Precise Analysis of Epidemic Algorithms

Anatolii Kostrygin

NNT : 2017SACLX042

Thèse de doctorat de l'Université Paris-Saclay préparée à l'École Polytechnique

Ecole doctorale n°580
Sciences et technologies de l'information et de la communication (STIC)
Spécialité de doctorat: Informatique

par

M. Anatolii Kostrygin

Precise Analysis of Epidemic Algorithms
Analyse précise des algorithmes épidemiques

Thèse à présenter et soutenir au LIX, École Polytechnique, le 29 août 2017.

Composition du Jury :

Mme	Joanna Tomasik	Professeur SUPÉLEC Département Informatique	(Présidente du jury)
M.	Nicolas Hanusse	Directeur de Recherche LaBRI, Université de Bordeaux	(Rapporteur)
M.	Amos Korman	Chargé de Recherche Université Paris-Diderot	(Rapporteur)
Mme	Johanne Cohen	Directeur de Recherche LRI Université de Paris-Sud	(Examinatrice)
M.	Adrian Kosowski	Chargé de Recherche IRIF, University Paris Diderot	(Examinateur)
M.	Marvin Künnemann	Postdoctoral Scolar Max Planck Insitute for Informatics	(Examinateur)
M.	Benjamin Dœrr	Professeur LIX, École Polytechnique	(Directeur de thèse)

Acknowledgements

My deepest gratitude is to my advisor, Professor Benjamin Doerr, whose expertise and knowledge added considerably to my graduate experience. I am fortunate to have an advisor who gave me the freedom to explore on my own, and at the same time gave an advice whenever it was necessary.

I am grateful to Professor Luca Castelli Aleardi and Éric Fusy from the Laboratoire d'Informatique d'École Polytechnique, the supervisors of my master internship, for guiding my first steps in research and teaching me how to write a good mathematical text.

I owe my gratitude to all members of the Laboratoire d'Informatique d'École Polytechnique for the enjoyable four years I spent here. I am grateful to Evelyne Rayssac, Sylvie Jabinet, and Vanessa Molina Magana for their efficient help with all administrative procedures I had to pass as a foreign student.

I would also like to thank my fellow Ph.D. students, especially Alina and Dorian for our interesting and stimulating discussions on various topics, for good laughs and tricky mathematical problems we shared.

Last but not least, my family have always been a source of strength, advice, and encouragement through all endeavors I ever undertook. And, of course, the continuous support of my closest friends, helped me to overcome many obstacles and setbacks.

Résumé Substantiel

Thème de recherche

Un problème tout aussi fondamental et central que la recherche de consensus est la dissémination collaborative d'une information d'un agent à tous les autres agents d'un système distribué. Comme pour la recherche de consensus, ce problème est particulièrement important lorsque l'on veut obtenir des algorithmes distribués qui à la fois sont robustes et fonctionnent dans un cadre anonyme, c'est-à-dire sans supposer que les agents possèdent des identifiants distincts connus. Ce problème, connu sous le nom de *problème de propagation de rumeur*, est à la base de nombreux algorithmes de communication sur des réseaux de capteurs sans-fil [DKM$^+$10] ou des réseaux mobiles ad-hoc, et est aussi une brique de base centrale pour de nombreux algorithmes distribués avancés, e.g., [MS08].

Les méthodes les plus connues pour surmonter les défis de robustesse et d'anonymat sont les algorithmes basés sur les ragots (*gossip-based algorithms*). Il s'agit d'algorithmes dans lesquels les agents contactent d'autres agents aléatoires pour échanger de l'information. Bien que dans ce projet nous nous concentrions sur le problème de la dissémination d'une seule unité d'information, nous devons mentionner que les algorithmes basés sur les ragots ont des applications vagues dans les problèmes plus complexes, comme maintenir la cohérence d'une base de données distribuée.

Algorithmes de propagation de rumeur

Les algorithmes de propagation de rumeur sont basés sur la paradigme que les agents contact aléatoirement les autres agents pour envoyer ou récupérer l'information. Dans ce travail, nous nous concentrons uniquement sur les protocoles de temps discret, c'est-à-dire, nous supposons que tous les agents ont l'accès à une horloge commune. Cette horloge partitionne la durée du procès en tours de communication discrets. Un nombre important des algorithmes différents a été proposé même avec cette restriction. Ils diffèrent par la manière comment les agents lancent

la communication, comment ils font le choix des partenaires de communication et quelles hypothèses sont faites sur le réseau.

Initiation du contact. En fonction des applications l'une des trois configurations peut être raisonnable : soit les agents déjà informés contactent les autres (*protocole de push*), soit les agents non informés cherche les nouvelles en appelant les autres (*protocole de pull*), soit tous les agents peuvent faire des appels indépendamment de leur statut (*protocole de push-pull*).

Nombre des communications par tour. Bien que dans l'hypothèse typique chaque agent puisse contacter au plus un partenaire de la communication, il était proposé d'autoriser plus d'un contact par tour. Pour les certains réseaux de communication (par exemple, les réseaux sociaux) il a été observé une différence considérable entre les deux approches [DFF12c].

