases:
v is 0-blocked By definition v is incident to already immunized and/or uninformed nodes only. Then all its currently uninformed neighbors become immunized by the end of round j'.

v is k-blocked for $k > 0$ By definition v has exactly k infected and $\deg_v - k$ many uninformed or immunized neighbors. Note that for each preventive message received by v in round $j' - 1$ at least one node w among these $\deg_v - k$ many nodes must exist that has transmitted said message to v. Let M be such a preventive message and let k_M be the number of nodes that transmitted M to v in round $j' - 1$. Then it follows that the remaining $\deg_v - k - k_M$ non-infected nodes receive M by the end of round j'. Further, by definition, the k infected neighbors eventually become immunized after at most d many additional rounds.

Let k^+ be the number of nodes among v's k many infected neighbors u satisfying $|p^+(w, u)| > 0$, for some node $w \in \Lambda^+_{i'}$, and let the set $S^+ = \{v_1, ..., v_{k^+}\}$ contain all these nodes. Further, let the set $S^- = \{v_1, ..., v_{k^-}\}$ contain all v's infected neighbors u satisfying $|p^-(w', u)| > 0$, for some node $w' \in \Lambda^-_{i'}$. Notice that $k = k^+ + k^-$. In the following we consider S^+ and S^- separately.

By assumption there is a plus-path p^+ for each node $u \in S^+$. Since each newly infected node is automatically immunized after d many rounds and Λ^+ is forwarded along p^+, it follows that each node in $\Lambda_{j'}$ on this path is reached by Λ^+ by round $j' + 1$. Recall that the nodes in $\Lambda_{i'}$ are allowed to send for the first time in round $i' + 1$. Then after at most d many additional rounds each node $u \in S^+$ is reached. Consequently, Λ^+ reaches each node in S^+ and the previously blocked transmissions are resumed after at most

$$d + 2$$

additional rounds in total.

On the other hand, the situation for the nodes in S^- is more complicated. W.l.o.g. let $u \in S^-$ be an arbitrary infected neighbor of v in round j'. Then there must be a minus-path $p^-(w, u) = (w, ..., u_0, ..., u_j)$, for some node $w \in \Lambda^-_{i'}$ and $u_j = u$, of length at most $j' - i' + d$, where $u_0 \in \Lambda_{j'}$ and $1 \leq j \leq d$. Claim 5.5 implies that Λ^+ reaches u_0 by round

$$\overbrace{3(i' - d) + d + 1}^{\substack{\text{delay until } \Lambda^+ \text{ reaches} \\ \text{each node in } \Lambda^-_{i'}}} + \overbrace{(j' - i')}^{\substack{\text{path length of the minus-path} \\ \text{from some } w \in \Lambda^-_{i'} \text{ to } u_0}} = 2(i' - d) + j' + 1.$$

Therefore, all transmissions blocked by u in round j' are resumed after at most

$$2(i' - d) + d + 2$$

additional rounds in total.

Due to the induction hypothesis we know that $\Gamma_i(z^*) \cap \left(\bigcup_{j \leq (2(i'-d)+d+2)\cdot i} \Lambda_{(i'+1)+j}\right) = \Gamma_i(z^*)$. Further, note that $j' \leq ((2(i'-d)+d+2)\cdot i)$, since $j' > ((2(i'-d)+d+2)\cdot i)$ would imply that at least one node on the path to v would have suffered a delay larger than $2(i'-d)+d+2$. By combining the above arguments we conclude that

$$\Gamma_{i+1}(z^*) \cap \left(\bigcup_{j \leq (2(i'-d)+d+2)\cdot(i+1)} \Lambda_{(i'+1)+j} \right) = \Gamma_{i+1}(z^*).$$

□

At this point we are ready to conclude. Considering the dissemination of Λ, it is obvious that the worst case occurs whenever all transmitting nodes of Λ are blocked by infected ones. However, these nodes become immunized after at most d many rounds and resume all previously blocked transmissions at most $2(i'-d)+2$ rounds later. It follows that an infected neighbor of a node $v \in \Lambda$ can prevent Λ^+ from reaching u for at most $2(i'-d)+d+2 < 3i'$ many rounds, since i' is large enough.

Since G is a random power law graph with power law exponent $2 < \alpha < 3$, we know that the average distance between two nodes in G is upper bounded by $\nu \cdot \log\log n$, for some constant ν (cf. [24], [70]). Further, recall that the very first infected node is chosen uniformly at random, which implies that the majority of all nodes is within distance $\nu \cdot \log\log n$ of said node with probability $1 - o(1)$.

Let $\beta = \max\{\nu d, 10c_1\}$, and $c_1 = \max\{(\nu\alpha+1)(\alpha-1)+1, (100+1)(\alpha-1)+1\}$. Recall that there are some nodes in $\Lambda_{i'}^+$ with degree at least $\log^{c_1-1/\alpha-1} n$. Consequently, the lower bound for the amount of rounds a preventive message that was created by a node with the highest degree in $\Lambda_{i'}^+$ is allowed to be forwarded is given by

$$\beta \left(\log\left(\log^{\frac{c_1-1}{\alpha-1}-1} n\right)\right)^2 = \beta \left(\left(\frac{c_1-1}{\alpha-1} - 1\right)\log\log n\right)^2 > 3i' \cdot 5\nu \log\log n.$$

Hence, all nodes in Λ^+ can easily reach the majority of the nodes within G with probability $1 - o(1)$. □

5.2.3 Glossary

At this point we list some of the most important definitions we used in the above analysis.

Definitions concerning the network

G	G is an undirected random graph, also called network, our dissemination processes work on.
n	The number of nodes in G is denoted by n.
m	The number of edges in G is denoted by m.
$\deg_v$	Degree of node $v \in V$.
$\Gamma(v)$	Neighborhood of node v.
δ	The minimum degree of G.
α	The power law exponent used to determine G's degree distribution, where $2 < \alpha < 3$.
$G_k, G_{k,i}^{\mathcal{U}}$	G_k is the set (also called group) containing all nodes with degree k, whereas $G_{k,i}^{\mathcal{U}} := G_k \cap \mathcal{U}(i)$.
$s(v), s_i^v$	$s(v)$ is the set of all *stubs* for a node $v \in V$, with s_i^v being the ith stub in an ordered sequence.
$s_f(v), \overline{s_f(v)}$	The set $s(v)$ is divided into a set of free stubs $s_f(v)$, and a set of already paired stubs $\overline{s_f(v)}$. Hereby, free stubs are available for the pairing process, whereas paired stubs are already connected to edges in G.
$s(V')$	$s(V') = \bigcup_{v' \in V'} s(v')$ defines the (overall) set of stubs for some subset $V' \subseteq V$.
$s_f(V')$	$s_f(V') = \bigcup_{v' \in V'} s_f(v')$ defines the set of free stubs for some subset $V' \subseteq V$.
$S_f(G)$	Set of all free stubs for a corresponding round. Since said round is always clear from the context, we omit to specify it explicitly.

Definitions concerning the node states (SIR-model)

$\mathcal{U}$	$\mathcal{U}$ contains the nodes that are not notified yet in any way.
$\mathcal{I}$	$\mathcal{I}$ contains the nodes that were already notified exclusively by process $\mathcal{V}$ and did not recover so far.
$\mathcal{R}$	$\mathcal{R}$ contains all nodes that recovered or that were already notified by process Φ or Λ and thus are neither a part of $\mathcal{U}$ nor $\mathcal{I}$.

Definitions concerning the dissemination processes

$\mathcal{V}$	$\mathcal{V}$ is a dissemination process trying to notify as many nodes of the network as possible with a malicious piece of information. Depending on the model in use, said process is counteracted by either Φ or Λ.
$\gamma(\mathcal{V})$	The maximum number of different neighbors process $\mathcal{V}$ can use for spreading.
d	The delay between the notification of a node $v \in \mathcal{U}(i-d-1)$ by $\mathcal{V}$ and the activation of Φ (or Λ) on said node in round i.
Φ	Φ is a dissemination process counteracting $\mathcal{V}$. Whenever Φ is activated on a node $v \in \mathcal{I}(i)$ in round i, v and all its non-infected neighbors are immunized.
Λ	Λ is a more sophisticated version of Φ. It behaves as follows. Whenever Λ is activated on a node $v \in \mathcal{I}(i-1)$ in round i, v is immunized and creates a new notification tree Υ_v.
$\Upsilon_{v,i}$	The corresponding level of a *notification tree* Υ_v with root v in round i.
$t(v)$	The round $v \in \mathcal{U}$ was initially infected by $\mathcal{V}$.

Parts of this chapter have been published in:

[36, 37]

Chapter 6

The Dynamic Case

In this chapter we examine epidemic spreading in *continuously evolving networks*. That is, in contrast to Chapter 5, we are not restricted to (mostly) fixed networks anymore. On the contrary, now we allow that a network observed at a specific time may be significantly different from its appearance a few moments ago. Note that this may benefit or hinder the dissemination of the epidemic/information, depending on the field of application of the network. For example, an infected individual may reach (and therefore infect) more potential hosts by being mobile. Indeed, the impact of an epidemic process is amplified by dynamic networks. Therefore, distributing an antidote/anti-rumor by a similar (yet slightly more powerful) process as the malicious one will certainly prevent parts of the network from being harmed. Due to conformity, we first present some examples of networks that are considered throughout this chapter. Afterwards, we give a short outline for the upcoming sections.

In order to present some examples of networks that fit in the context of this chapter, we use the same categorization as in Chapter 5. Namely:

- Technical Context, and
- Social Context.

We begin with the technical context. Recall that we focus on the dissemination of epidemic processes in dynamic networks throughout this chapter. Therefore, many wireless networks belong here. Note that a significant amount of devices within a dynamic network is usually mobile. Consequently, the network's structure is, to some extend, constantly in transition. Let us consider vehicular ad hoc networks as an example. Clearly, the moving vehicles provoke a constant change of the topology.

Note that we are not limited to networks created and maintained by artificial devices only. On the contrary, contact networks formed by mobile individuals are of special interest to us. As a consequence, the scope of our research with respect to the amount of scenarios and use cases worth considering expands drastically as soon as we take individuals in the real world and their interactions into account. For example, contact based graphs representing interpersonal encounters are often essential in the field of

epidemiology. Often it is crucial to know exactly who contacted whom at a specific time and place in order to efficiently counteract an epidemic.

The rest of this chapter is structured as follows. In Section 6.1 we analyze the dissemination of an epidemic in a constantly changing (urban) environment with multiple locations. This environment represents an instance of a virtual city inhabited by mobile individuals. Each moving individual chooses its target location independently at random according to the location's attractiveness. Said attractiveness is randomly assigned according to a power law distribution. Due to the individuals' movement, the contact graph is constantly in transition.

In this setting we gain two main insights:

1. a deadly epidemic does not manage to extinct the entire population under realistic (yet weak) assumptions with high probability, even if the behavior of the susceptible entities is not influenced by the epidemic outbreak in any way, and

2. already very limited pharmaceutical and non-pharmaceutical countermeasures suffice to drastically mitigate the impact of an epidemic.

In Section 6.2 we extend the setting of Section 6.1 and examine the implications on a large scale. However, here we analyze our model empirically. In short, the setting is as follows. The environment approximates the geography of Germany. Each individual may not only move within its current city but may also travel between cities, of which there are plenty. All individuals within a city interact according to the probabilistic model presented in Section 6.1 in a distributed manner.

In essence, we gain the following insights:

1. our model shows significant similarities to real-world observations of the course of the influenza virus, therefore indicating the usefulness of said model while simultaneously backing up the findings of [39], and

2. the effect of some non-pharmaceutical countermeasures that are usually adopted if an epidemic outbreak occurs can be easily approximated by adjusting the parameters of our model properly.

6.1 Epidemic Spreading in Urban Environments

In this section we consider an epidemic process operating in a dynamically changing environment. This environment represents an urban topology inhabited by mobile susceptible entities. Consequently, due to the (random) movement of these entities, the vicinity of each one is constantly changing. In contrast to Chapter 5, all types of mobile networks covered by this chapter are contact based and/or wireless. We refer the reader to Section 6.1.1 for the definition.

In order to analyze the above network, two models are presented:

1. the regular model (REG), where the behavior of the susceptible entities is not influenced by an epidemic outbreak in any way, and
2. the restricted model ($REST$), where an epidemic outbreak does influence the behavioral pattern of the susceptible entities. Such an influence may be represented by a mobility restriction (such as a curfew, isolation or school closure) or some kind of pharmaceutical intervention (such as self medication or vaccination), for example.

In the regular model we show that a polynomial fraction of the susceptible entities remains uninfected, even if each infected entity would transmit the infection for $o(\log n)$ many rounds. That is, it is very unlikely for an epidemic to wipe out the entire population, provided that the environment is large enough. Interestingly, our result provides an analytical evidence for a conjecture expressed in a historical documentation about the plague in the mid ages [46].

In the restricted model, on the other hand, we show that only a polylogarithmic amount of entities is infected until the epidemic is completely embanked, provided that suitable countermeasures are taken (see Section 6.1.1 for more details). We show that said embankment occurs within the first $\mathcal{O}((\log \log n)^4)$ rounds with probability $1 - o(1)$.

6.1.1 Theoretical Model

The outline of this section is the following. First, we describe our model and some basic definitions in more detail. Then we state some properties entailed by said model.

We present our model with respect to four main components:

1. the topology/environment,
2. the susceptible entities,
3. the epidemic dissemination procedure, and

4. the society models (i.e., the response pattern to an epidemic outbreak).

The Topology The topology is modeled by a complete graph $G = (V, E)$ (also called network) with κn nodes (also called *cells*), where $\kappa \in \mathbb{R}_{\geq 0}, n \in \mathbb{N}$. These cells represent locations a susceptible entity (also called agent or individual) can visit. Each cell may accommodate agents depending on the cells so-called *attractiveness*. The *attractiveness* d_v of a cell v is chosen randomly with probability proportional to $1/i^\alpha$, where $\alpha > 2$ is a constant. We define the highest attractiveness to be $\sqrt[\alpha]{\kappa n}$ and the lowest to be 2. Since all agents move randomly (see below), it follows that the expected amount of agents accommodated in a specific cell v with attractiveness d_v is given by $n \cdot c' \cdot d_v / \sum_{i=2}^{\sqrt[\alpha]{\kappa n}} c\frac{\kappa n}{i^\alpha} i$, where c and c' are some normalizing constants such that $\sum_{i=2}^{\sqrt[\alpha]{\kappa n}} c\frac{\kappa n}{i^\alpha} = \kappa n$ and $n \cdot \sum_{v \in V} \left(c' \cdot d_v / \sum_{i=2}^{\sqrt[\alpha]{\kappa n}} c\frac{\kappa n}{i^\alpha} i \right) = n$.

The Susceptible Entities The amount of susceptible *entities/agents* in the environment is given by $n \in \mathbb{N}$. These mobile agents may move from one cell to another. That is, in each round each agent chooses its next location independently with probability proportional to the location's attractiveness. Further, each agent has a status indicating its state of knowledge with respect to the epidemic. This status partitions the set of agents S into three disjoint subsets:

1. $\mathcal{I}(j)$, meaning the set of all infected agents in round j,
2. $\mathcal{U}(j)$, meaning the set of all uninfected (susceptible) agents in round j, and
3. $\mathcal{R}(j)$, meaning the set of all resistant/immunized agents in round j.

Consequently, $S := \mathcal{I}(j) \cup \mathcal{U}(j) \cup \mathcal{R}(j)$. Whenever it is clear from the context, we simply write $\mathcal{I}$, $\mathcal{U}$, and $\mathcal{R}$, respectively.

The Epidemic Dissemination In general, epidemics may be spread in various ways in an urban environment. However, to emphasize the importance of mobility, we assume that locations are not contagious if they do not accommodate at least one infected agent. Under this assumption the dissemination procedure can be described as follows. If in some round j an uninfected agent v visits a cell which also contains an agent w of $\mathcal{I}(j)$, v becomes infected with probability 1 and carries the disease further. Note that one can easily extend this model to the case where the disease is spread with some probability not equal to 1. Further, we define the time until an infected individual is cured again as τ. That is, a time period τ is assigned to the epidemic, which means that an individual is contagious for τ consecutive rounds only.

Society Models We specifically consider two different society models:

1. the regular model (*REG*), where the behavior of the susceptible entities is not influenced by an epidemic outbreak in any way, and

2. the restricted model (*REST*), where an epidemic outbreak does influence the behavioral pattern of each susceptible entity.

The regular model primarily serves the purpose of highlighting the importance of movement patterns we are already used to in such an urban setting. That is, in the regular model we show how our own habits aid epidemics in order to spread more efficiently. On the other hand, the restricted model represents our ability to cope with and counteract such an outbreak despite our lifestyle. Formally, this translates to the following parameter settings. Note that the individual values were chosen according to publicly available literature where possible and our best judgment otherwise.

Society model	α	κ	τ
The regular society model (*REG*)	$\alpha \in (2,3)$	1	$o(\log n)$
The restricted society model (*REST*)	$3 < \alpha \in \mathcal{O}(1)$	$\Theta(1)$	$\Theta(1)$

6.1.2 Analysis

At this point we present some basic observations. As we show below, the number of cells with a high attractiveness decreases with an increasing α. On the other hand, the area in which infected agents may infect others decreases while κ increases (cf. Figure 6.1). Recall that the amount of agents is independent of κ.

Observation 1. *Let $C := \{v_1, ..., v_j\}$ be the set of all cells with attractiveness $d \in \{2, ..., \sqrt[\alpha]{\kappa n}\}$. Then the expected amount of agents choosing a specific cell $v_i \in C$ and an arbitrary cell in C is proportional to d and $nd^{-\alpha+1}$, respectively.*

Proof. The probability of choosing a specific cell $v \in C$ with attractiveness d_v is given by

$$\frac{c' \cdot d_v}{\sum_{i=2}^{\sqrt[\alpha]{\kappa n}} c\frac{\kappa n}{i^\alpha} i} = \Theta\left(\frac{d_v}{n}\right),$$

where c and c' are the normalizing constants defined in the previous section. Similarly, the probability of choosing an arbitrary cell with attractiveness d is given by

$$\frac{\Theta\left(n/d^{\alpha-1}\right)}{\sum_{i=2}^{\sqrt[\alpha]{\kappa n}} c\frac{\kappa n}{i^\alpha} i} = \Theta\left(d^{-\alpha+1}\right).$$

Since the amount of agents equals n, the claim follows. □

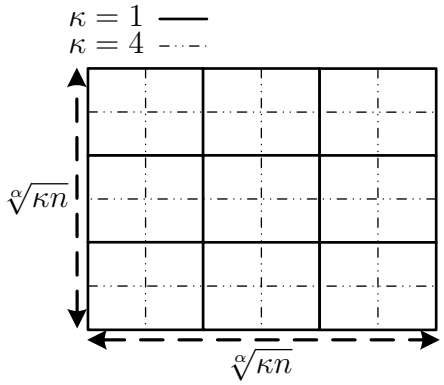


Figure 6.1: *The solid line indicates a grid-like visualization of a topology A (complete graph), where* $\kappa = 1, n = 9$*, and* $\alpha = 2$*. The broken line indicates the corresponding topology B with* $\kappa = 4$*, resulting in* 36 *cells.*

6.1.2.1 The Regular Society Model

Suppose a non curable course of a deadly disease. Consequently, all infected individuals decease eventually. In the following we show that at least a polynomial fraction of the uninfected agents stays at locations which are not occupied by infected ones due to the urban environment. This implies that the population in such a setting would outlast even the worst epidemic. Theorem 6.1 is the main statement of this section.