Choix des partenaires de communication. Notamment dans les réseaux avec les communications non fiables, il est risqué de compter sur les actions précédentes. Pour cette raison, nous admettons l'hypothèse typique suivante : les agents agissent stochastiquement indépendantes chaque tour. Néanmoins, un certain nombre d'améliorations a été obtenu pour les réseaux fiables en proposant des protocoles plus sophistiqués.

Fiabilité du réseau. La structure du réseau est loin d'être fiable dans les applications typiques de l'algorithme basé sur les ragots comme les réseaux de capteurs sans-fil ou les réseaux ad-hoc. Pour cette raison l'hypothèse réaliste est que tous les appels atteignent leurs cibles avec une certaine probabilité positive. Faute d'un meilleur modèle, nous supposons que les échecs de la communication sont distribués de manière indépendante.

Jusqu'à présent, toutes les analyses des algorithmes basés sur les ragots étudient les combinaisons spécifiques de ces variantes. Dans tous les cas la plupart des recherches étude de manière individuelle la loi de probabilité décrivant le progrès réalisé en un seul tour. L'objectif de notre travail est d'extraire les conditions générales qui couvrent la plupart des protocoles et qui permettent de prouver les garanties de performance basées seulement sur ces conditions. Cette approche peut simplifier un grand corps des travaux existants aussi bien que prédire les performances de nouveaux protocoles.

Nos résultats

Nous avons réussi à analyser une large classe des algorithmes basés sur les ragots. Pour la classe naturelle des protocoles uniformes qui traitent les agents symétriquement et dans laquelle chaque agent peut communiquer avec tous les autres agents, nous avons déterminé trois conditions suivantes qui couvrent la plupart des protocoles qui ont été étudiés précédemment.

Conditions de Croissance et de Contraction

À partir d'ici supposons que n est le nombre d'agents dans notre système multi-agents. Supposons que tout agent non informé s'informe pendant le tour initié par $k < n$ agents informés avec la probabilité p_k indépendant du choix d'un agent. Les conditions ci-dessous sont exprimées en forme des probabilités de succès p_k et de covariance c_k des variables indicatrices aléatoires pour les événements que deux agents différents s'informent simultanément (voir Définition II.12).

Conditions de la Croissance Exponentielle. Soit γ un constant positif. Soit $a, b, c \geq 0$ et $0 < f < 1$ tels que $af < 1$. Nous disons que le protocole satisfait les *conditions de la croissance exponentielle*, si pour tout $n \in \mathbb{N}$ assez grand, $k < fn$ implique

(i) tout agent non informé s'informe pendant le tour initié par k agents informés avec la probabilité $p_k = \gamma\frac{k}{n} \cdot \left(1 \pm a\frac{k}{n} \pm \frac{b}{\ln n}\right)$;

(ii) $c_k \leq c\frac{k}{n^2}$.

Conditions de la Contraction Exponentielle. Soit $0 < g < 1$ et $\rho > 0$. Soit $a, c \geq 0$. Nous disons que le protocol satisfait les *conditions de la contraction exponentielle* si pour tout $n \in \mathbb{N}$ assez grand, $u < gn$ implique que

(i) tout agent non informé reste non informé pendant le tour initié par u agents non informés avec la probabilité $1 - p_{n-u} = e^{-\rho} \pm a\frac{u}{n}$;

(ii) $c_{n-u} \leq \frac{c}{u}$.

Conditions de la Contraction Double Exponentielle. Soit $g, \alpha \in]0, 1[$ et $\ell > 1$. Soit $a, a', c \geq 0$. Nous disons que le protocole satisfait les *conditions de la contraction double exponentielle* si pour tout n assez grand, pour tout $u \in [n^{1-\alpha}, n - gn]$ les conditions suivantes sont satisfaites pour le tour avec u agents non informés initialement.

1. La probabilité $1-p_{n-u}$ qu'un agent non informé reste non informé est borné entre $a\left(\frac{u}{n}\right)^{\ell-1}$ et $a'\left(\frac{u}{n}\right)^{\ell-1}$;

2. $c_{n-u} \leq c\frac{n}{u^2}$.

Il se trouve que la grande majorité des protocoles considérés dans la littérature peut être décrite par ces trois conditions. Plus précisément, presque tous les des protocoles satisfont les *conditions de la croissance exponentielle*, souvent $\gamma = 2$ or 3. La plupart des protocoles de push satisfont les *conditions de la contraction exponentielle* tandis que dans le cas de communications fiables les protocoles de pull et de push-pull satisfont les *conditions de la contraction double exponentielle*.

Estimations Precises de la Performance

Tous les trois conditions assurent l'analyse précise de la performance des algorithmes basés sur les ragots. Elles nous permettent l'espérance du temps nécessaire pour informer tous les agents ainsi que les valeurs qui bornent ce temps de dissémination avec une probabilité forte (voir Théorème II.14, II.15 et II.16).

Résultat Principal

(i) Si le protocole satisfait les conditions de la croissance exponentielle, alors l'espérence du temps jusqu'à fn sont informés est $\log_{1+\gamma} n \pm O(1)$.

(ii) Si le protocole satisfait les conditions de la contraction exponentielle, alors l'espérence du temps nécessaire pour informer tous les agents à partir du moment que au moins $(1-g)n$ agents sont informés, est $\frac{1}{\rho}\ln n \pm O(1)$.

(iii) Si le protocole satisfait les conditions de la contraction double exponentielle, alors l'espérence du temps nécessaire pour informer tous les agents à partir du moment que au moins $(1-g)n$ agents sont informés, est $\log_\ell \ln n \pm O(1)$.

Nous notons que ces estimations sont extrêmement précises, le temps d'exécution est borné jusqu'à un constant additif. Jusqu'à présent aucune estimation n'a été autant précise sauf pour le cas basique du protocole de push. Donc en plus d'une méthode d'analyse très générale, notre travail donne des meilleures estimations pour la plupart des protocoles existants.

Nous notons aussi que les estimations du temps d'exécution sont indépendants de la plupart des paramètres introduits dans les conditions. Seulement le taux de croissance γ pour la croissance exponentielle, le taux de la contraction ρ pour la contraction exponentielle et l'exposant ℓ pour la contraction double exponentielle apparaissent dans les estimations. Donc les autres paramètres ont l'influence qui

est limitée par un nombre constant de tours et peuvent être négligés facilement. La preuve de notre résultat principal est décrite en Chapitre III.

Le résultat principal donne immédiatement les estimations du temps pour les protocoles concrètes. Dans le Chapitre IV nous appliquons les conditions ci-dessus pour les différents algorithmes de propagation de rumeur.

Protocoles de push. Dans le cas basique connu comme le protocole du push, chaque agent informé appelle un autre agent aléatoire une fois par tour. Ce protocole satisfait les conditions de la croissance exponentielle avec $\gamma = 1$ et les conditions de la contraction exponentielle avec $\rho = 1$. Par conséquent, le temps d'exécution pour le protocole de push est $\log_2 n + \ln n + O(1)$ (Théorème IV.1). Si on permet aux agents d'appeler un nombre fixe m d'agents par tour, alors γ et ρ sont égales à m, c'est-à-dire, le temps d'exécution est $\log_{1+m} n + \frac{1}{m} \ln n + O(1)$ (Théorème IV.7). Il n'y a aucune différence si chaque agent fait les appels aux m agents distincts ou aux m agents choisi aléatoirement. Si nous supposons que notre réseau ne soit pas fiable, c'est-à-dire chaque appel atteint sa destination avec la probabilité p, alors le temps d'exécution est $\log_{pm+1} n + \frac{1}{pm} \ln n + O(1)$. Par ailleurs il n'y a aucune différence si les appels échoués sont notés et répétés le prochain tour.

Protocoles de pull. Si chaque tour chaque agent non informé cherche l'information en appelant un autre agent aléatoire par tour, alors les conditions de la croissance exponentielle sont toujours satisfaites avec $\gamma = 1$. Désormais, les conditions de la contraction double exponentielle sont satisfaites avec $\ell = 2$, ce que prévoit la finition très rapide du procès avec la durée total de $\log_2 n + \log_2 \ln n + O(1)$ tours (Théorème IV.2). Pour chaque appel supplémentaire que les agents effectuent par tour, les taux du procès augment de 1. Donc pour le cas où un agent peut appeler m voisins, le temps d'exécution est $\log_{m+1} n + \log_{m+1} \ln n + O(1)$.

Cependant, tout cela est vrai seulement si le réseau est complètement fiable. Si nous supposons que les appels atteignent leurs destinations avec la probabilité $p < 1$, alors les conditions de la contraction double exponentielle ne sont plus satisfaites. Par contre, ce protocole satisfait les conditions de la contraction exponentielle. On en déduit que le temps d'exécution est $\log_{pm+1} n + \frac{1}{m \ln(1/(1-p))} \ln n + O(1)$.

Protocoles de push-pull. Le protocole de push-pull basique avec tous les agents appelant d'autres agents aléatoires satisfait les conditions de la croissance exponentielle avec $\gamma = 2$ (deux fois plus efficace que protocoles de push ou de pull tout seuls) et les conditions de la contraction double exponentielle avec $\ell = 2$ (aussi efficace que le protocole du pull). Nous en déduisons que le temps d'exécution est $\log_3 n + \log_2 \ln n + O(1)$ (Théorème IV.3). Avec m appels par agent par tour,

le temps s'améliore à $\log_{2m+1} n + \log_{m+1} \ln n + O(1)$ (Théorème IV.8). Si nous supposons une fraction constant des échecs de la communication, la contraction double exponentielle sera remplacée par l'exponentielle simple, exactement comme dans le cas du protocole de pull. Avec le taux de réussite des appels p, le temps d'exécution est $\log_{pm+1} n + \frac{1}{m(\ln(1/(1-p))+1)} \ln n + O(1)$.

La méthode proposée ci-dessus améliore plusieurs résultats clés obtenus au passé par différents groupes de chercheurs (par exemple, [Pit87], [KSSV00], and [DHL13]). Elle permet aussi de déterminer facilement le temps d'exécution de nouveaux algorithmes et aide ainsi le développement d'algorithmes de dissémination supérieurs.