Theorem 6.1. *Let* $\kappa = 1$ *and let* τ *be a slow-growing function in* n *(i.e.,* $\tau = o(\log n)$*). Then a polynomial fraction of the population remains uninfected when* $\mathcal{I} = \emptyset$ *with high probability.*

The theorem implies that there must be some point in time where only a small fraction of the population is still alive and uninfected while another (rather small) fraction still carries the infection further. We model this situation (in some round j) by assuming that $|\mathcal{I}(j)| \cdot |\mathcal{U}(j)| \leq n^{2\epsilon}$, for some small constant $\epsilon \leq 1/4$. Now the most interesting question is whether or not this epidemic process manages to infect the remaining healthy individuals, therefore exterminating the entire population. The following lemma copes with this situation and is a vital part of the proof of Theorem 6.1.

Lemma 6.1. *Let* $|\mathcal{I}(j)| \cdot |\mathcal{U}(j)| \leq n^{2\epsilon}$*, where* $\epsilon \leq 1/4$ *is an arbitrarily small constant. Further, let* κ *be a constant and* τ *a slow-growing function in* n *(i.e.,* $\tau = o(\log n)$*). Then there is no newly infected agent in round* j *with probability* $1 - n^{-\Omega(1)}$*.*

Proof. Let a pairing of two agents $i \in \mathcal{I}(j)$, $i' \in \mathcal{U}(j)$ describe the event that i and i' choose the same cell. Let $x_{i,i',d}$ be the event that agents i and i', where $i \neq i'$, choose the same (specific) cell with attractiveness $d \geq 2$. Then the probability $\Pr(x_{i,i',d})$ is

bounded by

$$\Pr(x_{i,i',d}) = \left(\frac{d}{\sum_{l=2}^{\sqrt[\alpha]{\kappa n}} c\frac{\kappa n}{l^\alpha} l}\right)^2 \leq \left(\frac{d}{c\kappa n \frac{1}{\alpha-2}\left(\frac{1}{2^{\alpha-2}} - o(1)\right)}\right)^2 \leq \left(\frac{d}{c\kappa n \frac{1}{\alpha-2}\frac{1}{2^{\alpha-2}}\frac{1}{2}}\right)^2 \leq \left(\frac{d\cdot(\alpha-2)\cdot 2^{\alpha-1}}{c\kappa n}\right)^2, \tag{6.1}$$

where c is defined as in Section 6.1.1. Further, let $x_{i,i'}$ be the event that agent i and i' meet in an arbitrary cell. Then $\Pr(x_{i,i'}) \leq \sum_{d=2}^{\sqrt[\alpha]{\kappa n}} c\frac{\kappa n}{d^\alpha} \Pr(x_{i,i',d})$ and we obtain that

$$\begin{aligned}\Pr(x_{i,i'}) &\leq \left(\frac{(\alpha-2)\cdot 2^{\alpha-1}}{c\kappa}\right)^2 \sum_{l=2}^{\sqrt[\alpha]{\kappa n}} \frac{c\kappa n}{l^\alpha}\cdot\frac{l}{n}\cdot\frac{\sqrt[\alpha]{\kappa n}}{n} \\ &= \mathcal{O}\left(\frac{1}{n^{1-1/\alpha}}\right),\end{aligned}$$

where we used that $\left(\frac{l}{n}\right)^2 \leq \frac{l\sqrt[\alpha]{\kappa n}}{n^2}$.

A fixed uninfected agent i' becomes infected with probability at most $|\mathcal{I}(j)|\Pr(x_{i,i'})$ and the expected amount μ of newly infected agents is bounded by

$$\mu \leq |\mathcal{U}(j)||\mathcal{I}(j)|\Pr(x_{i,i'}) \leq \mathcal{O}\left(\frac{1}{n^{1-1/\alpha-2\epsilon}}\right).$$

Since the agents of $\mathcal{U}(j)$ are assigned to the cells independently, we use Chernoff bounds (cf. Lemma 4.1) to obtain the desired result. With $(1+\delta)\mu = 1/2$ and X being the random variable describing the amount of newly infected agents, it follows that

$$\Pr[X > (1+\delta)\mu] \leq \left[\frac{e^\delta}{(1+\delta)^{(1+\delta)}}\right]^\mu \leq \left(\frac{e^{1/2\mu-1}}{(1/2\mu)^{1/2\mu}}\right)^\mu = \frac{(2\mu\cdot e)^{1/2}}{e^\mu} = n^{-\Omega(1)}.$$

□

At this point we are ready to prove the main statement of this section.

Proof of Theorem 6.1.

The proof consists of three phases. In the first phase we assume that $\mathcal{R} = \emptyset$. That is, we focus on the increase of the number of infected agents only. However, we will ensure that a sufficient number of uninfected agents will still be present. Note that this number may be very small compared to the size of the population. In the second phase then we distinguish between two situations: 1.) the spreading process runs out (i.e., $\mathcal{I} = \emptyset$) although the number of uninfected agents is polynomial in n, or 2.) we have a situation where the assumptions of Lemma 6.1 are fulfilled. In the latter case we apply Lemma 6.1 in a separate third phase. Let in the following $i \geq 0$ denote the current round with respect to the considered phase, and let t_1, t_2 describe the round *after* Phase 1 and Phase 2 have ended, respectively.

Phase 1 Throughout this phase we show that $|\mathcal{I}(i+1)| > |\mathcal{I}(i)|$ and $|\mathcal{U}(i+1)| = \Omega(|\mathcal{U}(i)|)$, for all considered rounds i. Further, we assume that $\mathcal{R} := \emptyset$. Thus, all infected agents remain infected during this phase and carry the disease to all agents they meet. According to the power law distribution of the attractiveness, a constant fraction of cells will have attractiveness $d = 2$[1]. Recall that $\kappa = 1$ and hence the number of all cells is n in total[2]. Therefore, we obtain that the probability for a cell to be empty and to have a constant attractiveness c' is given by

$$p := \frac{c}{(c')^{\alpha}} \cdot \left(1 - \frac{c'}{cn \sum_{i=2}^{\sqrt[\alpha]{\kappa n}} (1/i)^{\alpha-1}}\right)^{n} = \Theta(1),$$

where c is defined as in Section 6.1.1.

Now we use a simple balls into bins game to obtain the amount of these cells. Recall that there are $\Theta(n)$ many cells with constant attractiveness overall. We use two bins to determine whether one cell with constant attractiveness remains empty or not: the first one is chosen with probability p, representing that the cell remains *empty*, whereas the other one represents the opposite with probability $q := 1 - p$. Consequently, it follows that the number of empty cells with a constant attractiveness is $\Theta(n)$. Since each agent chooses a cell with probability proportional to its attractiveness (cf. Observation 1), it follows that the probability for an uninfected agent to remain uninfected in this round is constant. Further, because the uninfected agents are assigned to the cells independently, we apply Chernoff bounds to conclude that a constant fraction of $\mathcal{U}(i)$ remains uninfected after round i with probability $1 - n^{-\Omega(1)}$ as long as $|\mathcal{U}(i)| = n^{\Omega(1)}$.

Phase 2 Now we add the curing procedure to our analysis. Thus, $|\mathcal{R}(t_1 + i)|$ will potentially increase in each round. From Phase 1 we know that at least a constant fraction of the uninformed nodes remains uninformed in each round, i.e., that $|\mathcal{U}(j+1)| = \Theta(|\mathcal{U}(j)|)$ holds for any round j. Consequently, one of the following situations occurs:

1. the number of uninfected agents does not decrease below some polynomial in n while the epidemic runs out, or
2. there is some round $j \geq t_1$, such that $|\mathcal{I}(j) \cup \mathcal{R}(j)| = n - n^{\epsilon'}$ and $|\mathcal{U}(j)| = n^{\epsilon'}$ for some $\epsilon' > 0$ small enough.

In the following we focus on the second situation.

By definition it follows that $|\mathcal{R}(j+\tau)| \geq |\mathcal{I}(j)| + |\mathcal{R}(j)|$, since each infected agent is cured after τ rounds. Further, since $\tau = o(\log n)$ and $|\mathcal{U}(j+1)| = \Omega(|\mathcal{U}(j)|)$ for any individual round j, we obtain that $|\mathcal{U}(j+\tau)| \geq |\mathcal{U}(j)|/n^{o(1)}$ after τ rounds, with high probability. Note that $\mathcal{U}$ can only decrease if uninformed nodes are informed/infected during the corresponding round. Therefore, it follows that $|\mathcal{I}(j+\tau)| \leq |\mathcal{I}(j)| + |\mathcal{U}(j)|(1 - 1/n^{o(1)})$

[1] In fact, this holds for any constant value.
[2] One can easily extend this to an arbitrary constant κ.

with high probability. However, all nodes in $\mathcal{I}(j)$ are cured in round $j+\tau$. The statements so far lead to the following properties:

$$\frac{|\mathcal{U}(j)|}{n^{o(1)}} \leq |\mathcal{U}(j+\tau)| \leq |\mathcal{U}(j)| = n^{\epsilon'},$$
$$1 \leq |\mathcal{I}(j+\tau)| < n^{\epsilon'}.$$

Consequently,

$$|\mathcal{I}(j+\tau)| \cdot |\mathcal{U}(j+\tau)| \in \left(\frac{n^{\epsilon'}}{n^{o(1)}}, n^{2\epsilon'}\right),$$

where we assumed that $\mathcal{I}(j+\tau) \neq \emptyset$. Thus, there is some constant $\epsilon > 0$ small enough, such that $|\mathcal{I}(j+\tau)| \cdot |\mathcal{U}(j+\tau)| \leq n^{2\epsilon}$.

Phase 3 Since $|\mathcal{I}(t_2)| \cdot |\mathcal{U}(t_2)| \leq n^{2\epsilon}$ at this point, applying Lemma 6.1 for τ rounds gives the theorem. □

6.1.2.2 The Restricted Society Model

In the restricted model we combat the epidemic actively, meaning with actions such as isolation or medication, and passively, meaning with actions such as curfews or gathering bans. Further, we assume that each infected individual spreads the disease until the symptoms become visible. Afterwards, this individual is being isolated. Additionally, we assume that the disease has an incubation time not larger than some constant $\mathfrak{s}$, and model this by setting $\tau = \mathfrak{s}$. Our main statement for this section is as follows.

Theorem 6.2. *Let an epidemic disease be spread in a complete graph with κn cells as described in Section 6.1.1. Then, for all fixed $\tau \in \mathcal{O}(1)$, there exist suitable constants κ and α, such that $\mathcal{I} = \emptyset$ after $\mathcal{O}((\log\log n)^4)$ many rounds with probability $1 - o(1)$. Moreover, at the end (when $\mathcal{I} = \emptyset$) the set $\mathcal{R}$ has size $\log^{\mathcal{O}(1)} n$.*

In contrast to the previous model, we assume that an epidemic outbreak does influence the individuals' behavioral pattern. To be more specific, we assume that this effect is primarily caused by warnings communicated through the media. In our model a warning basically affects the constants α and κ due to the following observation. Whenever individuals are aware of such a warning, many of them will avoid places with a large number of persons, waive needless tours, and be more careful when meeting other people. Although one of these modifications alone may not be sufficient to stop the epidemic [64], we argue that a combination of these strategies, which is able to sufficiently influence the constants α and κ, is enough to stop the disease from spreading. Consequently, we assume that α and κ are suitable (large) constants throughout this section.

Before we prove our main statement, we first show that the distribution of the attractiveness of the cells implies an upper bound on the amount of new infections. Note that we represent the amount of infections by a function in n, i.e., we assume that $|\mathcal{I}(j)| = (f(n))^q$ in round j and some suitable constant q.

Lemma 6.2. *Let the set G_k contain all cells with attractiveness 2^k up to $2^{k+1}-1$ and let $|\mathcal{I}(j)| = f^q(n)$ for a specific round j, where $f(n)$ is some function with $\lim_{n\to\infty} f(n) = \infty$ and $q > 3$ is a constant. If $\alpha > \max\{q-1, 3\}$ and*

$$2 \leq k \leq \frac{1}{(\alpha-2)} \cdot \log\left(\frac{|\mathcal{I}(j)|}{f^3(n)}\right),$$

then the number of newly infected agents in G_k is bounded by

$$\mathcal{O}\left(\frac{|\mathcal{I}(j)|}{(2^k-1)^{\alpha-3}}\right)$$

with probability $1 - e^{-\Omega(f(n))}$.

Proof. The power law distribution of the attractiveness of the cells implies that the expected number of infected agents that choose cells in group G_k with $k \leq \frac{1}{(\alpha-2)} \cdot \log\left(\frac{|\mathcal{I}(j)|}{f^3(n)}\right)$ at the beginning of round j is bounded by

$$|\mathcal{I}(j)| \cdot \frac{\sum_{d=2^k}^{2^{k+1}-1} \frac{c\kappa n}{d^\alpha} d}{\sum_{d=2}^{\sqrt[\alpha]{\kappa n}} \frac{c\kappa n}{d^\alpha} d} = \Theta\left(\frac{|\mathcal{I}(j)|}{(2^k-1)^{\alpha-2}}\right) = \Omega(f^3(n)),$$

where c is the normalizing constant defined in Section 6.1.1.

We formulate the problem as a Vertex Exposure Martingale (cf. Definition 4.3) in order to prove the lemma. Let $z_{a,b}$ be the event that agent a and b choose the same cell. We define a graph $G' = (V', E')$ by setting $V' = \mathcal{I}(j) \cup L$, where $L = \{1, ..., n - |\mathcal{I}(j)|\}$ and $\mathcal{I}(j) \cap L = \emptyset$. The set of edges is $E' = \{(x_i, l) \mid x_i \in \mathcal{I}(j) \wedge l \in L \wedge z_{x_i,l}\}$. Let the considered vertex exposure sequence be given by $x_1, ..., x_{|\mathcal{I}(j)|} \in \mathcal{I}(j)$. Thus, each x_i represents an infected agent, which may establish edges connecting itself to the set L.

In order to estimate the degree of each x_i, we need to upper bound the amount of uninfected agents choosing the same cell as x_i. Recall that each agent chooses its target location for the next round independently at random with probability proportional to the location's attractiveness. Now we assume that no two infected agents choose the same cell. Therefore, if $\alpha > q - 1$, i.e., $2^k = 2^{1/(\alpha-2)\cdot\log(|\mathcal{I}(j)|/f^3(n))} = o(f(n))$, we obtain that at least $\Theta(f(n))$ many uninfected agents choose the same location with an attractiveness of at most $2^{k+1} - 1$ with probability at most

$$\binom{n - |\mathcal{I}(j)|}{\Theta(f(n))} \left(\frac{o(f(n))}{\Theta(n)}\right)^{\Theta(f(n))} \leq \frac{1}{e^{\Omega(f(n))}},$$

since $|\mathcal{I}(j)| = f^q(n)$. Consequently, an infected agent placed in a group up to G_k has at most $\mathcal{O}(f(n))$ edges with probability at least $1 - 1/e^{\Omega(f(n))}$. We conclude that all infected agents placed in cells of groups up to G_k have $\mathcal{O}(f(n))$ edges with probability at least $1 - \frac{f^q(n)}{e^{\Omega(f(n))}}$ by using the Union Bound (cf. Lemma 4.5).

Now we restrict the probability space to the events in which $\mathcal{O}\left(\frac{|\mathcal{I}(j)|}{(2^k-1)^{\alpha-2}}\right)$ many infected agents choose cells in G_k and each of these agents has $\mathcal{O}(f(n))$ edges. We obtain such an event with probability $1 - e^{-\Omega(f(n))}$. Let X be the random variable counting the number of edges to the nodes in $V' \cap \mathcal{I}_{G_k}(j)$, where $\mathcal{I}_{G_k}(j)$ is the set of infected agents choosing cells in G_k. Further, let X_i do the same but filtered by the knowledge of the edges incident to $x_1, \ldots, x_i$. In other words, $X_i = \mathbb{E}[X \mid \mathcal{F}_i]$, where $\mathcal{F}_i$ is a filter, such that the edges incident to $x_1, \ldots, x_i$ are known. Then the sequence $X_0, X_1, ...$ is a Martingale [65].

Note that we only expose the infected vertices lying in G_k. Given the assumption above, we have $|X_k - \mathbb{E}[X_k]| \leq \mathcal{O}(f(n))$ for any $k \leq \frac{1}{(\alpha-2)} \cdot \log\left(\frac{|\mathcal{I}(j)|}{f^3(n)}\right)$. Then, utilizing the Azuma-Hoeffding Bound (cf. Lemma 4.4), we obtain that

$$\Pr\left(|X_{|\mathcal{I}_{G_k}(j)|} - X_0| \geq \lambda\right) \leq 2 \cdot \exp\left(-\frac{\lambda^2}{2\sum_{k=1}^{t} \mathcal{O}(f^2(n))}\right) \leq \exp\left(-\frac{1}{\kappa}\frac{f^3(n)}{\mathcal{O}(f^2(n))}\right),$$

with $\lambda = \frac{|\mathcal{I}(j)|}{\kappa(2^k-1)^{\alpha-2}} \geq \frac{f^3(n)}{\kappa}$ and $t = |\mathcal{I}_{G_k}(j)|$. Since the expected number of uninfected agents in a cell of G_k is less than $2^{k+1} - 1$, we get $X_0 = \mathcal{O}\left(\frac{|\mathcal{I}(j)|}{(2^k-1)^{\alpha-2}}\right) \cdot (2^{k+1} - 1) = \mathcal{O}\left(\frac{|\mathcal{I}(j)|}{(2^k-1)^{\alpha-3}}\right)$, and the lemma follows. □

Finally, we are ready to prove our main statement. Depending on $\tau \in \Theta(1)$, there exists a configuration of $\kappa, \alpha \in \Theta(1)$, such that an epidemic runs out in $\mathcal{O}\left((\log\log n)^4\right)$ many rounds without infecting a large fraction of the network in the restricted model.

Proof of Theorem 6.2.

The proof consists of three parts. In short, the first part shows that the number of infected agents decreases after $\mathfrak{s}$ many rounds by at least a constant factor with probability at least $1 - 2\mathfrak{s}|\mathcal{I}|^{-1/q}$, where q is some suitable constant and $\tau = \mathfrak{s}$. Recall that $\tau \in \mathcal{O}(1)$. The second part states a result about the oscillating behavior (w.r.t. the number of infected agents) during the entire process, and the third part shows that it is sufficient to consider $\mathcal{O}((\log\log n)^4)$ additional rounds to eliminate the remaining infected agents with probability at least $1 - \log^{-\Omega(1)} n$, provided that the number of informed agents is small enough (although dependent on n). All parts combined imply the validity of the theorem.

To be more specific, we show the following properties:

- For all $j \geq \mathfrak{s} : |\mathcal{I}(j)| \leq {}^{|\mathcal{I}(j-\mathfrak{s})|}/_{c}$ with probability $1 - o(1)$, where $c > 1$ is a constant.
- Suppose that $|\mathcal{I}(i)| \leq \log^q n$. Then in each sequence of $\Theta((\log\log n)^4)$ consecutive rounds starting in round i there are at least $\Theta((\log\log n)^2)$ many rounds j such that $|\mathcal{I}(j)| \leq \log\log\log\log n$ holds with probability $1 - o(1)$.
- Suppose that $|\mathcal{I}(i)| \leq \log\log\log\log n$. Then in each sequence of $\Theta((\log\log n)^4)$ consecutive rounds starting in round i there are at least $\Theta((\log\log n)^2)$ many rounds where all agents of $\mathcal{I}$ spend τ consecutive rounds in otherwise empty cells with probability $1 - o(1)$.