Notons que malgré la ressemblance de certains résultats ci-dessus, les processus randomisés sont très différents. Par exemple, considérons le premier tour d'un processus de propagation de rumeur. Dans le cas du protocole de push, exactement un nouvel agent sera informé presque sûrement. Dans le cas du protocole de pull, l'espérance du nombre de nouveaux agents informés vaut aussi un. Par contre, la distribution est très différente, asymptotiquement elle est poissonnien. Désormais, personne ne sera informé avec une probabilité proche de $1/e$. Il est aussi possible avec une petite probabilité que plusieurs agents seront informés. La méthode simple qui peut traiter ces protocoles différents est un point fort de notre approche.

Finalement, dans le Chapitre V nous montrons que notre méthode a un potentiel de généralisation sur les processus de propagation de rumeur dans lesquels les agents peuvent se trouver dans plus que deux états, par exemple les certains agents informés peuvent devenir inactifs et s'arrêter à faire des appels.

Contents

Résumé Substantiel **iii**

Contents **ix**

List of Figures **xi**

List of Tables **xi**

I Introduction **1**
1 What Is Rumor Spreading 1
2 Motivation 3
3 State of the Art 6
4 Our Contribution 9
5 Plan of the Work 12

II Overview of Gossip Protocols **15**
1 Classic Results 15
1.1 Definition of Rumor Spreading 15
1.2 Independent Random Phone-Call Model 17
1.3 Asynchronous Rumor Spreading 19
1.4 Path, Square Lattice, Star, Necklace 20
1.5 Complete Graph. Overview of the Proofs 25
2 Precise Statements of Our Results 30
2.1 Tight Bounds via a Target-Failure Calculus 30
2.2 Uniform Treatment of Many Rumor Spreading Processes . 31
2.3 Precise Statement of the Technical Results 32
2.4 Applying the Above Technical Results 35
2.5 Limitations of the Phase Method 37

III Main Analysis Technique **39**
1 Homogeneous Rumor Spreading Processes 39
2 Exponential Growth Regime 42

2.1 Upper Bound 42
2.2 Lower Bound 49
3 Exponential Shrinking Regime 55
3.1 Upper Bound 55
3.2 Lower Bound 60
4 Double Exponential Shrinking Regime 63
4.1 Upper Bound. 63
4.2 Lower Bound. 67

IV Applications of our Method **71**
1 Classic Protocols 71
1.1 Push Protocol 72
1.2 Pull Protocol 73
1.3 Push-Pull Protocol 75
2 Robustness, Multiple Calls, and Dynamic Graphs 76
2.1 Transmission Failures 77
2.2 Multiple Calls 80
2.3 Dynamic Graphs 84
3 Limited Incoming Calls Capacity 92
3.1 Single Incoming Call Push-Pull Protocol 93
3.2 Single Incoming Call Pull-Only Protocol 95
3.3 Push-Pull Protocol with Transition Time 96

V Non-Homogeneous Rumor Spreading **99**
1 Two-State Multi-Parametric Process 100
2 Multi-State Rumor Spreading. Independent Stop Process 104
2.1 One Push Call per Node 105
2.2 Pull Protocol with $C \geq 1$ Push Calls: 107
2.3 Random Stop Decision 110
3 Multiparametric exponential growth 113

VI Summary **117**
1 Outlook 117
2 Open Problems 118

A Probabilistic notions **121**

B First Order Bounds **125**

C Coupon Collector and Ball into Bins **127**

Bibliography **129**

List of Figures

1 Regular square lattice 22
2 Necklace graph 25
3 Phase transition for the push protocol 28
4 Phase structure of Frieze and Gimmet's proof [FG85] 28
5 Phase structure of Pittel's proof [Pit87] 29
6 Phase structure of Karp's proof [KSSV00] 30

7 Rumor spreading time versus size of the network 109

List of Tables

1 New and previous-best results for rumor spreading time T of the classic rumor spreading protocols in complete graphs on n vertices... 7
2 New and previous-best results for the spreading time T of the variations of the classic rumor spreading protocols in complete graph on n vertices..... 9

Chapter I

Introduction

Randomized rumor spreading is one of the core primitives to disseminate information in distributed networks. It builds on the paradigm that nodes call random neighbors and exchange information with these contacts. This gives highly robust dissemination algorithms belonging to the broader class of gossip-based algorithms that, due to their epidemic nature, are surprisingly efficient and scalable. Randomized rumor spreading has found numerous applications (see Section 2), among others, in the consistency maintenance of replicated databases [DGH$^+$87], to disseminate large amounts of data in a scalable manner [MSF$^+$12], and to organize any kind of communication in highly dynamic and unreliable networks like wireless sensor networks and mobile ad-hoc networks [IvS10]. Randomized rumor spreading processes are also used to model epidemic processes like viruses spreading over the Internet [BBCS05], news spreading in social networks [DFF12b], or opinions forming in social networks [Kle08].

The importance of these processes not only has led to a huge body of experimental results, but, starting with the influential works of Frieze and Grimmett [FG85] and Karp, Shenker, Schindelhauer, and Vöcking [KSSV00] also to a large number of mathematical analyses of rumor spreading algorithms giving runtime or robustness guarantees for existing algorithms and, based on such findings, proposing new algorithms.

1 What Is Rumor Spreading

To better understand the idea of the rumor spreading algorithms, let us consider the following problem. There are n people that can communicate only by pairwise phone calls. Suppose one person knows some information usually called *rumor* but the others do not know who is this initially informed person. How fast can he

effectively broadcast the rumor to all remaining people?

Suppose that each person can call anyone, i.e., they form a complete network. In the most naive approach, the initially informed person can call all the people one by one. We naturally suppose that nobody can perform more than one outgoing call simultaneously. Thus, it takes time $O(n)$ to inform all people in the network. Yet a faster deterministic solution is possible, e.g., the first person calls two people and asks each one to inform a half of the remaining people, they do the same etc. In the complete network everybody will be informed in time $O(\log n)$ and it is not possible to make it faster (up to a constant factor) using only pairwise communications. Nevertheless, the deterministic method has several main disadvantages. First, each player must know for whom he is responsible, which requires additional memory and reduces the scalability of the algorithm. Second, such an algorithm is not robust against communication failures. If one of the first communications fails, $\Theta(n)$ players will never receive the information.

The both disadvantages may be overcome using the *rumor spreading* or *gossip* process. The basic idea that all nodes act independently of the others: at each moment they communicate with a neighbor chosen at random trying to forward or retrieve the rumor. In the most basic and best-studied example of such process called *push protocol* ([DGH$^+$87], [FPRU90], etc.) the rumor is propagated as follows: at each time step, every person that knows of the rumor chooses one of its neighbors uniformly at random, and informs it of the rumor. The push protocol needs $O(\log n)$ time steps or *rounds* to inform all people in the network. Such algorithm is simple and perfectly scalable: players do not need any additional memory or information about the network topology. Finally, the push protocol is also robust against communication failures.

In this work we discuss the common properties of different rumor spreading algorithms. Note that by rumor spreading we mean a distributed algorithm that spreads information over some network and satisfies the following properties (see also Chapter II, Section 1.1).

- All nodes are allowed to gossip via pairwise communication only.
- There is some sort of the randomness in the choice of the communication partners.
- The algorithm should be easily scalable on any size or topology of the network.
- The algorithm spreads the rumor reasonably fast compared to the deterministic algorithm.

We introduce two more important examples of rumor spreading. In the *pull protocol* in each round, every person that still does not know the rumor chooses

one of its neighbors uniformly at random and tries to retrieve the rumor from this person. The *push-pull protocol* is the composition of two previous one: both informed and uninformed players make calls trying to forward (push) or to retrieve (pull) the rumor. We use the word "*call*" to name the communication. Thus, the call initiated by informed player is *push* call, otherwise it is *pull* call.

Note that in this work we consider only synchronized rumor spreading where players call simultaneously in discrete time steps called *rounds*. Nevertheless, the asynchronous model also exists. In that case each node decides to make a call according to its inner clock independent on other nodes. The asynchronous versions of basic algorithms are presented in literature ([BGPS06], [ACMW14], [FPS12]), but less studied than synchronous ones. We discuss the difference between the synchronous and asynchronous model in Chapter II, Section 1.3).

2 Motivation

In this section we discuss some known applications of the rumor spreading processes. Although the basic goal of the gossip process is rumor-mongering, i.e., efficient broadcast of the information, we can find similar processes in different settings such as database maintenance or some distributed search protocols. Jelasity in his review [Jel11] highlighted three main branches of the applications: rumor-mongering, anti-entropy protocols and protocols that compute aggregates. We briefly discuss below these areas as well as the other applications, e.g., peer-to-peer membership management or modelling of the social networks.

Rumor-mongering: It is the most pure and natural application of rumor spreading. The gossip is used to spread information working by flooding agents in the network. All epidemic algorithms are perfect examples of rumor-mongering, as they involve nodes in process in some gossip-based way, rather similar to the way that a viral infection spreads in a biological population. Ideally they should produce bounded worst-case loads.

One of the key problem of rumor-mongering is to decide when everybody knows of the rumor, i.e., when informed players should stop flooding the network. This *stoppage problem* has been studied by Karp et al. [KSSV00] who proposed the median counter rumor spreading protocol. We proposed and analyzed *single incoming call protocol* which is another solution to the stop-problem ([DK17] or Chapter IV, Section 3). Also, we discussed some further solutions in Chapter VI, which are much simpler than the median counter rumor algorithm and provide the very similar performance.

Anti-entropy protocols: The gossiping can be used for repairing replicated data. In this case it may be hard to define a rumor, the core idea is that nodes communicate by comparing replicas and reconciling differences. Demers et al. [DGH$^+$87] were first who proposed an algorithm for maintenance of replicated databases that does not suffer from centralized control and guarantees that all updates reach all sites with high probability.

Protocols that compute aggregates: Kempe, Dobra, and Gehrke [KDG03] were the first who analyzed the protocols for computation of sums, averages, random samples, quantiles, and other aggregate functions, which are computable by fixed-size pairwise information exchanges, by sampling information between the nodes and combining their values. They showed that these protocols converge exponentially fast to the true answer when using uniform gossip, i.e., when each node can contact each other with the same probability.