Part 1 Let ν be an upper bound per node on the number of newly infected agents in the network. This bound holds with probability at least $1 - o(|\mathcal{I}(j)|^{-1/q})$ in round j and will be computed later. Note that each infected agent carries the infection further for a constant amount of consecutive rounds $\mathfrak{s}$ only, since $\tau = \mathfrak{s}$. Let a *super-step* be a sequence of $\mathfrak{s}$ consecutive rounds. By definition we know that all agents that were infected at the beginning of a super-step are cured (i.e., they are moved to the set $\mathcal{R}$) after said super-step. Let $j \geq \mathfrak{s}$ be an arbitrary round. Then

$$\begin{aligned} |\mathcal{I}(j)| &\leq |\mathcal{I}(j-\mathfrak{s})|(1+\nu)^{\mathfrak{s}} - |\mathcal{I}(j-\mathfrak{s})| \\ &\leq |\mathcal{I}(j-\mathfrak{s})|\left((1+\nu)^{\mathfrak{s}} - 1\right). \end{aligned} \tag{6.2}$$

To obtain a value for ν, we need to compute an upper bound on the amount of different cells that become occupied by infected agents. We therefore group the cells with respect to their attractiveness. The group G_k contains all cells with attractiveness 2^k up to $2^{k+1} - 1$. Let x_{ij} be the number of agents that are infected by $i \in \mathcal{I}(j)$ in round j and let $X_j := \sum_{i \in \mathcal{I}(j)} x_{ij}$. Further, we assume that $|\mathcal{I}(j)| = \log^q n$, with q being a suitable large constant. Note that we show below that the amount of infected agents does not rise above $\log^q n$ in one round with probability at least $1 - o(\log^{-1} n)$.

Let c be the normalizing constant defined in Section 6.1.1. Note that the expected amount of infected agents accommodated in cells of attractiveness at least $k := 2^{\frac{1}{\alpha-2}\log\left(\frac{|\mathcal{I}(j)|}{\log^3 n}\right)}$ is at most

$$|\mathcal{I}(j)| \cdot \frac{\sum_{k'=k}^{\sqrt[\alpha]{\kappa n}} \frac{c\kappa n}{k'^{\alpha-1}}}{\sum_{i=2}^{\sqrt[\alpha]{\kappa n}} \frac{c\kappa n}{i^{\alpha-1}}} \leq \log^3 n, \tag{6.3}$$

provided α is large enough (although constant). Recall that all agents choose their target location independently. Therefore, Chernoff bounds (cf. Lemma 4.1) imply that Equation 6.3 holds up to a constant factor with probability at least $1 - e^{-\Omega(\log n)}$.

On the other hand, if the attractiveness is larger than $l := 2^{\frac{1}{\alpha-2}\log\left(|\mathcal{I}(j)|\log^{1+\Theta(1)} n\right)}$, then it follows that the expected amount of infected agents in cells with attractiveness at

least l is at most

$$\mu := |\mathcal{I}(j)| \cdot \frac{\sum_{k'=l}^{\sqrt[\alpha]{\kappa n}} \frac{c\kappa n}{k'^{\alpha-1}}}{\sum_{i=2}^{\sqrt[\alpha]{\kappa n}} \frac{c\kappa n}{i^{\alpha-1}}} = o(\log^{-3} n). \tag{6.4}$$

Similar as before, utilizing Chernoff bounds with the parameter δ' being chosen such that $(1+\delta')\mu = 1/2$, we conclude that none of the infected agents chooses a cell with attractiveness above l with probability at least $1 - o(\log^{-1} n)$.

Let $f(n) = \log n$. In the following we differentiate between the first two groups G_1, G_2 and all the other ones in order to compute an upper bound on X_j. According to Lemma 6.2, we know that the number of newly infected agents in G_k for some round j is $\mathcal{O}\left(\frac{|\mathcal{I}(j)|}{(2^k-1)^{\alpha-3}}\right)$ with probability $1 - n^{-\Omega(1)}$. For $k = 1$, however, the situation is different. Let c be the normalizing constant as defined in Section 6.1.1. Then it follows that the amount of cells with attractiveness at least 4 is upper bounded by

$$c\kappa n \cdot \sum_{d=4}^{\sqrt[\alpha]{\kappa n}} 1/d^{\alpha} \leq \frac{c\kappa n}{(\alpha-1)\cdot 3^{\alpha-1}} \leq \frac{2^{\alpha-1}\kappa n}{3^{\alpha-1}}.$$

Consequently, the amount of cells within G_1 grows with rising α, whereas all other cells tend to a $o(1)$-fraction. Now we consider the probability for the nodes in $\mathcal{I}(j)$ to infect at least $|\mathcal{I}(j)|/c''$ many agents, for some constant $c'' > 1$, while being in a cell of group G_1 or G_2. According to Observation 1, this probability is upper bounded by

$$\binom{n - |\mathcal{I}(j)|}{|\mathcal{I}(j)|/c''} \left(\frac{7 \cdot |\mathcal{I}(j)|}{\sum_{d=2}^{\sqrt[\alpha]{\kappa n}} \frac{c\kappa n}{d^{\alpha-1}}}\right)^{|\mathcal{I}(j)|/c''} \leq \left(\frac{7ec''}{\kappa}\right)^{|\mathcal{I}(j)|/c''} \leq \frac{1}{2^{|\mathcal{I}(j)|/c''}},$$

whenever $\kappa \geq 14ec''$.

All statements thus far imply the following properties:

1. the amount of newly infected agents within G_1 or G_2 is upper bounded by $\frac{14e|\mathcal{I}(j)|}{\kappa}$ with probability at least $1 - (2^{-f^q(n)/c''})$,
2. according to Lemma 6.2, the amount of newly infected agents in G_k is bounded by $\mathcal{O}\left(\frac{|\mathcal{I}(j)|}{(2^k-1)^{\alpha-3}}\right)$ with probability $1 - e^{-\Omega(f(n))}$ for each $2 \leq k \leq \frac{1}{(\alpha-2)} \cdot \log\left(\frac{|\mathcal{I}(j)|}{f^3(n)}\right)$,
3. at most $\mathcal{O}(\log^3 n)$ many infected agents choose cells with attractiveness at least k with probability at least $1 - e^{-\Omega(\log n)}$, and
4. none of the infected agents chooses a cell with attractiveness at least l with probability at least $1 - o(\log^{-1} n)$.

Consequently, the amount of newly infected nodes is upper bounded by

$$\begin{aligned}
X_j &\leq \frac{14e|\mathcal{I}(j)|}{\kappa} + \sum_{k=3}^{\frac{1}{\alpha-2}\log\left(\frac{|\mathcal{I}(j)|}{\log^3 n}\right)} \mathcal{O}\left(\frac{|\mathcal{I}(j)|}{(2^k-1)^{\alpha-3}}\right) + \mathcal{O}(\log^3 n)\cdot l \\
&\leq \frac{14e|\mathcal{I}(j)|}{\kappa} + \frac{\Theta(|\mathcal{I}(j)|)}{\alpha-4}\cdot\frac{1}{(2^3-2)^{\alpha-4}} + \mathcal{O}(\log^3 n)\cdot\left(|\mathcal{I}(j)|\log^{1+\Theta(1)} n\right)^{\frac{1}{\alpha-2}} \\
&\leq \frac{14e|\mathcal{I}(j)|}{\kappa} + c'\cdot|\mathcal{I}(j)|\cdot\left(1+\frac{\mathcal{O}(\log^{3+\frac{1+\Theta(1)}{\alpha-2}} n)}{|\mathcal{I}(j)|^{1-\frac{1}{\alpha-2}}}\right) \\
&\leq \frac{14e|\mathcal{I}(j)|}{\kappa} + c'\cdot|\mathcal{I}(j)|\cdot\left(1+\frac{\mathcal{O}(\log^{3+\frac{1+\Theta(1)}{\alpha-2}} n)}{\log^{q(1-1/\alpha-2)} n}\right) \\
&\leq \frac{14e|\mathcal{I}(j)|}{\kappa}\cdot\left(1+\frac{\kappa c'}{14e}(1+o(1))\right) \\
&\leq \frac{28e|\mathcal{I}(j)|}{\kappa},
\end{aligned} \tag{6.5}$$

with probability $1-o(\log^{-1} n)$, whenever q is properly chosen and α large enough such that $c' \leq (14e)/\kappa \cdot 9/10$. Note that we only need to consider attractiveness' up to l, since none of the infected agents will be accommodated in cells with an attractiveness larger than l with probability $1-o(\log^{-1} n)$. In this context c' is an arbitrary small constant depending on α and κ (cf. Inequality 6.5).

Note that we obtain the general formulas, i.e., for the case where $|\mathcal{I}(j)| \neq \log^q n$, by replacing $\log(n)$ with $f(n)$ in the above statements. Further, note that $|\mathcal{I}(j)|$ can be represented by $f^q(n)$ at this point. Then Inequality 6.5 holds with probability $1-o(|\mathcal{I}(j)|^{-1/q})$. Thus, $\nu := \frac{28e}{\kappa}$ with the probability given above.

Consequently,

$$\begin{aligned}
((1+\nu)^{\mathfrak{s}}-1) &= \sum_{k=0}^{\mathfrak{s}}\left(\binom{\mathfrak{s}}{k}1^{\mathfrak{s}-k}\nu^k\right) - 1 \\
&= \sum_{k=1}^{\mathfrak{s}}\binom{\mathfrak{s}}{k}1^{\mathfrak{s}-k}\nu^k \leq \frac{(\nu\mathfrak{s})\frac{(1-\nu)^{\mathfrak{s}-1}}{(1-1/\mathfrak{s})^{\mathfrak{s}-1}}}{(1-\nu)^{\mathfrak{s}-1}} \\
&\leq \nu\mathfrak{s}e^{\frac{\mathfrak{s}-1}{\mathfrak{s}}} \ll 1,
\end{aligned} \tag{6.6}$$

if $\nu\mathfrak{s} \ll 1/e$. Recall that ν depends on α and κ (cf. Inequality 6.5). Hence, setting α and κ large enough in dependence of τ guarantees that $\nu\mathfrak{s} \ll 1/e$. Here the first inequality in Equation 6.6 can be obtained using the following estimation from [50]

$$\sum_{k=1}^{\mathfrak{s}}\binom{\mathfrak{s}}{k}1^{\mathfrak{s}-k}\nu^k(1-\nu)^{\mathfrak{s}-1} \leq \left(\frac{\nu}{1/\mathfrak{s}}\right)^{\frac{1}{\mathfrak{s}}\mathfrak{s}}\left(\frac{1-\nu}{1-1/\mathfrak{s}}\right)^{(1-1/\mathfrak{s})\mathfrak{s}}.$$

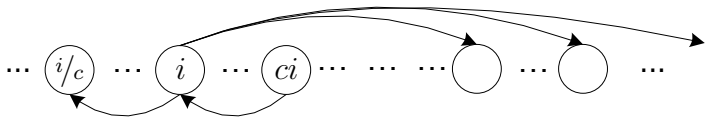


Figure 6.2: *The random walk used in Part 2 of Theorem 6.2, with $c = (\nu \mathfrak{s} e)^{-1}$. Only a small subset of all transitions, indicated by the edges, are depicted above to improve readability.*

Then Equation 6.2, 6.5 and 6.6 imply a decreasing number of infected agents after each super-step by a factor of $\nu \mathfrak{s} e$. Since Inequality 6.5 holds with probability $1 - o(|\mathcal{I}(i)|^{-1/q})$ for an arbitrary round i, the same statement holds for an arbitrary super-step starting in round j with probability at least

$$\begin{aligned}\prod_{i=j}^{j+\mathfrak{s}-1} \left(1 - o\left(|\mathcal{I}(i)|^{-1/q}\right)\right) &\geq 1 - o\left(|\mathcal{I}(j)|^{-1/q}\right)^{\mathfrak{s}} \\ &\geq 1 - 2\mathfrak{s}|\mathcal{I}(j)|^{-1/q} \\ &= 1 - o(1).\end{aligned}$$

Note that $|\mathcal{I}(j)| \leq |\mathcal{I}(j+1)| \leq ... \leq |\mathcal{I}(j+\mathfrak{s}-1)|$ holds, since the infected nodes at the beginning of each super-step remain infected until the end of said super-step.

Part 2 The results so far imply an oscillating behavior with respect to the number of infected agents. That is, the set $\mathcal{I}$ will mainly decrease (cf. Part 1). However, the number of infected agents may also increase. For instance, the set $\mathcal{I}$ increases drastically whenever an infected agent chooses a destination with plenty of uninformed agents. Thus, it remains to compute the probability for such an event to happen. To cope with this kind of behavior, we model the transitions of $\mathcal{I}(i)$ to $\mathcal{I}(i+1)$, for all considerable rounds i, by a random walk. The details can be found below. Intuitively, Part 1 implies a biased random walk towards a decreasing amount of infected agents.

We start by dividing the dissemination process into *phases*. Let us consider each phase separately, and assume that a specific phase begins in round j. A phase usually consists of $\mathfrak{s}$ rounds, and a round is divided into two substeps. In the first substep of round j', we allow each infected agent to transmit the disease to every other agent being in the same cell. In the second substep then all agents that were infected in round $j' - \mathfrak{s}$ are moved to $\mathcal{R}$ and stop transmitting the disease in the subsequent rounds.

A phase starting in some round j ends after $\mathfrak{s}$ rounds if in all these rounds $j' \leq \mathfrak{s}$ it holds that $|\mathcal{I}(j+j') \setminus \mathcal{I}(j)| \leq \nu \mathfrak{s} e |\mathcal{I}(j)|$. This holds with probability at least $1 - 2\mathfrak{s}|\mathcal{I}(j)|^{-1/q}$ (see above). If in some round $j' \leq \mathfrak{s}$ it holds that $|\mathcal{I}(j+j') \setminus \mathcal{I}(j)| > \nu \mathfrak{s} e |\mathcal{I}(j)|$, then this phase ends, and we start with a new phase in the next round.

At this point we are ready to describe the model of the dissemination process as a special random walk. Let $G' = (V', E')$, with $V' = \{\log\log\log\log n, \ldots, \log^q n\}$, be a

directed graph. A node $v \in V'$ corresponds to the case $|\mathcal{I}| = v$. From each v there is a transition to $\max\{v_{\min}, \nu\mathfrak{s}ev\}$ with probability $1-2\mathfrak{s}v^{-1/q}$, where $v_{\min} = \log\log\log\log n$ (all the other transitions are not relevant and can be arbitrary). Figure 6.2 depicts the directed graph corresponding to this random walk. We use the following lemma in order to show that the vertex $v_{\min}$ is visited $(\log\log n)^2$ many times within $\Theta((\log\log n)^4)$ rounds.

Lemma 6.3. *Let $(X_t)_{t=1}^{\infty}$ be a Markov chain using space $\{1, ..., m\}$ and fulfilling the following property:*

- *for any $t \in \mathbb{N}$ there are constants $c_1, c_2 < 1$, such that $\Pr[X_{t+1} \leq c_1 X_t] \geq 1 - X_t^{-c_2}$.*

Let $T = \min\{t \in \mathbb{N} \mid X_t = 1\}$. Then $\Pr[T = \mathcal{O}(\log^2 m)] \geq 1 - m^{-4}$.

Proof. Note that this lemma has certain similarities to the multiplicative drift theorem of [31]. There it is assumed that $\mathbb{E}[X_t - X_{t+1} \mid X_t] = \Omega(X_t)$, and then a result w.r.t. $\mathbb{E}[T]$ is derived. The proof of our lemma uses some of the arguments of Claim 2.9 from [30]. However, there are two main differences. First, the probability of not moving towards the target in [30] (which is m in their case and 1 in our case) is exponentially small w.r.t. the current state $-X_t$, while we only have some polynomial probability in X_t^{-1}. Secondly, in their case the failure in moving towards the target most likely occurs at the other end of the graph (close to 1), whereas in our case a failure mainly occurs at some state close to the target.

We define a round to be decreasing if $X_{t+1} \leq c_1 X_t$. Furthermore, a decreasing round is successful if $X_{t+1} = 1$ after some $X_t > 1$. Let Y be a random variable which denotes the number of consecutive decreasing rounds without reaching a successful one. Note that $t := \log_{1/c_1} m$ is an upper bound on Y, since $X_t \leq c_1^{\log_{1/c_1} m} m = 1$.

Now we divide the Markovian process X_t into consecutive epochs. Every time we fail to have a decreasing round, we start a new epoch. Furthermore, we stop if a round is successful. Now we show that an epoch is successful with some constant probability. Let P_i be the probability that some epoch i contains a successful round. Recall that an epoch i is successful if each round j of i satisfies $X_j \leq c_1 X_{j-1}$. Thus, $c_1^j m$ represents an upper bound of $X_{i'}$ for some round i' (i.e., the jth round of epoch i). Then

$$P_i \geq \prod_{j=0}^{\log_{1/c_1} m - 1} \left(1 - (c_1^j m)^{-c_2}\right) \geq \prod_{j=0}^{\log_{1/c_1} m - 1} \rho^{-(c_1^j m)^{-c_2}} = \rho^{-a},$$

where ρ is a proper constant with $1 < \rho \leq (1 - c_1^{c_2})^{-c_1^{-c_2}}$. To show the upper bound of ρ, we consider the borders of j. With $j = \log_{1/c_1} m - 1$ it holds that $\left(1 - (c_1^j m)^{-c_2}\right) = (1 - c_1^{c_2})$. Consequently, we obtain that $\rho^{-(c_1^j m)^{-c_2}} \leq \left(1 - (c_1^j m)^{-c_2}\right)$ if $\rho \leq (1 - c_1^{c_2})^{-c_1^{-c_2}}$.

Further, for the exponent a we obtain that

$$a = \sum_{j=0}^{\log_{1/c_1} m-1} (c_1^j m)^{-c_2} = \sum_{j=0}^{\log_{1/c_1} m-1} \frac{(1/c_1^{c_2})^j}{m^{c_2}}$$
$$= \frac{1-(1/c_1^{c_2})^{\log_{1/c_1} m}}{(1-1/c_1^{c_2})m^{c_2}} = \frac{m^{c_2}-1}{(1/c_1^{c_2}-1)m^{c_2}} = \Theta(1).$$

Let T' be the first epoch in which we reach a successful round. As in [30], we know that an epoch has such a round with constant probability. Hence, similar as in [30], $\Pr[T' = \mathcal{O}(\log m)] \geq 1 - m^{-4}$. In other words, we reach such an epoch T' after at most $\mathcal{O}(\log m)$ many unsuccessful ones. Since an epoch can only last for $\mathcal{O}(\log m)$ rounds, we conclude that a successful round is reached within $\mathcal{O}(\log^2 m)$ rounds with probability $1 - m^{-4}$. □

According to the lemma above, node $v_{\min}$ in the graph G' is visited at least once within $\Theta((\log\log n)^2)$ many rounds with probability at least $1 - \log^{-4q} n$. Therefore, node $v_{\min}$ is visited at least $\Theta((\log\log n)^2)$ many times within $\Theta((\log\log n)^4)$ rounds with probability $1 - \mathcal{O}\left((\log\log n)^2 \cdot \log^{-4q} n\right)$. Furthermore, if $|\mathcal{I}(j)| = \log^q n$ for some round j, then $|\mathcal{I}(j+\mathfrak{s})| = \nu\mathfrak{s}e|\mathcal{I}(j)|$ with probability at least $1 - 2\mathfrak{s}\log^{-1} n$. Consequently, each sequence of $\Theta((\log\log n)^2)$ consecutive rounds node $v_{\min}$ is visited in contains decreasing rounds only with probability at least $1 - \mathcal{O}\left((\log\log n)^2 \log^{-1} n\right)$. Therefore, we conclude that the above random walk visits $v_{\min}$ at least $(\log\log n)^2$ many times (without visiting $\log^q n$ twice in any two consecutive phases) within $\Theta((\log\log n)^4)$ many rounds with probability at least

$$\left(1 - \mathcal{O}\left(\frac{(\log\log n)^2}{\log^{4q} n}\right)\right) \cdot \left(1 - \mathcal{O}\left(\frac{(\log\log n)^2}{\log n}\right)\right) \geq 1 - o(\log^{-0.99} n).$$

Part 3 Observe that $|\mathcal{I}|$ does not increase to some value larger than $\log^q n$ within one round with probability $1 - o(\log^{-1} n)$ (cf. Inequality 6.5). Furthermore, given that $|\mathcal{I}|$ is at most $\log^q n$, we know that there are $\Theta\left((\log\log n)^2\right)$ many rounds within each sequence of $\Theta((\log\log n)^4)$ consecutive rounds in which the number of infected agents is at most $\log\log\log\log n$ with probability $1 - o(\log^{-0.99} n)$. To show that the disease is eliminated, we consider the probability for the infected agents not to meet anyone else for τ consecutive rounds.

Due to our previous statements, we only consider rounds in which the number of infected agents is at most $\log\log\log\log n$. First, we assign the uninfected agents to the cells. Recall that the probability for an agent to choose a specific cell with constant attractiveness is proportional to $\Theta\left(1/\kappa n\right)$ (cf. Observation 1). Thus, a cell with constant attractiveness remains empty with probability $\left(1 - \Theta\left(1/\kappa n\right)\right)^n = e^{-\Theta(1)}$. With the same argument as in the proof of Theorem 6.1, we conclude that $\Theta(n)$ many cells remain

empty overall. Let $t := (\log\log n)^2$ be the number of so-called phases, where each phase consists of τ rounds. Then an (infected) agent chooses an empty cell for τ consecutive rounds with probability $e^{-\Theta(\tau)}$, and, according to the Union Bound (cf. Lemma 4.5), all agents of $\mathcal{I}$ choose empty cells in τ consecutive rounds with probability $e^{-\Theta(\tau|\mathcal{I}|)}$. Thus, in all t phases there is at least one agent $v \in \mathcal{I}$ that does not choose an empty cell in at least one of the rounds with probability at most

$$\left(1 - \frac{1}{e^{\Theta(\tau|\mathcal{I}|)}}\right)^t \leq \exp\left(-\frac{t}{e^{\Omega(|\mathcal{I}|)}}\right) \leq \log^{-\Omega(1)} n. \tag{6.7}$$

Hence, after $\Theta((\log\log n)^4)$ rounds there was at least one phase in which all agents of $\mathcal{I}$ spent τ consecutive rounds in otherwise empty cells with probability $1 - \log^{-\Omega(1)} n$.

Consequently, the network becomes completely healthy again, i.e., there are no infected agents anymore, after $\Theta((\log\log n)^4)$ rounds with probability $1 - o(1)$. □

6.1.3 Glossary

At this point we list some of the most important definitions we used in the above analysis.

Definitions concerning the network

G	Complete graph representing the topology.
κn	Number of nodes in G, where $\kappa \in \Theta(1)$.
α	The power law exponent used to determine the attractiveness of the cells in G.
d_v	Attractiveness of $v \in V$.

Definitions concerning the agents states

$\mathcal{U}(i)$	Set of uninfected agents in round i.
$\mathcal{I}(i)$	Set of infected/contagious agents in round i.
$\mathcal{R}(i)$	Set of recovered/immunized agents in round i.
τ	Represents an amount of consecutive rounds an individual is contagious in.

6.2 Epidemics in Urban Environments on a National Scope

In Section 6.1 we stated some theoretical insights concerning the spread of epidemics in an urban environment on a small scale. There we used a simple, yet quite realistic, movement model. Now we extend the previous analysis to an experimental one on a large scale, i.e., we present and empirically analyze a dynamic model for the spread of epidemics on a national scope. To be more specific, we investigate the spread of different epidemics in a general simulation software by conducting hundreds of experiments. The tool in use is agent-based, i.e., the individuals (or groups of such) are represented by agents interacting with each other. The environment all epidemics spread in approximates the geography of Germany, in which agents may travel between cities. Within a city the agents interact according to the probabilistic model presented in Section 6.1 in a distributed manner. One of our objectives is to find the right parameters that lead to realistic settings. Our second objective is to evaluate similarities between countermeasure approaches in our model and the real world. We use empirical data for the comparison.

We gain several insights. First, we show that the accuracy of our results in the scenarios we have tested heavily depends on the number of agents in use. This implies a fragile balance between the parameters and the amount of agents, which may lead to problems with respect to the calibration of the system for yet unknown (or largely unexamined) epidemics. The second main result of this section states that one can approximate the effect of some non-pharmaceutical countermeasures (such as school closures or isolation) that are usually adopted if an epidemic outbreak occurs, by setting the parameters properly. This observation is supported by the empirical study of [64]. Interestingly, the right choices of the parameters in our experiments seem to be in line with previous observations in the real world (e.g., the power law exponent seems to be in the range of 2.6-2.9 [39]). To analyze the effect of the countermeasures mentioned above, we integrate the corresponding mechanisms on a smaller scale and verify their impact on a larger scale, too.

6.2.1 Theoretical Model and Algorithmic Framework

In Chapter 6.1 we considered an asymptotic analysis of the spread of a disease in a large urban environment. Now we expand our focus to a much larger scale, including several hundred cities. Hereby the cities are chosen from a list in descending order of their population size. Note that we do not consider cities inhabited by less than one agent on expectation. Consequently, the amount of considered cities is limited by the overall number of agents.

It is intuitively clear that an agent in our model may not only move between locations within a city but between cities as well. Furthermore, due to simplicity, the agents are not categorized (i.e., they do not provide further properties like gender etc.). Note that we are not interested in the evaluation of such details anyway. Due to readability reasons, we present our model with respect to four main components:

1. The Environment on a Large Scale (Inter-City Connectivity)
2. The Environment on a Small Scale (Intra-City Connectivity)
3. The Epidemic Model
4. The Countermeasure Model

The Environment on a Large Scale Let $G(d) = (V, E)$ be a complete graph with $m \in \mathbb{N}$ nodes $V := \{c_1, ..., c_m\}$ and parameter $d := \{d_{c_1}, ..., d_{c_m}\}$, whereas each city $c_i \in V$ is represented by a complete graph G_{c_i} itself. For now we focus on the graph $G(d)$ connecting all cities. The intra-city topology, i.e., the topology on a small scale, is defined below. Further, let $p_{c_i} \in \mathbb{N}_0$ be the real-world population size of c_i and $p := \{p_{c_i}, ..., p_{c_m}\}$. Then we define the attractiveness $d_{c_i} \in \mathbb{R}_{\geq 0}$ of city c_i as

$$d_{c_i} = \frac{p_{c_i}}{\sum_{1 \leq i \leq m} p_{c_i}}.$$

Note that the attractiveness d_{c_i} is proportional to the real-world population size (w.r.t. the whole population) of city c_i.

The above definition implies that $G(d)$ represents the topology on a large scale. Each $c_i \in V$ of this graph corresponds to a city of Germany. However, depending on the size, not every city of Germany is represented by a node in V[3]. The population is represented by $n = \sum_{1 \leq i \leq m} nd_{c_i}$ agents, where the number of agents assigned to city c_i is given by nd_{c_i}. Note that nd_{c_i} is proportional to the cities real-world population.

Let $A_{i,s,t}$ be the event that agent i travels from its current location $s \in V$ to $t \in V$. Further, let p' be the probability that an agent decides to travel at all, and let $dist(s, t)$ be the Euclidean distance between the cities s and t. Then we define the probability that event $A_{i,s,t}$ occurs as

$$\Pr(A_{i,s,t}) = p' \cdot \frac{d_t \cdot (dist(s,t))^{-1}}{\sum\limits_{(s,j) \in E} d_j \cdot (dist(s,j))^{-1}}.$$

[3]The amount of overall agents in use (n) determines how many cities are represented by V. To be more specific, we sort the list of all cities in descending order of their real-world population size. Then, starting from the top, we include the currently considered city c_i to V if and only if the assigned amount of agents to said city is at least 1. The latter amount is given by $n \cdot d_{c_i}$.

Note that the probability for an arbitrary agent i to relocate from s to t depends on: 1.) t's real-world population size, 2.) the distance between s and t, and 3.) whether or not i (randomly) decides to relocate at all.

The Environment on a Small Scale Let $G_{c_i}(d(c_i)) = \{V_{c_i}, E_{c_i}\}$ be a complete graph representing the topology of city $c_i \in V$ with $m_{c_i} = \lceil \kappa_{c_i} \cdot nd_{c_i} \rceil$ many nodes (also called *cells*) and parameter $d(c_i) := \{d_{v_1}, ..., d_{v_{m_{c_i}}}\}$. Note that $\kappa_{c_i} > 0$ is a constant with respect to c_i, which will be specified in the upcoming experiments. Whenever it is clear from the context, we omit the index c_i. Further, note that κ_{c_i} does not affect the amount of agents but the amount of cells only. The attractiveness of each cell $v_i \in V_{c_i}$ within city $c_i \in V$ is given by $d_{v_i} \in \mathbb{R}_{\geq 0}$. Said attractiveness is randomly chosen from the set $\{1, ..., \sqrt[\alpha]{m_{c_i}}\}$ according to a power law distribution with parameter $\alpha > 2$. Note that the constant α may vary in our experiments.

In other words, each city is modeled by a clique of cells. These cells correspond to locations within a city an agent can visit. Consequently, each cell may contain a varying amount of agents (also called individuals) depending on the cell's attractiveness. In the case that an agent decides to stay within its current city, this agent moves to a randomly chosen cell according to the distribution of the attractiveness among the cells. Otherwise, said agent simply chooses its next target city in the graph $G(d)$ according to the model described above. Note that the same procedure is used to determine the first cell an agent is accommodated in after entering a city.

Epidemic Model The amount of susceptible *entities/agents* is given by $n \in \mathbb{N}$. Further, each agent has a status indicating its state of knowledge with respect to the epidemic. This status partitions the set of agents S into three disjoint subsets:

1. $\mathcal{I}(j)$, meaning the set of all infected agents in round j,
2. $\mathcal{U}(j)$, meaning the set of all uninfected (susceptible) agents in round j, and
3. $\mathcal{R}(j)$, meaning the set of all resistant/immunized agents in round j.

Consequently, $S := \mathcal{I}(j) \cup \mathcal{U}(j) \cup \mathcal{R}(j)$. Whenever it is clear from the context, we simply write $\mathcal{I}$, $\mathcal{U}$, and $\mathcal{R}$, respectively. Suppose an uninfected agent i visits a cell in round j that also contains agents of $\mathcal{I}(j)$. Let A be the event that agent i becomes infected under these circumstances. Then

$$\Pr(A) = 1 - (1 - \gamma)^{|\mathcal{I}_j^i|},$$

where $\mathcal{I}_j^i$ represents the set of infected agents accommodated in the same cell as i. We refer to the specific value of γ in the upcoming experiments.

Countermeasure Models The countermeasure models described below are tailored to take advantage of the parameters α and κ. That is, high values of these two parameters represent a high level of countermeasures and vice-versa. With countermeasures applied, individuals avoid places with a large number of persons more often, waive needless tours, and are more careful when meeting other people. While α is mostly responsible for a decreasing number of visitors within a cell (and thus for the avoidance of crowded areas, for example), the total space available for the individuals is determined by κ. That is, more cells are created (and thus available) for the individuals whenever κ increases. Thus, encounters of agents within the same cell will get less frequent. As pointed out in [64], a single countermeasure alone is most likely not sufficient to stop an epidemic. Therefore, we assume a combination of them powerful enough to influence the parameters α and κ to be in place. Note that our countermeasure models apply to each city $c_i \in V$ individually.

Although the model described thus far already provides everything necessary to simulate countermeasures, one aspect is still missing. When do we activate/deactivate the countermeasures and to what extent shall they be applied? We use two different types of countermeasure-models for this purpose:

1. a (multi-tier) *level based approach (LM)* considering the percentage of infected agents in the current round, and

2. a *ratio based approach (RM)* considering the amount of newly infected agents in the current round compared to the one in the round before.

In the following we use α_0 and κ_0 as initial values for α and κ, respectively. Note that the specific values for both countermeasure models are specified in Section 6.2.2.5.

The *level based model (LM)* provides multiple *levels*. Each level defines to what extend the considered countermeasures are applied to counteract the epidemic. That is, for each level we define a combination of parameters α and κ that is used to counteract the epidemic if said level is considered active. A level l in the level based model is considered active if the percentage of infected agents is above a threshold defined by the levels $0, ..., l-1$, i.e., the percentage of infected agents is to high to be effectively counteracted by the levels $0, ..., l-1$. The (complete) graph representing the city c_i is (re-)constructed using the parameters provided by level l at the very beginning of round j, in the case that the active level l in round j for some city c_i differs from the active level of the previous round. Afterwards, all other actions scheduled for round j are executed. The details can be found below.

Let LM_m represent the level based model with $m \in \mathbb{N}$ countermeasure levels $L = \{l_1, ..., l_m\} \cup \{l_0\}$. Further, let $l_i = \{\alpha_i, \kappa_i\}$, for $0 \leq i \leq m$, be the set of parameters α and κ that are applied at level i. Note that the set L contains $m+1$ items. However, countermeasures are used only at the levels $l_1, ..., l_m$, since we define l_0 to be the level without any countermeasures in use. Further, let $T = \{l_0^d, l_0^u, l_1^d, l_1^u, ..., l_m^d, l_m^u\}$, with $l_i^d, l_i^u \in \{-\infty, \infty, 0\%, 1\%, ..., 100\%\}$ for all $0 \leq i \leq m$, be the set of transition points for

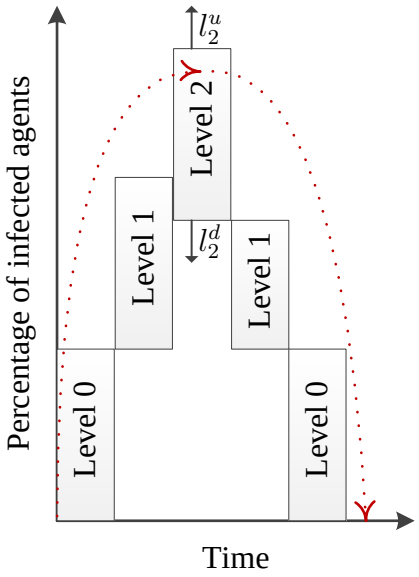
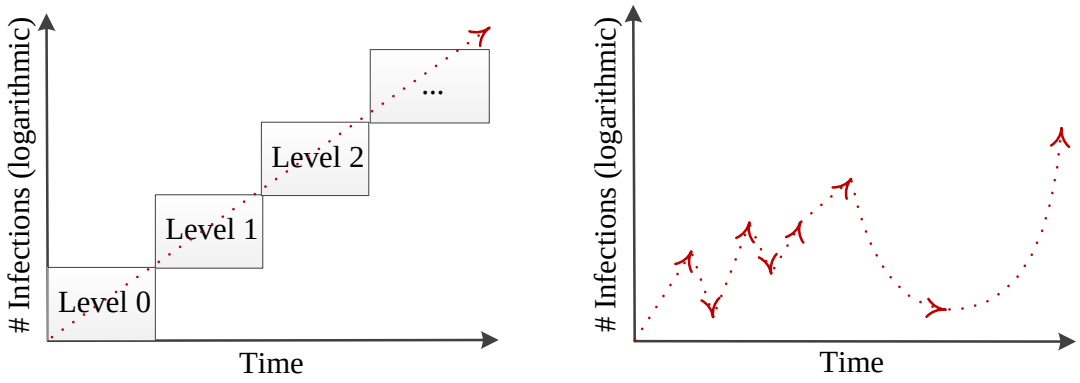


Figure 6.3: *This example demonstrates a possible configuration of LM_2. The dotted line depicts the percentage of infected agents in the city, while the scope of activity of the countermeasure levels is represented by the rectangles.*

all levels. Generally speaking, a transition point of level l_i defines the point in time (with respect to the current percentage of infected agents) where level i is deactivated and level $i+1$ or $i-1$ is activated, respectively. To be more specific, l_i^d defines the point in time (with respect to the current percentage of infected agents) where level i is deactivated and level $i-1$ is activated. Similarly, l_i^u defines the transition point from level i to $i+1$. Note that we do not require $l_i^u = l_{i+1}^d$.

Model LM_2 uses 2 countermeasure levels, for instance. Let us assume that $l_0^d = -\infty, l_0^u = l_1^d = 10\%, l_1^u = l_2^d = 20\%, l_2^u = \infty$. That is, should less than 10 percent of the population of a city c_i be infected, then the complete graph representing c_i is (re-)constructed using the parameters α_0 and κ_0. Should the number of infected individuals grow to at least 10 percent, then α_1 and κ_1 are used to (re-)construct the complete graph representing c_i. In the case that the amount of infections increases to at least 20 percent, both parameters are raised to α_2 and κ_2, respectively. Similarly, both parameters are lowered to the previous values and the complete graph representing c_i is (re-)constructed accordingly whenever the amount of infections falls below the corresponding transition values. Figure 6.3 depicts an example situation.

In contrast, the *ratio based model (RM)* uses a non static approach. Let the set of newly infected nodes of a city c_i in round j be denoted by $\mathcal{I}_{c_i}^*(j)$. Furthermore, let α_j and κ_j denote the corresponding parameters used in round j. In the case that $\frac{|\mathcal{I}_{c_i}^*(j)|}{|\mathcal{I}_{c_i}^*(j-1)|} \geq a$, for



(a) *This example demonstrates a possible configuration of RM. The dotted line depicts the amount of total infections in the city (on a logarithmic scale), while the scope of activity of the countermeasure levels is represented by the rectangles.*

(b) *The oscillating sequence of total infections provokes significant adjustments of α and κ.*

Figure 6.4: *Basic examples for the ratio based model.*

some constant a, we set $\alpha_{j+1} = \alpha_j + z_1$ and $\kappa_{j+1} = \kappa_j + z_2$, where z_1, z_2 are some small constants specified in the corresponding experiment, and (re-)construct the complete graph representing the corresponding city. Similarly, we set $\alpha_{j+1} = \max\{\alpha_0, \alpha_j - z_1\}$ and $\kappa_{j+1} = \max\{\kappa_0, \kappa_j - z_2\}$, whenever $\frac{|\mathcal{I}^*_{c_i}(j)|}{|\mathcal{I}^*_{c_i}(j-1)|} \leq 1/a$. Figure 6.4a depicts an example situation. Since the ratio based model depends on relative numbers rather than absolute numbers, an oscillating amount of infections may provoke a problem. For example, a very low amount of infections may satisfy $\frac{|\mathcal{I}^*_{c_i}(j)|}{|\mathcal{I}^*_{c_i}(j-1)|} \geq a$ in round j, while satisfying $\frac{|\mathcal{I}^*_{c_i}(j+1)|}{|\mathcal{I}^*_{c_i}(j)|} \leq 1/a$ in round $j+1$. Certainly, such an oscillating behavior implies significant adjustments of α and κ over time. Figure 6.4b demonstrates this problem.

6.2.2 Experimental Analysis

As already stated in Section 6.2, the environment approximates the geography of Germany. Our experiments incorporate several hundred cities as visitable areas spread all over the country depending on the number of total agents in use. Each city is assumed to be reachable from any other city. However, an agent may travel at most 1,000 km within a single round. Each round represents an entire day in the real world. Consequently, an agent moving from one city to another has to wait until its destination is reached before it can interact with other agents at said destination.

6.2.2.1 Setup and Implementation

The experiments were mostly conducted on a computation node of the Doppler cluster at the University of Salzburg. This node integrates a quad-socket AMD Opteron 6274 CPU (16 physical cores at 2.1 GHz each) with 512 GB RAM (32 x 16GB DIMMs). An Intel Xeon E5430 CPU (2.66 GHz with 4 physical cores) has been used as a secondary evaluation unit.

Michael Meier implemented the simulation software under the author's supervision. Said software is written in Java and uses an agent-based core architecture, where each agent represents several inhabitants during the simulation. Overall, we are able to conduct large scale simulations with more than 100 million agents in a reasonable time frame. Since such a huge amount of agents may have a significant impact on the performance, several parallelization approaches were implemented to decrease the computational time.