Another known application of computing the aggregates is peer-to-peer content search within a decentralized and unstructured network such as Gnutella or KaZaA. The idea of the search algorithm is the following. Suppose that nodes of the network contain some set of strings and we are looking for a string which better corresponds some given search pattern queried from some node. Initially we communicate a new query containing some pattern to one of the nodes. Periodically they communicate in gossip manner broadcasting the query. If two communicating nodes are both aware of the pattern, they compare their current found results and both keep the one that better corresponds to the search pattern. The difference from the anti-entropy protocol discussed above is that two different rumor spreading protocols act simultaneously: rumor-mongering for the queries and anti-entropy that computes the most relevant result. That makes challenging the precise analysis of such algorithm.

Nevertheless, this idea has been developed in many works. Stoica et al. [SMK$^+$01] provide a scalable system that can answer queries even if the network is continuously changing, e.g., by joining and leaving nodes. Voulgaris et al.[VKMvS04] and Tang, Xu, and Dwarkadas [TXD03] optimized the search process by creating a semantic overlay, i.e., linking the "semantically close" which are interested in similar documents.

Peer-to-peer membership management: Although basic probabilistic rumor spreading protocols are proven scalable for the message dissemination, they rely on a nonscalable membership protocol, i.e., each node should know every other node. Ganesh, Kermarrec, and Massoulié [GKM03] proposed a fully decentralized gossip-based protocol which provides each node with a partial view of the membership. The core idea is that each new member sends a subscription request to one node

of the network. This node starts gossiping the request. If a node receives the request, then it adds the requesting node to its partial list of the membership with probability depending on the current size of the list, so that the view size is concentrated with high probability around $O(\log n)$, where n is the current number of nodes in the network. Note that the diameter of the network in this model is also bounded by $O(\log n)$ with high probability.

Mathematical theory of epidemics: The mathematical theory of epidemics has been started many years before the idea of using epidemics for information spreading. The logarithmic estimates for the time elapsed between infecting the first individual and involving the whole population (which is supposed to form a complete network) are known for a long time [KM27]. The key difference is that in epidemic theory people substantially study the continuous approximation of the discrete epidemic process and the corresponding differential equations. Nevertheless the epidemic models can be more sophisticated and allow more states for each individual (infected, not infected, dead, cured, immune, etc.) than the rumor spreading processes in which nodes typically have only two states: either informed or uninformed. A great overview of the area is provided in the book by Norman T. J. Bailey [Bai57].

The first attempt to connect epidemics and the dissemination of information refers to Goffman and Newill [GN64]. They proposed to use epidemic processes to design an information retrieval system as an aid to a given population of scientists.

In one of the core articles on rumor spreading Demers et al. [DGH$^+$87] use the mathematical theory of epidemics to relate the number of gossip targets to the fraction of group members who eventually receive the gossip message in the setting with non-reliable communications, i.e., the probability that an arbitrary group member will receive the message.

Modelling of social networks: Since rumor spreading protocols are inspired by human communication, they should simulate reasonably well the behavior of information in social networks. Also, better understanding the propagation of information in social networks will open new ways in the notification systems from advertising to alerting.

In the work [DFF12c] some numerical results were obtained for the case of preferential attachment graph, random attachment and complete graph. As a result, complete and random-attachment graphs showed the same rate of rumor spreading. Orkut network is described very well by rumor spreading in preferential attachment graph which is a little faster than two previous cases. But in Twitter network the rumors spread surprisingly much faster than all previous cases.

3 State of the Art

One of the core problem for studying rumor spreading protocols is to compute the spreading time, i.e., the time elapsed since the appearance of the rumor and before all nodes are informed. Since the spreading time is a random variable, we distinguish between its expectation, i.e., *average spreading time*, and *guaranteed spreading time* – the value that bounds the spreading time with high probability. In this section we summarize the known results for the synchronous rumor spreading in the complete or dynamic graphs with n vertices being the starting point of our work.

Classic protocols, robustness: We start from the most basic push, pull, and push-pull protocols. We recall that in the *push protocol*, in each round each informed node calls a random node and sends a copy of the rumor to it. In the *pull protocol*, in each round each uninformed node calls a random node and tries to obtain the rumor from it. In the *push-pull protocol*, all nodes contact random ones and send the rumor in the direction needed.

For all three protocols precise results are well-known The best known bound for the push protocol was obtained by Doerr and Künnemann [DK14]:

$$\lfloor \log_2 n \rfloor + \ln n - 1.116 \leq \mathbb{E}[T] \leq \lceil \log_2 n \rceil + \ln n + 2.765 + o(1).$$

Robustness properties for the basic push protocol against the communication failures are known. Doerr, Huber, and Levavi [DHL13] analyzed the rumor spreading process in which each communication can be lost with constant probability p. They proved that the rumor spreading time differs from the bound above by a constant factor depending on p (see the second row of Table 1). Some other theoretical works study the robustness against the Byzantine failures[1]. Hence, Malkhi et al. [MRRS01] showed that if b nodes suffer from the Byzantine failure, then it is possible to inform all remaining nodes b times slower than in the reliable case using the simple idea that a node becomes informed if it receives the same rumor from $b + 1$ different sources.