6.2.2.2 Simulations

We present and evaluate our results in the upcoming sections with focus on

1. the impact of the number of agents on the characteristics of the simulated epidemic compared to real-world data, and

2. the impact of non-pharmaceutical countermeasures on the behavior of an epidemic (e.g., social distancing, school closures, and isolation [64]).

Furthermore, we also analyze our parameter settings. Although this is only a short part, our settings seem to coincide with the real-world observations of [39], and thus provide an additional valuable insight. Note that the figures presented below show values based on the real-world population and not the number of agents.

6.2.2.3 Relevance of the Chosen Parameters

Based on real-world observations (e.g., [39]), we chose $\alpha = 2.8$ and $\kappa = 1$ as a starting point for a series of simulations concerning α and κ, respectively. Each plot represents the average value of 50 different experiments utilizing 10 million agents. Note that the corresponding confidence intervals were omitted, since a deviation between the values of the individual experiments was basically non existing. All relevant parameters are explained in Table 6.1.

Figure 6.5a and 6.5b show the impact of α and κ on the behavior of the simulated epidemic. We compare these results to the characteristics of a typical Influenza case

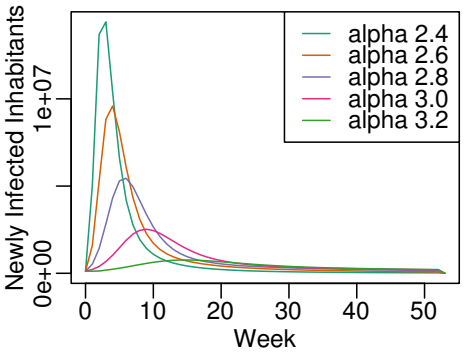
Parameter	
γ	Each agent $v \in V$ becomes infected by each individual $w \in \mathcal{I}$ occupying the same cell at the same time independently with probability γ.
τ	The amount of rounds an agent $v \in \mathcal{I}$ is contagious in.
C_0	Initial amount of cities the infection is being placed in.
$\lvert\mathcal{I}(0)\rvert$	Amount of initially infected agents. These agents are placed in C_0 different cities.
α	The power law exponent used to compute the attractiveness' of the cells within each city.
κ	A multiplicative factor to increase/decrease the amount of cells within a city with respect to the initial amount of agents assigned to said city. With $\kappa = 1$, the total amount of agents within a city equals the number of cells within said city.

Table 6.1: *The parameters used for the evaluation.*

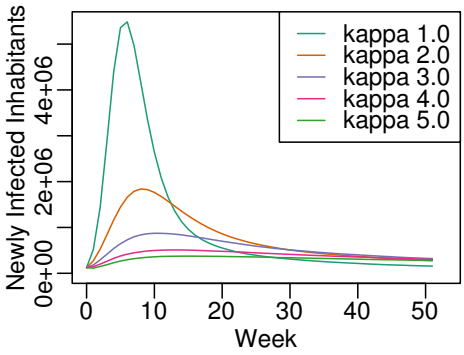
reported by the RKI[4], as depicted by Figure 6.5c. Our results and the data provided by the RKI are presented separately due to readability reasons. All parameters except the subject of the sensitivity analysis were set according to Case 1 (see below). Note that the probability of infection γ was set to a value low enough to model an Influenza epidemic, but high enough to provoke an outbreak. This seemed reasonable due to the (at least to our knowledge) lack of specific values deduced from real-world observations.

Among all five α-values, $\alpha = 2.8$, which has also been obtained in [39] in a different context, represents the best tradeoff between curve similarity and amount of infections. A similar phenomenon can be observed for κ. By increasing κ (including fractional values), the characteristics of the curve (i.e., the amount of infected individuals at the peak vs. the overall number of infections and duration of the outbreak) become less and less accurate. The results indicate that already a minor increase of κ results in a curve that does not follow the characteristics of the real-world observations reported by the RKI anymore.

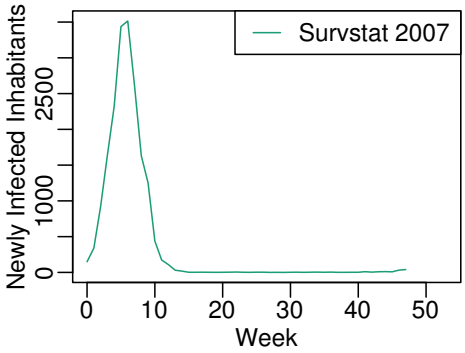
[4]The Robert Koch Institute (RKI) is the central federal institution in Germany responsible for disease control and prevention, and is therefore the central federal reference institution for both, applied and response-orientated research. (Source: http://www.rki.de/EN/Home/homepage_node.html)



(a) $2.4 \leq \alpha \leq 3.2$



(b) $1 \leq \kappa \leq 5$



(c) *Real-world data provided by the RKI for the Influenza epidemic of 2007.*

Figure 6.5: *A sensitivity analysis for varying α (6.5a) and κ (6.5b) values. All other parameters are identical to Case 1 (cf. Section 6.2.2.4). Each result represents the average value of 50 different experiments for the topology of Germany with 10 million agents in use. Due to readability reasons, the real-world data provided by the RKI is shown separately.*

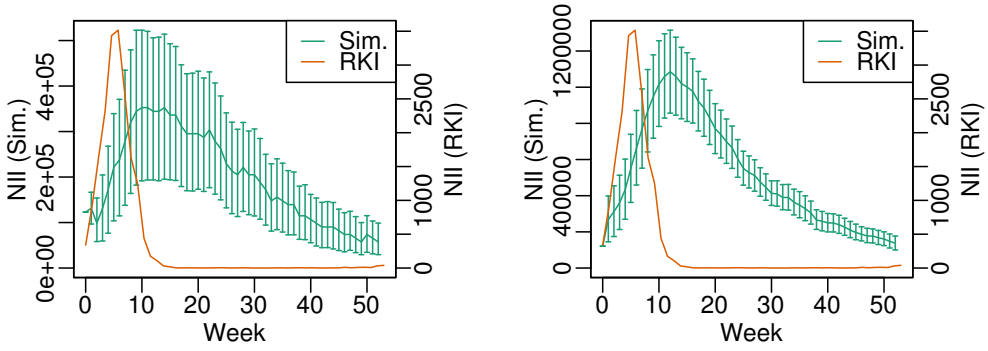
6.2.2.4 Case 1 - Number of Agents

Before we present the results for this case, we first introduce the relevant sources for comparison. In Subcase a) we compare our results to real-world data provided by the RKI SurvStat system [73] for the year 2007. We set the parameters according to reference data provided by the RKI[4] [8] when possible, or to reasonable ones otherwise (cf. Section 6.2.2.3 for more details). In Subcase b) then, an analysis of a fictional epidemic is presented.

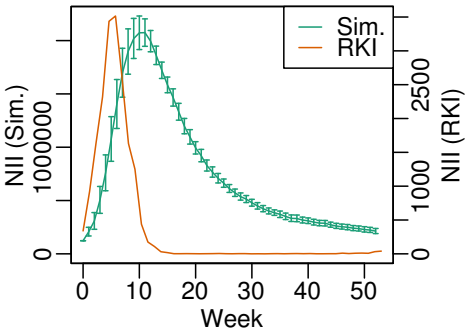
RKI[4]: Basis of Comparison In the following we compare our results to the real-world data provided by the RKI. We use two different data sources for this purpose: 1.) the official report of the Influenza epidemiology of Germany for 2010/2011 [8], and 2.) an online database containing obligatory reports for the Influenza virus called SurvStat [73].

Relevance The data for the SurvStat database and the report of 2010/2011 itself was obtained from more than 1% of all primary care doctors spread all over Germany. This indicates the significance according to international standards. Unfortunately, there are some drawbacks resulting from the type of data ascertainment itself. For example, not every infected person consults a doctor, which implies that the data of the SurvStat system contains only the serious courses of the disease. Nonetheless, these sources provide a valuable tool to obtain insight into the spread and persistence of the Influenza virus in Germany. Further, due to the data's significance, it is possible to estimate the number of infections within certain areas as well. Since the spread of infections in the real world is influenced by factors such as seasonal fluctuations or the virus' aggressiveness, the number of total infections may differ significantly from year to year (cf. [73] for different years). However, the course of the curve usually does not vary. Therefore, we do not focus on absolute values in our simulations, but on the *characteristics* of our results. These characteristics remain, up to some scaling factor, identical over the entire data set provided by the RKI.

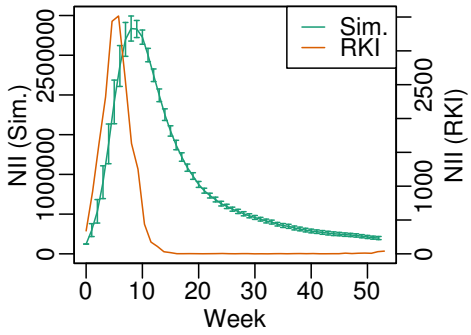
Subcase a) The parameters for this subcase are as follows. We set α to 2.8, κ to 1, γ to 7%, τ to 5 days, the amount of initially infected cities C_0 to 1 (namely Berlin), and the amount of initially infected agents $|\mathcal{I}(0)|$ to 0.0015% of the total amount of agents.



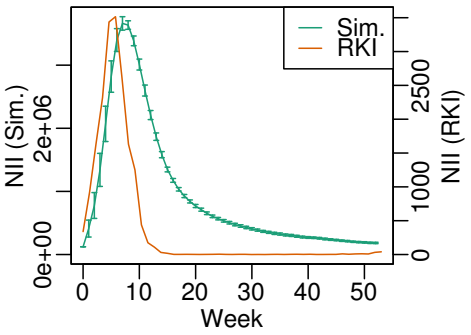
(a) *10,000 Agents* **(b)** *25,000 Agents*



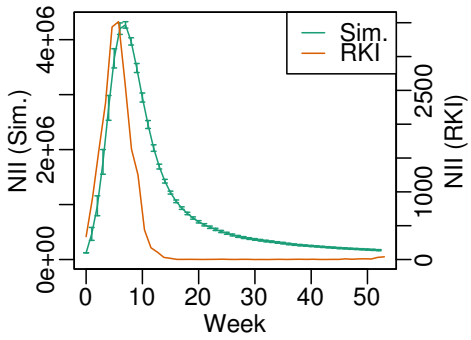
(c) *50,000 Agents*



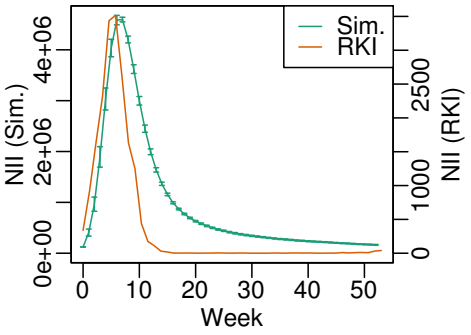
(d) *100,000 Agents*



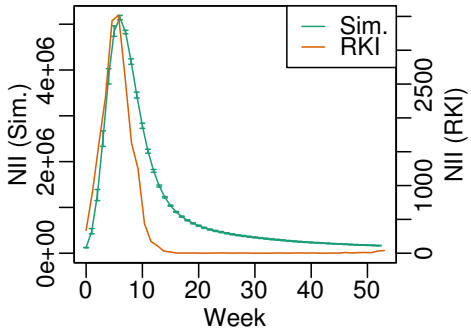
(e) *250,000 Agents*



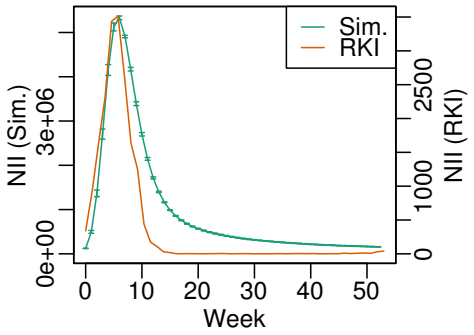
(f) *500,000 Agents*



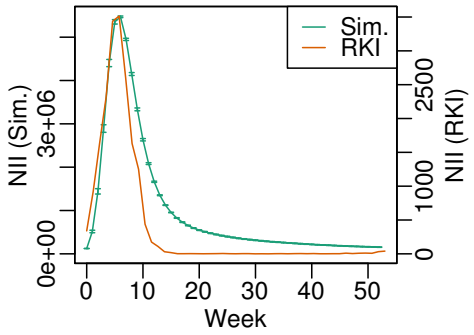
(g) *1,000,000 Agents*



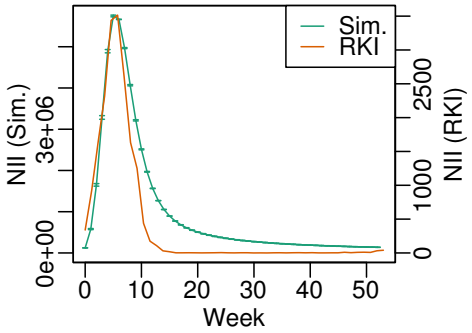
(h) *2,500,000 Agents*



(i) *5,000,000 Agents*



(j) *10,000,000 Agents*



(k) *100,000,000 Agents*

Figure 6.6: *Simulation results for Case 1a) (green) in comparison to real-world data (red) provided by the RKI for a varying amount of agents. The abbreviation* NII *is short for* Newly Infected Inhabitants. *Each result represents the average value of 50 different experiments, where the reliability of each value is indicated by the corresponding confidence interval (*95%*).*

Each obtained plot represents the average value of 50 different experiments. Figure 6.6 shows the results for the first subcase. Here the red curve represents the real-world data provided by the RKI for the year 2007, while the green curve represents our simulation results. Note that both curves vary significantly in terms of absolute numbers. However, this is not our focus here. Due to the level of abstraction in our model, and since the RKI data contains reported cases only (see above), the absolute numbers do not coincide. Additionally, as stated above, the data provided by the RKI also differs significantly (in terms of absolute numbers but not the disease characteristics) from year to year (cf. [73]). Therefore, we focus on the course of the disease and the resulting characteristics of the plotted curves. Note that we moved the original position of the epidemic outbreak contained in the RKI data to the origin to create a comparable situation. Further, note that this relocation does not have any influence on the data or the evaluation itself, since the time of an outbreak varies in the real world as well (cf. [73]) and the beginning of the disease is not important to us.

Obviously, the more agents are used, beginning from Figure 6.6a up to Figure 6.6k, the more the curve characteristics converge. Moreover, the accuracy of each experiment

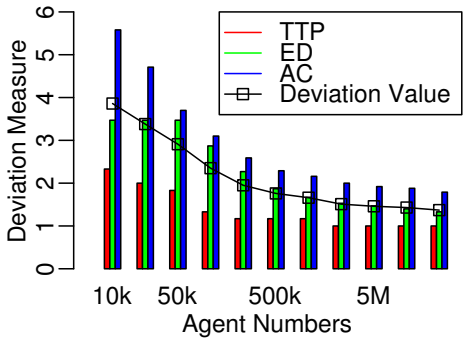


Figure 6.7: *A visual representation of Table 6.2.*

increases as well (cf. the confidence intervals in Figure 6.6). With at least $500,000$ agents in use, both curves become similar.

We define three measures to obtain a mathematical evaluation:

- the time to peak (TTP),
- the epidemic duration (ED), and
- the area of the curve (AC).

The time to peak describes the week with the maximum amount of newly infected agents of the corresponding curve. The area of the curve is defined as the sum of the area between the origin and the endpoint EP (defined by the epidemic duration). Finally, the endpoint of the epidemic duration is set to the week in which only a minor amount of new infections are observed. Let round j be the last round of the simulation. Then minor infections start in the earliest round i, such that for all rounds $i \leq i' \leq j$ it holds that the amount of newly infected agents does not exceed 9% of the maximum value. The latter value was deduced from the data provided by the RKI. Additionally, we compute a global deviation value utilizing the formula $\frac{1}{3} \cdot \mathrm{TTP} + \frac{1}{3} \cdot \mathrm{AC} + \frac{1}{3} \cdot \mathrm{ED}$. For simplicity we consider each individual measurement uniformly weighted.

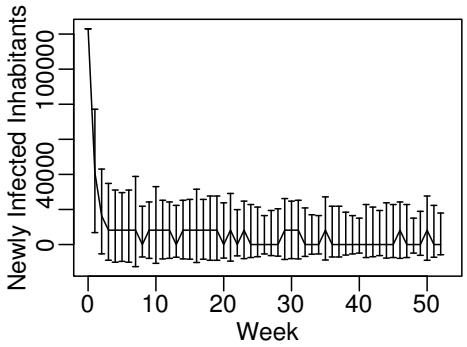
The results, which confirm our previous observations, are shown in Table 6.2 and, in a more graphical manner, in Figure 6.7. Note that the numbers given in Table 6.2 represent the ratio between the value obtained in our simulations and the value provided by the RKI. For example, a value of 4.00 for TTP implies that the time to peak in our case divided by the time to peak in the real-world data is 4. Further, note that we used a properly scaled version of our results for this comparison (cf. Figure 6.6).

All obtained results and previous statements imply a fragile balance between the accuracy, the parameter setting, and the amount of agents in use.

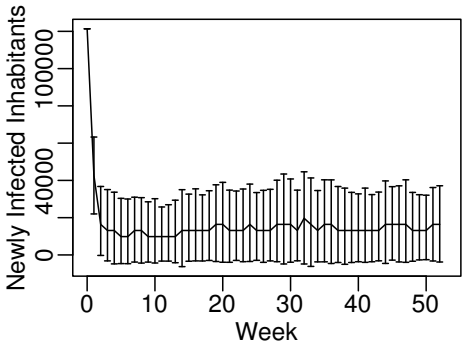
Agents	Deviation			Deviation
	TTP	*ED/EP (week)*	*AC*	Value
10 K	2.33	3.53/53	5.58	3.81
25 K	2.00	3.53/53	4.71	3.41
50 K	1.83	3.47/52	3.70	3.00
100 K	1.33	2.87/43	3.10	2.43
250 K	1.17	2.27/34	2.59	2.01
500 K	1.17	1.87/28	2.29	1.77
1 M	1.17	1.67/25	2.16	1.66
2.5 M	1.00	1.53/23	2.00	1.51
5 M	1.00	1.47/22	1.92	1.46
10 M	1.00	1.40/21	1.88	1.43
100 M	1.00	1.33/20	1.79	1.37

Table 6.2: *This table shows a quantitative comparison of the experiments conducted in Case 1a) with respect to the data provided by the RKI. The results refer to the following properties: the time to peak (TTP), the epidemic duration (ED), and the area of the curve (AC). Here the area of the curve (AC) starts at the origin and ends at the endpoint (EP) defined by the epidemic duration (ED). All of the above values are given as a relative deviation with respect to the data provided by the RKI, whereas the deviation value itself is computed utilizing the formula:* $\frac{1}{3} \cdot \mathrm{TTP} + \frac{1}{3} \cdot \mathrm{AC} + \frac{1}{3} \cdot \mathrm{ED}$.

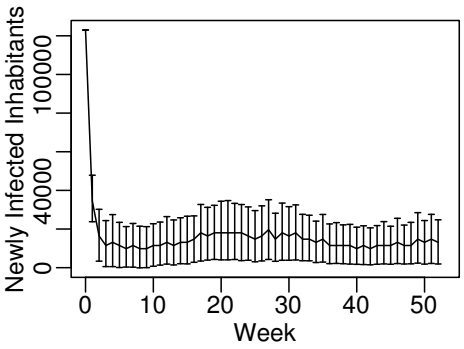
Subcase b) In contrast to Subcase a), the epidemic we focus on here is a fictional one. We set α to 2.8, κ to 1, γ to 1.5%, τ to 5 days, the amount of initially infected cities C_0 to 1 (namely Berlin), and the amount of initially infected agents $|\mathcal{I}(0)|$ to 0.0015% of the total amount of agents.