For the push-pull protocol we do not know any analysis precise apart from an additive constant. Nevertheless, Karp et al. [KSSV00] estimated the guaranteed spreading time by $T = \log_3 n \pm O(\log \log n)$. They also presented the *median-counter algorithm* – an improved version of the classic push-pull protocol having the same runtime in the complete graph, but guaranteeing that all calls will be stopped a few rounds later after the last node is informed. They showed that

[1]Byzantine fault tolerance refers to the well-known Byzantine Generals' Problem. In the rumor spreading we suppose that some nodes do not respect the protocol trying to sabotage the rumor spreading. Clearly, the node that initiates the rumor spreading should be reliable

	no transmission failures	calls fail indep. with prob. $p \in (0,1)$
push protocol	$\mathbb{E}[T] = \log_2 n + \ln n \pm O(1)$ $\lfloor \log_2 n \rfloor + \ln n - 1.116 \leq \mathbb{E}[T] \leq \lceil \log_2 n \rceil + \ln n + 2.765 + o(1)$ [DK14]	$\mathbb{E}[T] = \log_{1+p} n + \frac{1}{p} \ln n \pm O(1)$ $T = \log_{1+p} n + \frac{1}{p} \ln n \pm o(\log n)$ whp. [DHL13]
pull protocol	$\mathbb{E}[T] = \log_2 n + \log_2 \ln n \pm O(1)$	$\mathbb{E}[T] = \log_{1+p} n + \frac{1}{\ln \frac{1}{1-p}} \ln n \pm O(1)$
push-pull protocol	$\mathbb{E}[T] = \log_3 n + \log_2 \ln n \pm O(1)$ $T = \log_3 n \pm O(\log \log n)$ whp. [KSSV00]	$\mathbb{E}[T] = \log_{1+2p} n + \frac{1}{p + \ln \frac{1}{1-p}} \ln n \pm O(1)$

Table 1: New (red) and previous-best results for rumor spreading time T of the classic rumor spreading protocols in complete graphs on n vertices. The first line of each table entry contains the result that follows easily from the method proposed in this work, the second line states the best previous result (if any).

assuming up to F *node failures*, the median-counter algorithm informs all but $O(F)$ nodes in $O(\ln n)$ rounds with high probability.

Surprisingly, the basic pull protocol is not really studied in the literature despite its simplicity. Anyway, using the same arguments as in [KSSV00], one can easily see that the guaranteed spreading time for the basic pull protocol is equal to $\log_2 n + O(\ln \ln n)$.

Multiple calls: The variation of the basic push protocol when nodes are allowed to make more than one call was first mentionned by Pittel [Pit87] by the remark that his original proof is suitable for the setting when each informed node makes a constant number $c \geq 1$ simultaneous independent random calls. The guaranteed rumor spreading time in such setting is $\log_c n + c^{-1} \ln n + O(1)$. A general case was analyzed by Panagiotou, Pourmiri, and Sauerwald [PPS15] who proposed a variation of the classic protocols in which the number of calls (always to different nodes) each node performs when active is a positive random variable R. They mostly assume that for each node, this random number is sampled once at the beginning of the process. For the case that R has constant expectation and variance, they show that the rumor spreading time of the push protocol is $\log_{1+\mathbb{E}[R]} n + \frac{1}{\mathbb{E}[R]} \ln n \pm o(\log n)$ with high probability and that the rumor spreading time of the push-pull protocol is $\Omega(\log n)$ with probability $1 - \varepsilon$, $\varepsilon > 0$. When R follows a power law with exponent $\beta = 3$, the push-pull protocol takes $\Theta(\frac{\log n}{\log \log n})$ rounds, when $2 < \beta < 3$, it takes $\Theta(\log \log n)$ rounds.

Dynamic networks: Clementi et al. [CCD+16] have shown that when the network in each round is a newly sampled Erdős-Rényi $G(n,p)$ random graph for some fixed p, then with high probability the rumor spreading time is $\Theta(\log n/(\hat{p}))$,

where $\hat{p} = \min\{p, 1/n\}$. For the more general case that the network is edge-Markovian graph[2] with constant parameters p, q, Clementi et al. have shown that the rumor spreading time for the push protocol is $\Theta(\log n)$, even if the network is disconnected at every time step.

Answering single calls only: While in all protocols above (apart from the one of [PPS15]) it is assumed that each node can call at most one other node per round, we tacitly suppose in the pull and push-pull protocols that nodes can answer all incoming calls. For complete graphs with n vertices, the classic balls-into-bins theory[3] immediately gives that in a typical round there is at least one node that receives $\Theta(\frac{\log n}{\log \log n})$ calls. So unlike for the outgoing traffic, nodes are implicitly assumed to be able to handle very different amounts of incoming traffic in one round.