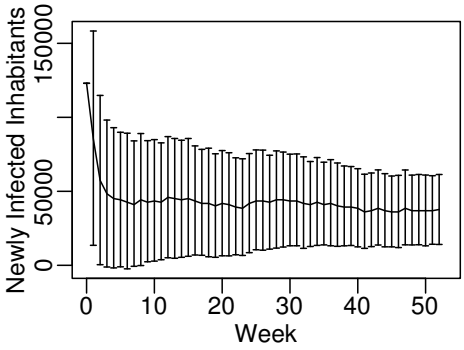
(a) *10,000 Agents*



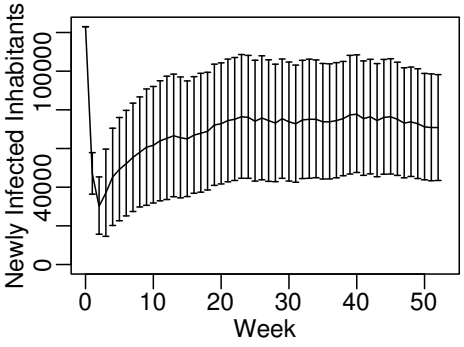
(b) *25,000 Agents*



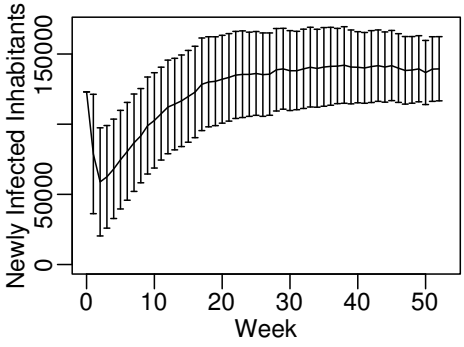
(c) *50,000 Agents*



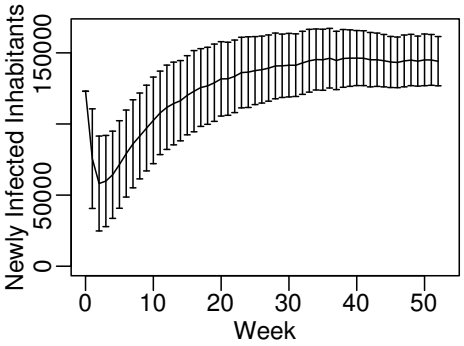
(d) *100,000 Agents*



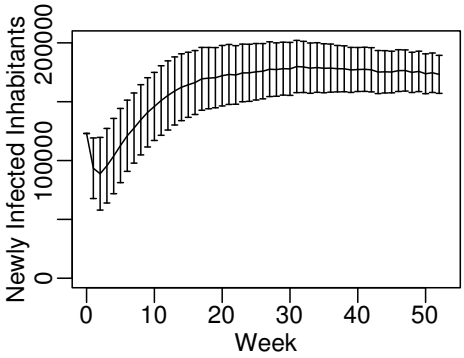
(e) *250,000 Agents*



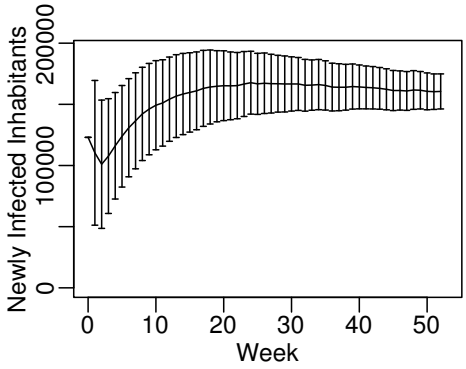
(f) *500,000 Agents*



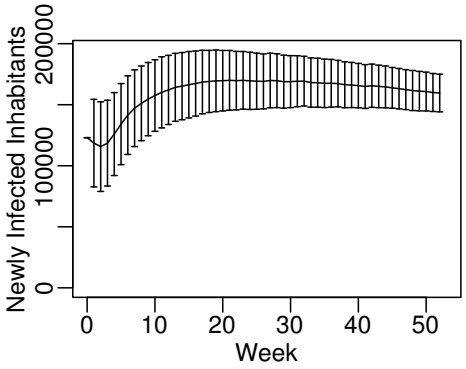
(g) *1,000,000 Agents*



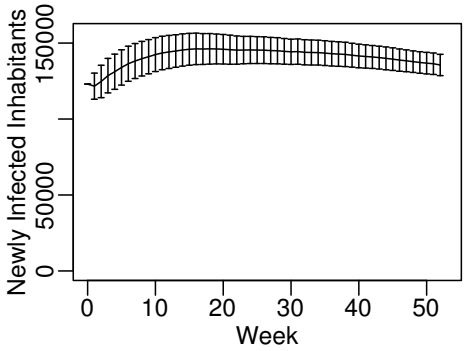
(h) *2,500,000 Agents*



(i) *5,000,000 Agents*



(j) *10,000,000 Agents*



(k) *100,000,000 Agents*

Figure 6.8: *Simulation results for Case 1b) for a varying amount of agents. Each result represents the average value of 50 different experiments. The reliability of each value is indicated by the corresponding confidence interval (*95%*).*

Note that we chose γ relatively low for this subcase in order to emphasize the emerging problem. The figures show a significantly varying course of the disease, which even exceeds the distortion of Subcase a). One can identify multiple curve characteristics (cf. Figure 6.8). To be more specific, we observe:

1. the *extinction of the disease almost instantly* after the outbreak (cf. Figure 6.8a, Figure 6.8b, and Figure 6.8c),

2. followed by a completely *unpredictable behavior* including an increasing as well as a decreasing influence of the disease (cf. Figure 6.8d, and Figure 6.8e),

3. up to a relatively *stable spreading of the disease* (cf. Fig. 6.8j and 6.8k).

Since no other parameter was altered, only the varying amount of agents could have had such a strong influence on the results.

In combination both subcases indicate a significant influence of the amount of agents in use on the spreading process itself. One main goal in such simulations is to gain a speedup by using a lower number of agents while keeping the collectible data (such as

Level	LM_1				LM_2				LM_3			
	l_i^u	l_i^d	α_i	κ_i	l_i^u	l_i^d	α_i	κ_i	l_i^u	l_i^d	α_i	κ_i
0	6%	$-\infty$	2.8	1.0	4%	$-\infty$	2.8	1.0	2%	$-\infty$	2.8	1.0
1	∞	4%	3.3	1.2	6%	2%	3.1	1.1	4%	1%	3.0	1.0
2	-	-	-	-	∞	4%	3.3	1.2	6%	2%	3.1	1.1
3	-	-	-	-	-	-	-	-	∞	4%	3.3	1.2

Figure 6.9: *Parameters for the level based countermeasure models used in Case 2.*

RM Model				
a	z_1	z_2	α_0	κ_0
2	0.2	0.1	2.8	1.0

Figure 6.10: *Parameters for the ratio based countermeasure model used in Case 2.*

the number of infections) identical in terms of accuracy. However, our results show that this can be a very fragile approach, which may not lead to accurate results. Note that we did not explicitly tune the settings for the different amount of agents, since such a possibility may not be given for unknown or (largely) unexamined diseases. However, to some extent, this is implicitly considered by our model due to the dependency of the amount of cells in a city and the amount of agents in use.

6.2.2.5 Case 2 - Non-Pharmaceutical Countermeasures

Now we extend our analysis to incorporate non-pharmaceutical countermeasures such as school closures and social distancing. Again we use a fictional epidemic to amplify the impact of the countermeasures in use. Note that the increase of γ to 12% implies a way faster spreading progress. Consequently, the application of countermeasures (or the lack thereof) results in a higher impact on the amount of new infections compared to Case 1. Further, we set τ to 5 days, the amount of initially infected cities C_0 to 1 (namely Berlin), and the amount of initially infected agents $|\mathcal{I}(0)|$ to 0.00075% of the total amount of agents (to compensate the higher γ at the beginning). All relevant parameters regarding the countermeasure models, which where chosen according to our best judgment, can be found in Table 6.9 and 6.10.

We assume that non-pharmaceutical countermeasures basically affect the parameters α and κ, since the individuals will most likely avoid places with a large number of persons, waive needless tours, and be more careful when meeting other people. Although one of these modifications alone is most likely not sufficient [64], we assume a combination of them powerful enough to influence the parameters α and κ to be in place. Obviously, due to the lack of publicly available data, we cannot compare our results to medical research incorporating the effects of different epidemics on the human body with varying countermeasures in use. We use the work of Markel et al. [64] for this purpose instead.

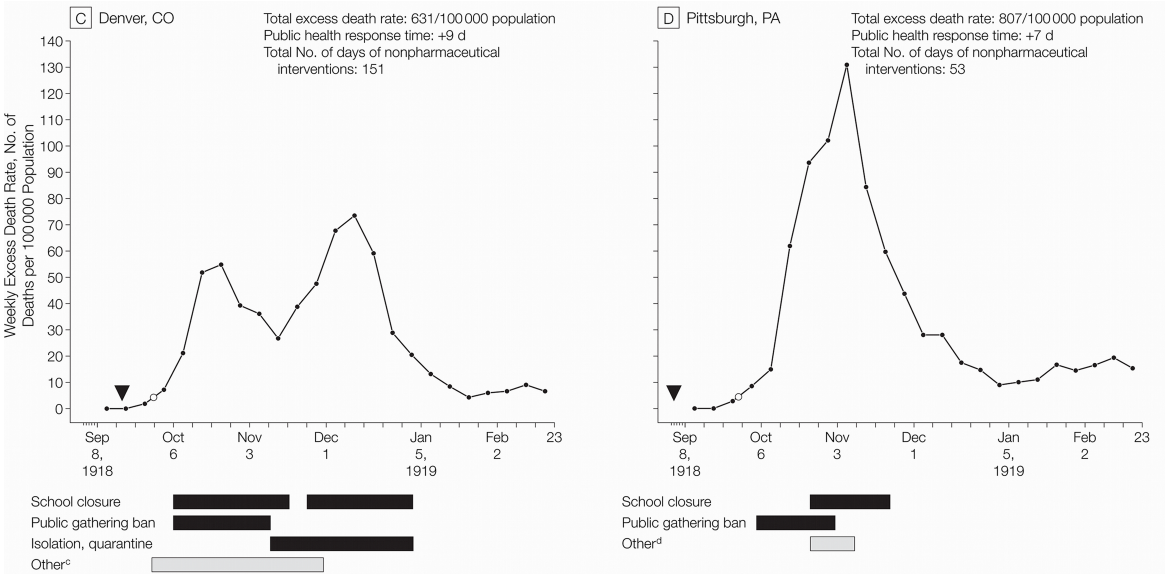


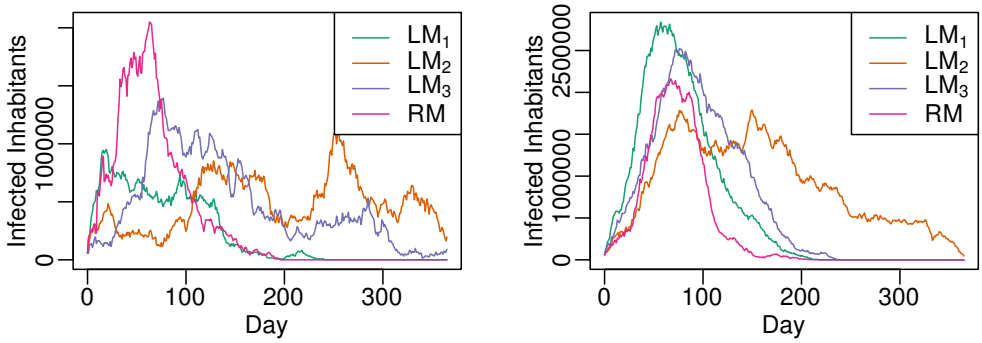
Figure 6.11: *Weekly excess death rates from September 8, 1918, through February 22, 1919 [64, Figure 3].*

Especially Figure 6.11 is of particular interest to us. There the correlation between the application of countermeasures and a decreasing amount of new infections (and vice versa) given implicitly by the excess death rate is well highlighted. We observed an identical effect in our simulations (cf. Figures 6.13, 6.14, 6.15, and Table 6.3). Note that different combinations of countermeasures used in [64] entail different kinds of impact on the death rates. In contrast, we do not focus on specific combinations but on sufficient ones to actually achieve an immediate impact on the epidemic.

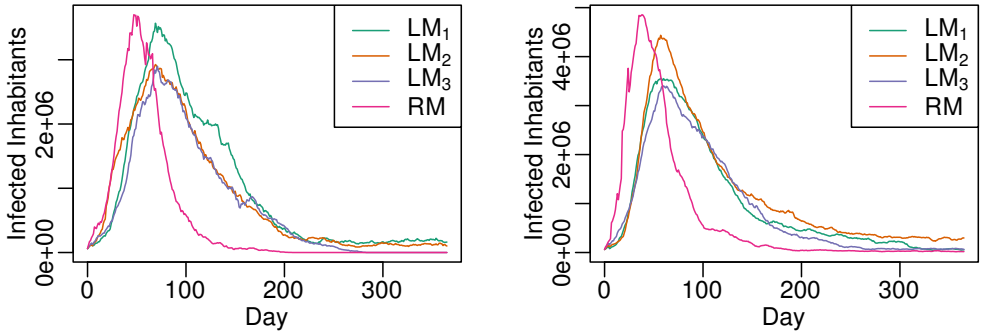
As already described in Section 6.2.1, we use two different countermeasure approaches: 1.) the level based (LM_1, LM_2, and LM_3), and 2.) the ratio based (RM), respectively. Note that the parameters used by the different level based models at their highest level are identical. Therefore, the main advantage provided by the additional levels is the ability to adjust the parameters in a more fine-grained way. Both approaches use different mechanisms and parameters to achieve the embankment of the epidemic. Recall that all transition points in the level based model are chosen w.r.t. the ratio between the amount of currently infected individuals and the population size of the city. Figure 6.12 compares all models to each other on the national scope, i.e., by comparing the overall amount of infected inhabitants in the entire country.

Agents	Infected Agents *(in percent)*				
	$NOCM$	LM_1	LM_2	LM_3	RM
10 k	55.45	40.36	24.38	25.26	31.93
25 k	58.57	16.72	1.56	15.18	33.43
50 k	76.20	13.49	21.50	28.40	42.51
100 k	75.08	11.75	−0.29	16.50	27.88
250 k	81.37	11.13	18.95	16.38	40.30
500 k	85.77	18.05	16.15	20.04	39.26
1 M	86.12	1.84	14.53	12.16	37.41
2.5 M	88.22	12.47	14.66	13.42	35.47
5 M	87.39	11.33	8.50	8.90	29.81
10 M	88.06	8.0	9.30	9.15	32.08
100 M	90.13	2.0	2.82	5.72	32.84

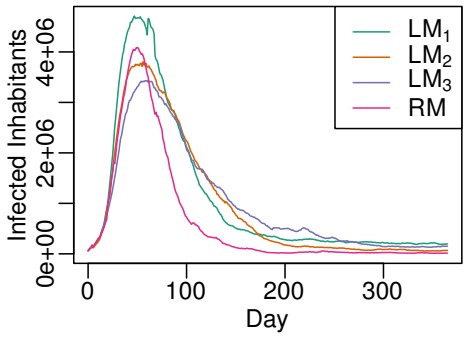
Table 6.3: *This table shows the percentage of the overall infected inhabitants (Case 2) with respect to a specific countermeasure model. Column NOCM represents the percentage of overall infections without any countermeasures in use, whereas the other columns state the gain compared to the case without countermeasures for the corresponding level based and ratio based model, respectively.*



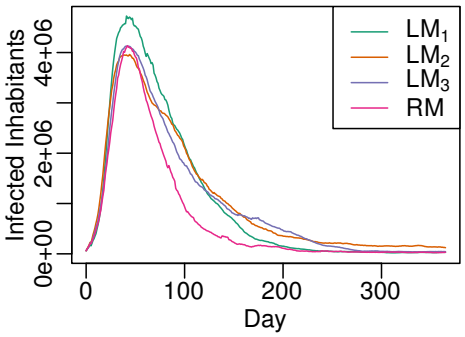
(a) *10,000 Agents* **(b)** *25,000 Agents*



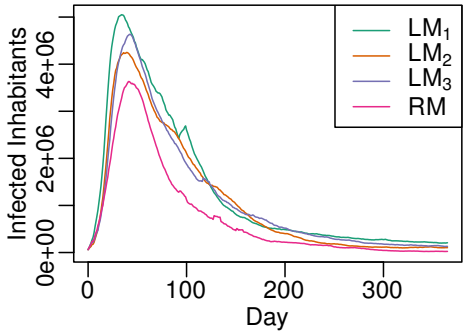
(c) *50,000 Agents* **(d)** *100,000 Agents*



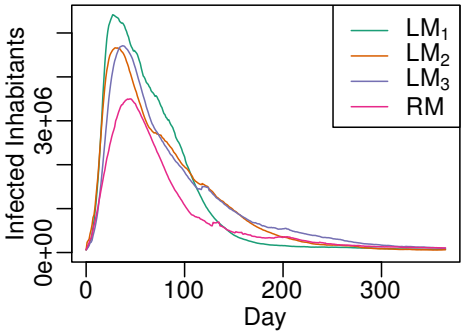
(e) *250,000 Agents*



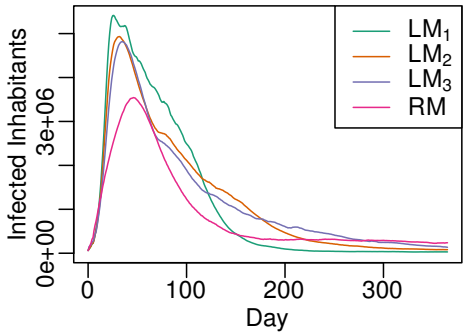
(f) *500,000 Agents*



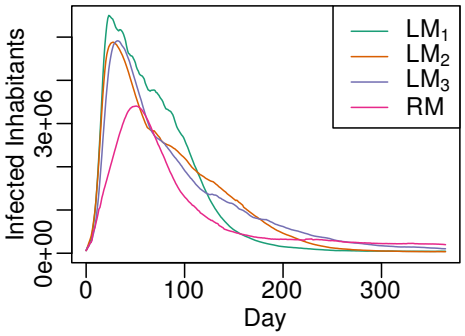
(g) *1,000,000 Agents*



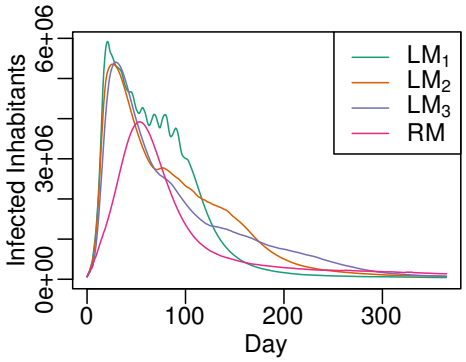
(h) *2,500,000 Agents*



(i) *5,000,000 Agents*



(j) *10,000,000 Agents*



(k) *100,000,000 Agents*

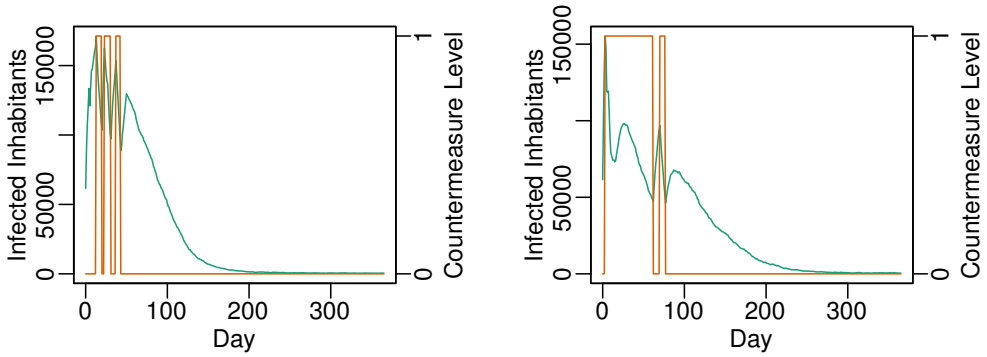
Figure 6.12: *Comparison of different countermeasure models on a large scale. Each curve depicts the situation for the corresponding countermeasure model in all considered cities of Germany combined.*

Although the level based approach is completely different compared to the ratio based approach, the achieved results are similar. However, the overall increase of α and κ by the ratio based approach may be noticeably higher, especially if a large number of agents is used (cf. Figure 6.15d, for example). That is, while all LM-models use $\alpha \leq 3.3$, the RM-model goes way above 4. This implies that the LM-models are more cost efficient, since both α and κ are kept lower and therefore less effort is needed to achieve and maintain said values.

Additionally, to be able to compare our results to the findings in [64], we examine the following three cities in more detail:

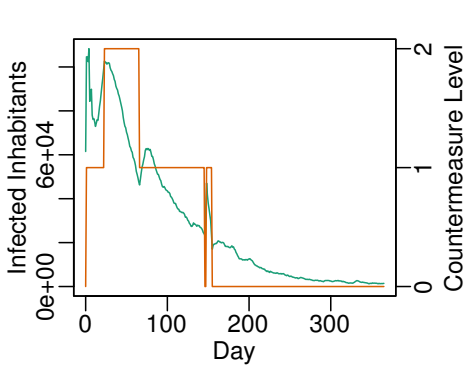
- Berlin (BER, population: ≈ 3.5 million in 2012)
- Hamburg (HAM, population: ≈ 1.8 million in 2012)
- Paderborn (PB, population: $\approx 147,000$ in 2012)

Figures 6.13, 6.14, and 6.15 represent a composition of some interesting results for each city and model in use. Note that the red curve in these figures represents the countermeasure level for the LM models at the corresponding time, and indicates the number of times z_1, z_2 have increased α, κ in the RM model. Recall that level j in the RM-model implies that $\alpha_i = \alpha_0 + \sum_{k=1}^{j} z_1$ and $\kappa_i = \kappa_0 + \sum_{k=1}^{j} z_2$ in round i.

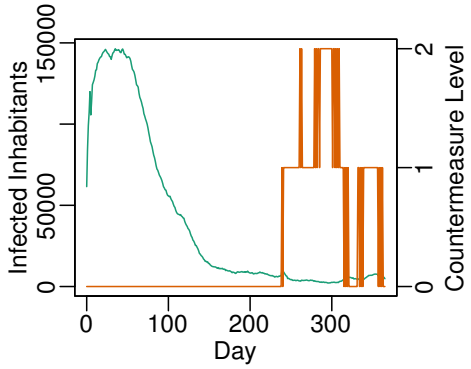


(a) *Model LM_1, 10 million agents*

(b) *Model LM_2, 10 million agents*

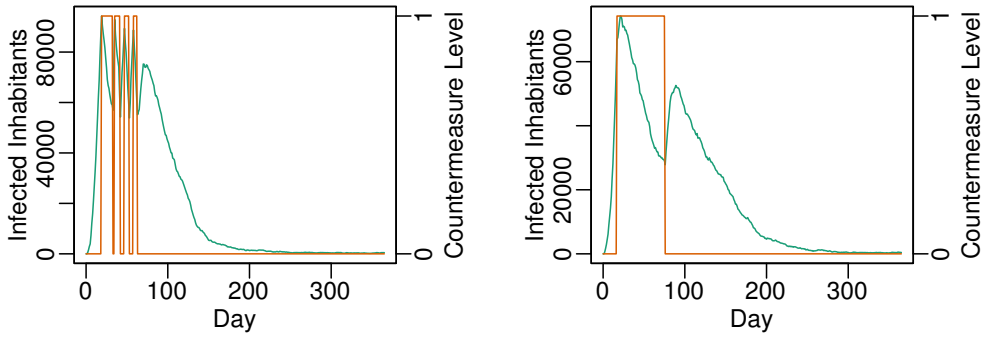


(c) *Model LM_3, 10 million agents*



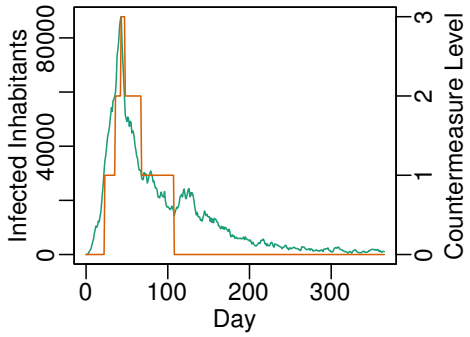
(d) *Model RM, 10 million agents*

Figure 6.13: *Example results for BER in Case 2. The green curve (Y-axis on the left side of the graph) represents the amount of infected inhabitants in the corresponding city, while the red curve (Y-axis on the right side of the graph) indicates the activated countermeasure level with respect to the countermeasure model in use. The number of agents refers to the total amount of agents used in the experiment.*

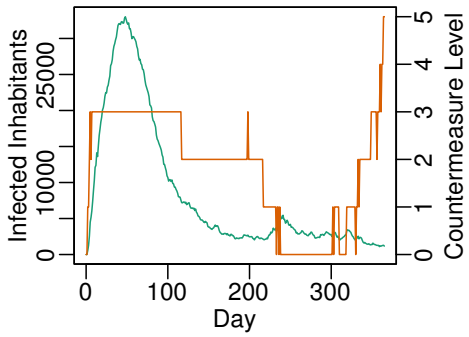


(a) *Model LM_1, 10 million agents*

(b) *Model LM_2, 10 million agents*

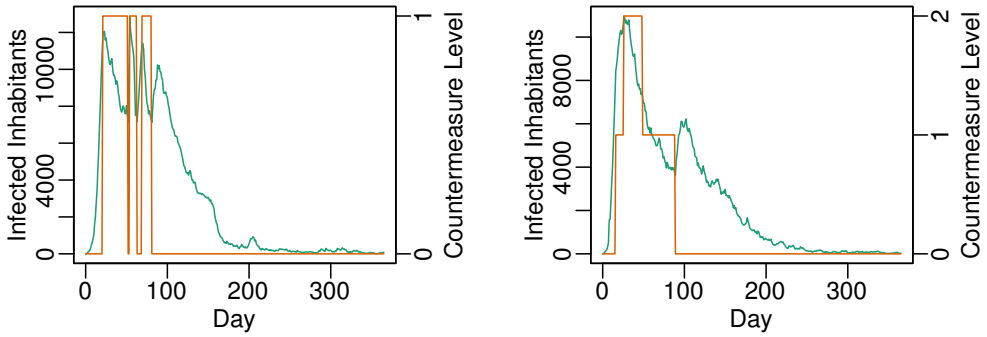


(c) *Model LM_3, 1 million agents*



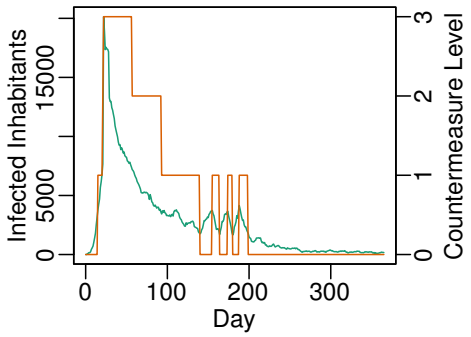
(d) *Model RM, 10 million agents*

Figure 6.14: *Example results for HAM in Case 2. The green curve (Y-axis on the left side of the graph) represents the amount of infected inhabitants in the corresponding city, while the red curve (Y-axis on the right side of the graph) indicates the activated countermeasure level with respect to the countermeasure model in use. The number of agents refers to the total amount of agents used in the experiment.*

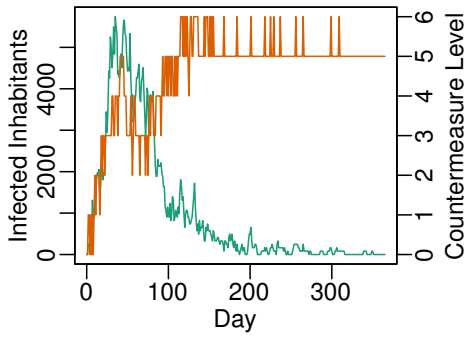


(a) *Model LM_1, 10 million agents*

(b) *Model LM_2, 10 million agents*



(c) *Model LM_3, 10 million agents*



(d) *Model RM, 1 million agents*

Figure 6.15: *Example results for PB in Case 2. The green curve (Y-axis on the left side of the graph) represents the amount of infected inhabitants in the corresponding city, while the red curve (Y-axis on the right side of the graph) indicates the activated countermeasure level with respect to the countermeasure model in use. The number of agents refers to the total amount of agents used in the experiment.*

Our results confirm the impact of different countermeasures observed in the real world [64]. Compared to Figure 6.11, our simulations show a similar behavior (i.e., more than one peak during the epidemic). It is easy to see that the countermeasures presented in [64] directly influence the course of the epidemic. The same property can be observed in our results (cf. Figure 6.13c, 6.14b, and 6.15c). There we observed that the number of infections increases or decreases depending on the active countermeasure level.

Furthermore, we observe that small adjustments of the two parameters α and κ entail a significant impact on the number of overall infections (cf. Table 6.9, Table 6.10, and Table 6.3). Among others, we already proved that even a very aggressive epidemic (γ equal to 100%) will affect no more than a polylogarithmic number of the population if the power law exponent and κ are assumed to be some suitable large constants. Our findings now back up these observations.

In conclusion, we showed the impact of different countermeasures on the behavior of a population (and therefore on the dissemination of an epidemic itself) w.r.t. our model. Although some complexity of the real world is lacking, the similarities to real-world observations are significant. By choosing settings for the environment, and therefore implicitly the individuals' behavior, based on real-world observations (cf. Section 6.2.2.3), relatively mild countermeasures were sufficient to embank or at least significantly suppress an outbreak. Essentially the same properties were already observed in the real world (cf. [64]). This underlines the importance of behavioral and environmental models based on power law distributions.

6.2.3 Glossary

At this point we list some of the most important definitions we used in the above analysis.

In our model there are only three main variables which can vary. One is α, which governs the probability distribution of the attractiveness of different locations within a city. Another is κ, which manipulates the total number of cells within a city, i.e., the amount of locations the agents may visit, as a multiplicative factor. Additionally, the time until an infected individual is cured again can also vary. That is, a time period τ is assigned to the epidemic, which means that an individual is contagious for τ consecutive rounds only. With τ very large (i.e., $\omega(\log n)$), we essentially obtain a model without any recovery.

Definitions concerning the environment

n	Total amount of agents.
α	The power law exponent used to determine the attractiveness of each cell.
$G(d)$	Complete graph representing the topology on a large scale, i.e., connecting all cities.
$G_{c_i}(d)$	Complete graph representing the topology on a small scale, i.e., representing the topology within city c_i.
c_i	Represents the ith city.
d_{c_i}	Attractiveness of city c_i.
d_{v_i}	Attractiveness of cell v_i, whereas $v_i \in V_{c_i}$ for some c_i.
p_{c_i}	Real-world population size of city c_i.
m_{c_i}	Total amount of cells within city c_i.

Definitions concerning the epidemic model

$\mathcal{U}(i)$	Set of uninfected agents in round i.
$\mathcal{I}(i)$	Set of infected/contagious agents in round i.
$\mathcal{R}(i)$	Set of recovered/immunized agents in round i.
τ	The amount of consecutive rounds an individual is contagious in.

Definitions concerning the countermeasure models

LM_m	Abbreviation for the level based model with m countermeasure levels.
T	The set of transition points for the corresponding level based model.
l_i^d, l_i^u	Transition point from level i to $i-1$ and i to $i+1$, respectively.
RM	Abbreviation for the ratio based model.
z_1, z_2	Constants to adjust α and κ with respect to a specific round j in the ratio based model.
a	Constant to determine whether or not an adjustment of α and κ in the corresponding round for the RM-model is in order.

Chapter 7

Conclusion and Future Work

In this thesis we considered different types of dissemination processes in multiple types of networks. We divided the types of dissemination processes into information dissemination processes eager to spread a (malicious) piece of data within the network, and epidemiological processes struggling to survive by infecting new hosts. Obviously, due to our lifestyle, such processes are an inherent part of our daily lives in various kinds of ways and will only become more prominent in the future.

For instance, we rely more and more on fast advancing technologies in order to keep up with the constantly increasing pace of life. As a consequence, we as individuals become part of a fast growing and well-connected world. Consequently, studying technological networks, i.e., networks created and maintained mainly due to artificial devices (such as laptops or smartphones), and social networks, i.e., networks mainly driven by human behavior and actions, seems a promising endeavor. In this thesis we studied the dissemination of a (malicious) piece of information in static networks, i.e., in networks that do not change much over a large period of time, and the spread of epidemics in, due to the individuals' movements, highly dynamic environments. Note, however, one can easily transfer the results for epidemiological processes to information dissemination processes (and vice versa), since both types of dissemination processes have a lot in common.

All static networks we examined share at least the following property: the power law degree distribution. We showed that such networks provide the possibility for fast information dissemination, and that already rudimentary strategies are sufficient to counteract a (malicious) dissemination process by using another (similar) dissemination process.

In the case of dynamic networks, we took a similar yet different path though. In our theoretical analysis we examined one (large) urban environment providing numerous locations the individuals may visit. The similarities of our environment and the urban topologies we live in then ultimately lead to comparable outcomes: aggressive and deadly epidemics (such as the pestilence, for example) are not likely to wipe out the population entirely, and countermeasures used in the real world have a comparable impact in our model. As a second step, we analyzed such epidemic outbreaks empirically

on a large scale and observed significant similarities in the course of the disease between real-world epidemics and our model with and without the aid of countermeasures.

In the following we give a more detailed summary of this thesis, which also motivates the recommendations for future work we conclude with.

7.1 Summary in More Detail

The above summary contains only the most important intuitions and basic ideas to solve the described problems. At this point we complement these intuitions by a more detailed high level description of the approaches we used to obtain our results.

In **Section 5.1** we used an adaptation of the Random Phone-Call model on random power law graphs with n nodes. Our graphs were chosen uniformly at random from the space of all (simple) graphs with power law degree distribution, where the smallest degree is in $\Theta\left((\log\log n)^2\right)$. In each round every node was allowed to call on ρ different neighbors chosen uniformly at random, and to establish communication channels to these nodes. We showed that this algorithm completes broadcasting in time $O(\log n)$ with high probability by using $O(n \log\log n)$ many message transmissions.

In Section 5.2 we dealt with the situation where a malicious dissemination process is combated by another dissemination process. The harmful dissemination process in our model was called $\mathcal{V}$. Its goal was to infect as many nodes of the network (n in total) as possible, and therefore inflict damage and/or costs. On the other hand, there were two different remedying processes Φ and Λ, representing different strategies used to counteract $\mathcal{V}$. By definition, their goal was to combat $\mathcal{V}$ by reactively curing all infected nodes as well as proactively immunizing possible contact points of $\mathcal{V}$. Obviously, the extinction of $\mathcal{V}$ was the ultimate goal. We showed that, with high probability, an infection disseminated by $\mathcal{V}$ does not infect more than $o(n)$ many nodes if the neighborhood of each infected node is immunized after a constant delay, i.e., in the round it was initially infected plus the delay, provided the minimum degree in the network is at least $\Omega(\log\log n)$. Hereby $\mathcal{V}$ spread over a constant amount of randomly chosen edges of each infected node.

In the case where the minimum degree in the network is given by a constant, we needed a more advanced strategy. In short, each infected node v executed the counteracting strategy Λ after a constant delay, i.e., in the round it was initially infected plus the delay. This resulted in the immunization of its $\beta \log^2(\deg_v)$-neighborhood, with $\deg_v$ denoting v's degree and β being a sufficiently large constant. In this context the k-neighborhood represents a set containing all nodes which are at most k hops away from v. Using this strategy, we showed that $\mathcal{V}$ does not infect more than $o(n)$ many nodes until it is eliminated, with high probability.

In Section 6.1 we examined an urban environment inhabited by n mobile individuals. In each round these individuals chose a location independently at random according to a power law distribution. First, we showed that a polynomial fraction of the individuals remains uninfected in a society model where the behavior of the susceptible entities is not influenced by an epidemic outbreak in any way, even if said individuals are allowed to transmit the (deadly) infection for $o(\log n)$ many rounds. That is, it is very unlikely for a (deadly) epidemic to wipe out the entire population.

Secondly, we showed that in a society model where an epidemic outbreak does influence the behavioral pattern of each susceptible entity and certain countermeasures are applied, only a polylogarithmic amount of agents is infected until the epidemic is stopped. Such countermeasures may include the isolation of individuals after a (large) constant number of rounds, for example. Ultimately, we showed that the epidemic is embanked within $\mathcal{O}((\log \log n)^4)$ rounds with probability $1 - o(1)$.

Section 6.2 primarily extends Section 6.1 to a large scope. We presented and empirically analyzed a dynamic model integrating several hundred of cities for the spread of epidemics on a national scope. The environment in said model approximates the geography of Germany, in which individuals may travel between cities. Within a city the individuals interact according to the probabilistic model presented in Section 6.1 in a distributed manner.

In this setting we showed that one can approximate the effect of some non-pharmaceutical countermeasures, which are usually adopted if an epidemic outbreak occurs, by setting the parameters properly. This observation is supported by the empirical study of [64]. Interestingly, the right choices of parameters in our experiments seem to be in line with previous observations in the real world (the right power law exponent seems to be in the range of 2.6-2.9 [39], for example).

7.2 Recommendations for Future Work

In the following we compose a list of promising areas for further research. Due to the vast scope of information dissemination and epidemics in general, we limit the upcoming recommendations primarily to the ones being either logical or technical extensions of the results, models, and methods we presented thus far. Of course, this includes the weaknesses and limitations of these. Clearly, also the most recent related work must be considered.

Information Dissemination in Power Law Networks Obviously, one key aspect in our analysis considering the information dissemination in Section 5.1 was the randomness. Although there are deterministic and randomized algorithms to disseminate information in power law graphs, it is not clear whether or not partial randomness could push the

limits even further. For example, nodes with low degree may work most efficiently in a deterministic manner, whereas nodes with high degree may spread the information over longer distances more often using random decisions.

A similar argument can be made with respect to the amount of different communication partners each individual node may choose. In our case, this amount did not depend on the node's degree. However, it is intuitively clear that the dissemination process would highly benefit from informed hubs, i.e., nodes with a very high degree, forwarding the information to as many uninformed nodes as possible. Of course, this is no trivial task, since the amount of informed nodes changes drastically over time and many connections of each hub may point to already informed nodes only. Using a randomly chosen set of communication partners depending on the node's degree may increase the amount of newly informed nodes (especially) in the beginning of the dissemination process. Afterwards, one could use the age of the corresponding information to estimate its dissemination status.

Another very promising approach may be to establish a meta structure primarily containing the hubs in the network. This idea is based on [9]. Provided we could manage to inform a good portion of the hubs first, all uninformed nodes then could simply call on such a hub and get informed in one additional step. Interestingly, one could say that these hubs already form a meta structure dictated by the network's topology. The main challenge would be to favor these nodes in the dissemination process in a fully decentralized way.

Strategies to Counteract Dissemination Processes An intuitive approach in the case of two competing information dissemination processes is certainly to strengthen the remedying process whilst keeping the induced costs by the malicious one as low as possible. The arising questions and problems are diverse. Suppose an efficient strategy to counteract the malicious dissemination process exists. One could wonder if it is possible to adapt said strategy according to the aggressiveness of the malicious dissemination process. This way one may be able to (drastically) lower the cost of using the remedying process whilst still maintaining a reasonable time frame for defeating the malicious dissemination process.

However, the above statements assumed that an efficient strategy to counteract the malicious dissemination process does exist. But how to design one? This is the second type of problems we suggest for further studies. Especially adaptive strategies with a lesser involvement of individual nodes are certainly interesting, yet challenging. For example, nodes that are relatively close to the infected ones could use more connections than others acting proactively only. This way one would combine the effectiveness of local immunization whilst reaching far into the distance, which instantly implies another fascinating approach: immunize/cure as few nodes as possible in your own neighborhood, then reach out into the distance for only one node (several hops away) to be immunized, and finally let this node immunize its vicinity. Such a strategy would certainly create

large immunized areas all over the network. If executed cleverly, methods like these may show us new ways to strategically counteract (malicious) dissemination processes.

Epidemic Spreading This topic is a very exciting one, since it involves our daily lives in an obvious way. Consequently, many possibilities for improvements can be deduced by observing our very own behavior and lifestyle. These improvements primarily involve modeling the individuals, their movement on a daily basis, and the environment itself.

For example, different types of standard movement models are conceivable such as Levy flight, periodic mobility model, and grid like movement. These movement models should incorporate the possibility for the individuals to move in a more realistic way, i.e., to meet other individuals on the way to a destination, as well as being able to represent a schedule. For most parts our daily life is determined by external circumstances such as the type of job we have or the activities we like to participate in during our spare time. As a result we tend to frequent the same locations most of the time. This kind of behavior combined with a realistic model of the environment should definitely be considered on an individual basis.

In order to make the environmental model more realistic, it is important to see that different individuals may have different preferences. Further, to the best of our knowledge, all models presented thus far allow disease spreading either only within locations or do not provide such anyway. There is no model allowing the disease to be spread on the way to a destination chosen by the infected individual (e.g., in the pedestrian zone or the subway).

Although there is still a long way to go, we feel that all of the above recommendations could have a significant impact in their respective fields and are therefore certainly worth further studies.

List of Figures

List of Tables

Bibliography

[1] L. A. Adamic and B. A. Huberman. Power-law distribution of the world wide web. *Science*, 287(5461):2115, 2000.

[2] W. Aiello, F. Chung, and L. Lu. A random graph model for massive graphs. In *Proceedings of the 32nd annual ACM symposium on Theory of computing*, STOC '00, pages 171–180. ACM, 2000.

[3] M. Ajelli, B. Goncalves, D. Balcan, V. Colizza, H. Hu, J. Ramasco, S. Merler, and A. Vespignani. Comparing large-scale computational approaches to epidemic modeling: Agent-based versus structured metapopulation models. *BMC Infectious Diseases*, 10(190), 2010.

[4] R. Albert, H. Jeong, and A. Barabási. Internet: Diameter of the world-wide web. *Nature*, 401(6749):130–131, 1999.

[5] R. Albert, H. Jeong, and A.-L. Barabási. Error and attack tolerance of complex networks. *Nature*, 406(6794):378–382, July 2000.

[6] E. Alirol, L. Getaz, B. Stoll, F. Chappuis, and L. Loutan. Urbanisation and infectious diseases in a globalised world. *The Lancet Infectious Diseases*, 11(2):131 – 141, 2011.

[7] L. A. Amaral, A. Scala, M. Barthelemy, and H. E. Stanley. Classes of small-world networks. *Proceedings of the National Academy of Sciences*, 97(21):11149–11152, 2000.

[8] Arbeitsgemeinschaft Influenza. Bericht zur Epidemiologie der Influenza in Deutschland Saison 2010/11. http://www.rki.de/DE/Content/InfAZ/I/Influenza/PK_AGI_2011_AGI-Saisonbericht_10_11.html?nn=2370434, September 2011.

[9] C. Avin and R. Elsässer. Faster rumor spreading: Breaking the logn barrier. In *Proceedings of Distributed Computing - 27th International Symposium*, DISC '13, pages 209–223. Springer, 2013.

[10] D. Balcan, H. Hu, B. Goncalves, P. Bajardi, C. Poletto, J. J. Ramasco, D. Paolotti, N. Perra, M. Tizzoni, W. V. den Broeck, V. Colizza, and A. Vespignani. Seasonal transmission potential and activity peaks of the new influenza A(H1N1): a Monte Carlo likelihood analysis based on human mobility. *BMC Medicine*, 7:45, 2009.

[11] A. L. Barabási. Scale-free networks: A decade and beyond. *Science*, 325:412, 2009.

[12] P. Berenbrink, R. Elsässer, and T. Friedetzky. Efficient randomised broadcasting in random regular networks with applications in peer-to-peer systems. In *Proceedings of the 27th ACM Symposium on Principles of Distributed Computing*, PODC '08, pages 155–164. ACM, 2008.

[13] P. Berenbrink, R. Elsässer, and T. Sauerwald. Communication complexity of quasirandom rumor spreading. In *Proceedings of the 18th Annual European Conference on Algorithms: Part I*, ESA'10, pages 134–145. Springer, 2010.

[14] P. Berenbrink, R. Elsässer, and T. Sauerwald. Randomised broadcasting: Memory vs. randomness. In *Proceedings of the 9th Latin American Conference on Theoretical Informatics*, LATIN '10, pages 306–319. Springer, 2010.

[15] L. Bettencourt, J. Lobo, D. Helbing, C. Kuhnert, and G. West. Growth, innovation, scaling, and the pace of life in cities. *Proceedings of the National Academy of Sciences*, 104(17):7301–7306, 2007.

[16] G. V. Bobashev, D. M. Goedecke, F. Yu, and J. M. Epstein. A hybrid epidemic model: combining the advantages of agent-based and equation-based approaches. In *Proceedings of the Winter Simulation Conference*, WSC '07, pages 1532–1537. IEEE Press, 2007.

[17] C. Borgs, J. Chayes, A. Ganesh, and A. Saberi. How to distribute antidote to control epidemics. *Random Struct. Algorithms*, 37(2):204–222, September 2010.

[18] F. Carrat, J. Luong, H. Lao, A.-V. Sallé, C. Lajaunie, and H. Wackernagel. A 'small-world-like' model for comparing interventions aimed at preventing and controlling influenza pandemics. *BMC Medicine*, 4(1):1–14, 2006.

[19] M. Cernea. Social integration and population displacement : the contribution of social science. *International Social Science Journal*, 143(1):93–112, 1995.

[20] F. Chierichetti, S. Lattanzi, and A. Panconesi. Rumor spreading in social networks. In *Proceedings of the 36th Internatilonal Colloquium on Automata, Languages and Programming: Part II*, ICALP '09, pages 375–386. Springer, 2009.

[21] A. Clementi, P. Crescenzi, C. Doerr, P. Fraigniaud, M. Isopi, A. Panconesi, F. Pasquale, and R. Silvestri. Rumor spreading in random evolving graphs. In *Proceedings of the 21st European Symposium on Algorithms*, ESA '13, pages 325–336. Springer, 2013.

[22] A. Clementi and F. Pasquale. Information spreading in dynamic networks: An analytical approach. In *Theoretical Aspects of Distributed Computing in Sensor Networks*, Monographs in Theoretical Computer Science. An EATCS Series, pages 591–619. Springer, 2011.

[23] R. Cohen, K. Erez, D. Ben-Avraham, and S. Havlin. Resilience of the Internet to random breakdowns. *Physical Review Letters*, 85(21):4626–4628, November 2000.

[24] R. Cohen and S. Havlin. Scale-free networks are ultrasmall. *Physical Review Letters*, 90(5):058701, Feb 2003.

[25] B. S. Cooper, R. J. Pitman, W. J. Edmunds, and N. J. Gay. Delaying the international spread of pandemic influenza. *Public Library of Science Medicine*, 3(6):e212, 05 2006.

[26] C. Cooper, R. Elsässer, and T. Radzik. The power of two choices in distributed voting. In *Proceedings of the 41st International Colloquium on Automata, Languages, and Programming*, ICALP '14, 2014. to appear.

[27] X. De La Barra. Fear of epidemics: The engine of urban planning. *Planning Practice and Research*, 15(1-2):7–16, 2000.

[28] B. Doerr, M. Fouz, and T. Friedrich. Social networks spread rumors in sublogarithmic time. In *Proceedings of the 43rd Annual ACM Symposium on Theory of Computing*, STOC '11, pages 21–30. ACM, 2011.

[29] B. Doerr, M. Fouz, and T. Friedrich. Why rumors spread so quickly in social networks. *Communications of the ACM*, 55(6):70–75, 2012.

[30] B. Doerr, L. A. Goldberg, L. Minder, T. Sauerwald, and C. Scheideler. Stabilizing consensus with the power of two choices. In *Proceedings of the 23rd Annual ACM Symposium on Parallelism in Algorithms and Architectures*, SPAA '11, pages 149–158. ACM, 2011.

[31] B. Doerr, D. Johannsen, and C. Winzen. Multiplicative drift analysis. In *Proceedings of the 12th Annual Conference on Genetic and Evolutionary Computation*, GECCO '10, pages 1449–1456. ACM, 2010.

[32] M. Draief and A. Ganesh. A random walk model for infection on graphs: spread of epidemics & rumours with mobile agents. *Discrete Event Dynamic Systems*, 21(1):41–61, March 2011.

[33] R. Durrett. *Random Graph Dynamics (Cambridge Series in Statistical and Probabilistic Mathematics)*. Cambridge University Press, 2006.

[34] R. Elsässer. On randomized broadcasting in power law networks. In *Proceedings of the 20th International Symposium on Distributed Computing*, DISC '06, pages 371–385. Springer, 2006.

[35] R. Elsässer and A. Ogierman. Efficient broadcasting in random power law networks. In *Proceedings of the 36th International Workshop on Graph Theoretic Concepts in Computer Science*, WG '10, pages 279–291. Springer, 2010.

[36] R. Elsässer and A. Ogierman. The impact of the power law exponent on the behavior of a dynamic epidemic type process. In *Proceedings of the 24th ACM Symposium on Parallelism in Algorithms and Architectures*, SPAA '12, pages 131–139. ACM, 2012.

[37] R. Elsässer, A. Ogierman, and M. Meier. Agent based simulations of epidemics on a large scale. In *Proceedings of the 3rd International Conference on Simulation and Modeling Methodologies, Technologies and Applications*, SIMULTECH '13, pages 263 – 274. SciTePress, 2013.

[38] R. Elsässer and T. Sauerwald. The power of memory in randomized broadcasting. In *Proceedings of the 19th Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '08, pages 218–227. Society for Industrial and Applied Mathematics, 2008.

[39] S. Eubank, H. Guclu, V. Kumar, M. Marathe, A. Srinivasan, Z. Toroczkai, and N. Wang. Modelling disease outbreaks in realistic urban social networks. *Nature*, 429(6988):180–184, 2004.

[40] M. Faloutsos, P. Faloutsos, and C. Faloutsos. On power-law relationships of the internet topology. In *Proceedings of the conference on Applications, technologies, architectures, and protocols for computer communication*, SIGCOMM '99, pages 251–262. ACM, 1999.

[41] N. M. Ferguson, D. A. T. Cummings, C. Fraser, J. C. Cajka, P. C. Cooley, and D. S. Burke. Strategies for mitigating an influenza pandemic. *Nature*, 442(7101):448–452, July 2006.

[42] N. Fountoulakis, A. Huber, and K. Panagiotou. Reliable broadcasting in random networks and the effect of density. In *Proceedings of the 29th Conference on Information Communications*, INFOCOM '10, pages 2552–2560. IEEE Press, 2010.

[43] N. Fountoulakis, K. Panagiotou, and T. Sauerwald. Ultra-fast rumor spreading in social networks. In *Proceedings of the 23rd Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '12, pages 1642–1660. Society for Industrial and Applied Mathematics, 2012.

[44] P. Fraigniaud and E. Lazard. Methods and problems of communication in usual networks. *Discrete Applied Mathematics*, 53(1-3):79–133, 1994.

[45] S. Funk, E. Gilad, C. Watkins, and V. A. A. Jansen. The spread of awareness and its impact on epidemic outbreaks. *Proceedings of the National Academy of Sciences*, 106(16):6872–6877, 2009.

[46] R. Gardner. The Plague. DVD/TV, History Channel, October 2005. http://www.imdb.com/title/tt0499545/combined.

[47] G. Giakkoupis. Tight bounds for rumor spreading in graphs of a given conductance. In *Proceedings of the 28th International Symposium on Theoretical Aspects of Computer Science*, STACS '11, pages 57–68. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2011.

[48] G. Giakkoupis. Tight bounds for rumor spreading with vertex expansion. In *Proceedings of the 25th ACM-SIAM Symposium on Discrete Algorithms*, SODA '14, pages 801–815. Society for Industrial and Applied Mathematics, 2014.

[49] A. Grabowski, N. Kruszewska, and R. A. Kosinski. Properties of on-line social systems. *The European Physical Journal B - Condensed Matter and Complex Systems*, 66(1):107–113, 2008.

[50] T. Hagerup and C. Rüb. A guided tour of Chernoff bounds. *Information Processing Letters*, 33(6):305–308, 1990.

[51] R. J. Hatchett, C. E. Mecher, and M. Lipsitch. Public health interventions and epidemic intensity during the 1918 influenza pandemic. *Proceedings of the National Academy of Sciences*, 104(18):7582–7587, 2007.

[52] S. M. Hedetniemi, S. T. Hedetniemi, and A. L. Liestman. A survey of gossiping and broadcasting in communication networks. *Networks*, 18(4):319–349, 1988.

[53] H. W. Hethcote. The mathematics of infectious diseases. *SIAM Review*, 42(4):599–653, December 2000.

[54] J. Hromkovic, R. Klasing, A. Pelc, P. Ruzicka, and W. Unger. *Dissemination of Information in Communication Networks: Broadcasting, Gossiping, Leader Election, and Fault-Tolerance*. Texts in Theoretical Computer Science. An EATCS Series. Springer, 2005.

[55] F. Iozzi, F. Trusiano, M. Chinazzi, F. C. Billari, E. Zagheni, S. Merler, M. Ajelli, E. D. Fava, and P. Manfredi. Little italy: An agent-based approach to the estimation of contact patterns- fitting predicted matrices to serological data. *Public Library of Science Computational Biology*, 6(12), 2010.

[56] S. W. Jaffry and J. Treur. Agent-based and population-based simulation: A comparative case study for epidemics. In *Proceedings of the 22th European Conference on Modelling and Simulation*, ECMS '08, pages 123–130. European Council for Modeling and Simulation, 2008.

[57] H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai, and A. L. Barabási. The large-scale organization of metabolic networks. *Nature*, 407(6804):651–654, October 2000.

[58] M. Jovanovic, F. Annexstein, and K. Berman. Scalability issues in large peer-to-peer networks - a case study of Gnutella. Technical report, University of Cincinatti, 2001.

[59] R. Karp, C. Schindelhauer, S. Shenker, and B. Vocking. Randomized rumor spreading. In *Proceedings of the 41st Annual Symposium on Foundations of Computer Science*, FOCS '00, pages 565–574. IEEE Computer Society, 2000.

[60] B. Y. Lee, V. L. Bedford, M. S. Roberts, and K. M. Carley. Virtual epidemic in a virtual city: simulating the spread of influenza in a us metropolitan area. *Translational Research*, 151(6):275 – 287, 2008.

[61] B. Y. Lee, S. T. Brown, P. C. Cooley, R. K. Zimmerman, W. D. Wheaton, S. M. Zimmer, J. J. Grefenstette, T.-M. Assi, T. J. Furphy, D. K. Wagener, and D. S. Burke. A computer simulation of employee vaccination to mitigate an influenza epidemic. *American Journal of Preventive Medicine*, 38(3):247 – 257, 2010.

[62] I. M. Longini, A. Nizam, S. Xu, K. Ungchusak, W. Hanshaoworakul, D. A. T. Cummings, and M. E. Halloran. Containing pandemic influenza at the source. *Science*, 309(5737):1083–1087, 2005.

[63] B. J. Marais, C. K. Mlambo, N. Rastogi, T. Zozio, A. G. Duse, T. C. Victor, E. Marais, and R. M. Warren. Epidemic spread of multidrug-resistant tuberculosis in johannesburg, south africa. *Journal of Clinical Microbiology*, 51(6):1818–1825, 2013.

[64] H. Markel, H. B. Lipman, J. A. Navarro, A. Sloan, J. R. Michalsen, A. M. Stern, and M. S. Cetron. Nonpharmaceutical interventions implemented by us cities during the 1918-1919 influenza pandemic. *The Journal of the American Medical Association*, 298(6):644–654, August 2007.

[65] R. Motwani and P. Raghavan. *Randomized Algorithms*. Cambridge University Press, 1995.

[66] M. E. J. Newman. The structure and function of complex networks. *SIAM Review*, 45(2):167–256, 2003.

[67] R. Pastor-Satorras and A. Vespignani. Epidemic spreading in scale-free networks. *Physical Review Letters*, 86(14):3200–3203, 2001.

[68] P. C. Pinto, P. Thiran, and M. Vetterli. Locating the source of diffusion in large-scale networks. *Physical Review Letters*, 109(6):068702, Aug 2012.

[69] S. Redner. How popular is your paper? An empirical study of the citation distribution. *European Physical Journal B*, 4(2):131–134, August 1998.

[70] Remco van der Hofstad, G. Hooghiemstra, and D. Znamenski. A phase transition for the diameter of the configuration model. *Internet Mathematics*, 4(1):113–128, 2007.

[71] R. Reyes, R. Ahn, K. Thurber, and T. Burke. Urbanization and infectious diseases: General principles, historical perspectives, and contemporary challenges. In *Challenges in Infectious Diseases*, Emerging Infectious Diseases of the 21st Century, pages 123–146. Springer, 2013.

[72] M. Ripeanu, I. Foster, and A. Iamnitchi. Mapping the gnutella network: Properties of large-scale peer-to-peer systems and implications for system. *IEEE Internet Computing Journal*, 6(1):50–57, 2002.

[73] Robert Koch Institute. SurvStat@RKI. www3.rki.de/SurvStat/QueryForm.aspx, 2012. A web-based solution to query surveillance data in Germany.

[74] F. D. Sahneh and C. M. Scoglio. Epidemic spread in human networks. In *Proceedings of the 50th IEEE Conference on Decision and Control and European Control Conference*, CDC-ECE '11, pages 3008–3013. IEEE Press, 2011.

[75] M. Tizzoni, P. Bajardi, C. Poletto, J. Ramasco, D. Balcan, B. Goncalves, N. Perra, V. Colizza, and A. Vespignani. Real-time numerical forecast of global epidemic spreading: case study of 2009 a/h1n1pdm. *BMC Medicine*, 10(1):165, 2012.

[76] R. M. Tripathy, A. Bagchi, and S. Mehta. A study of rumor control strategies on social networks. In *Proceedings of the 19th ACM International Conference on Information and Knowledge Management*, CIKM '10, pages 1817–1820. ACM, 2010.

[77] S. D. Valle, J. Hyman, H. Hethcote, and S. Eubank. Mixing patterns between age groups in social networks. *Social Networks*, 29(4):539 – 554, 2007.

[78] N. C. Valler, B. A. Prakash, H. Tong, M. Faloutsos, and C. Faloutsos. Epidemic spread in mobile ad hoc networks: Determining the tipping point. In *Proceedings of the 10th International IFIP TC 6 Conference on Networking - Volume Part I*, NETWORKING '11, pages 266–280. Springer, 2011.

[79] B. Wang, L. Cao, H. Suzuki, and K. Aihara. Epidemic spread in adaptive networks with multitype agents. *Journal of Physics A: Mathematical and Theoretical*, 44(3):035101, 2011.

[80] D. Watts and S. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393(6684):440–442, June 1998.