The first to discuss this issue are Daum, Kuhn, and Maus [DKM15] (also the SIROCCO 2016 best paper). Among other results, they show that if only one incoming call can be answered and if this choice is taken adversarially, then there are networks where a previously polylogarithmic rumor spreading time of the pull protocol becomes $\tilde{\Omega}(\sqrt{n})$. If the choice which incoming call is answered is taken randomly, then things improve and the authors show that for any network, the rumor spreading times of the pull and push-pull protocol increase by at most a factor of $O(\frac{\Delta(G)}{\delta(G)} \log n)$ compared to the variant in which all incoming calls are answered, where $\Delta(G)$ and $\delta(G)$ denote the largest and smallest degree of G.

Another possible solution was proposed by Kiwi and Caro [KC17]: nodes keep all extra pull request in queues and reply them later. They showed that there might be a very significant performance loss if messages are processed at each network node in first-in first-out order. In the worst case the slowdown is linear on the size of the buffer. If we suppose that the buffer is of infinite size, then Kiwi and Caro showed that there exists a maximum degree 4 graph, such that the guaranteed spreading time is $O(n \ln n)$ for the basic pull protocol and $\Omega(1)2^{n/3}$ for the pull protocol in the unbounded buffer model.

Note that both results consider the arbitrary graphs. Apart [DK17], we do not know any tight analysis of these protocols in the complete graphs.

---

[2] i.e., a dynamic graph that every time step evolves by the following rule: any existent edge disappears independently with probability q and any nonexistent one appears independently with probability p. Thus, a newly sampled $G(n, p)$ random graph is the edge-Markovian one with $q = 1 - p$.

[3] The balls-into-bins problem involves m balls and n bins. Each time, a single ball is placed into one of the bins uniformly at random. The problem is what is the maximum load, i.e., the number of balls in a single bin after all balls are in the bins. For the case $m = n$, with high probability the maximum load is $\frac{\log n}{\log \log n} \cdot (1 + o(1))$ [KSC78].

	Multiple outgoing calls $\mathbb{E}[R] = \Theta(1)$, $\mathrm{Var}[R] = O(1)$	Single incoming call
push protocol	$\mathbb{E}[T] = \log_{1+\mathbb{E}[R]} n + \frac{1}{\mathbb{E}[R]} \ln n \pm O(1)$ $T = \log_{1+\mathbb{E}[R]} n + \frac{1}{\mathbb{E}[R]} \ln n \pm o(\log n)$ whp. [PPS15]	—
pull protocol	—	$\mathbb{E}[T] = \log_{2-1/e} n + \log_2 \ln n \pm O(1)$ $T = O(\log^2 n)$ whp. [DKM15]
push-pull protocol	$\mathbb{E}[T] = \log_{1+2\mathbb{E}[R]} n + \log_{1+\ell} \ln n \pm O(1)$, if $\ell > 0$, otherwise, $\mathbb{E}[T] = \log_{1+2\mathbb{E}[R]} n + \frac{1}{\mathbb{E}[R] - \ln \mathbb{P}[R=0]} \cdot \ln n \pm O(1)$. $\forall \varepsilon > 0$, $T \geq \Omega(\log n)$ w.p. $1-\varepsilon$ [PPS15]	$\mathbb{E}[T] = \log_{3-2/e} n + \frac{1}{2} \ln n \pm O(1)$ without stopping time; $\mathbb{E}[T] = \log_{3-2/e} n + \frac{1}{2} \ln n \pm O(1)$ for the stopping time $R = \log_{3-2/e} n$

Table 2: New (red) and previous-best results for the spreading time T of the variations of the classic rumor spreading protocols in complete graph on n vertices. By ℓ for the multiple outgoing call push-pull protocol, we denote the smallest nonnegative integer such that $\mathbb{P}[R = \ell] > 0$.

4 Our Contribution

The typical analysis of a rumor spreading protocol in the papers cited above is based on the same idea, but uses different technical arguments. The idea is that any reasonable rumor spreading process in the complete graph can be split into two parts or *phases* with different behavior. While most of the network is uninformed, rumor spreads exponentially, i.e., each round the number of informed nodes multiplies by almost a constant. This is the *exponential growth* phase. Then, since most of the nodes are informed, the rumor spreading slows down. This phase is called *shrinking* phase, because in one of the typical scenarios, the number of uninformed nodes shrinks by almost a constant each round.

So far, each rumor spreading algorithm on each network topology was analyzed with individual arguments relying heavily on the particular rumor spreading algorithm. Even in fully connected networks (complete graphs), the existing analyses for the basic push protocol [FG85, Pit87, DK14], the push protocol in the presence of transmission failures [DHL13], the push protocol with multiple calls [PPS15], and the push-pull protocol [KSSV00] all employ highly specific arguments that cannot be used for the other processes. Note that the typical analysis of a rumor spreading protocol in the papers cited above needs between six and eight pages of proofs. Clearly, this hinders a faster development of the field.

In this work, we make a big step forward towards overcoming this weakness.

We propose a *general analysis method* for all symmetric and memoryless rumor spreading processes in complete networks. It allows to easily analyze all rumor spreading processes mentioned above and many new one. The key to this generality is showing that the rumor spreading times for these protocols are determined by the probabilities p_k of a new node becoming informed in a round starting with k informed nodes together with a mild bound on the covariance on the indicator random variables of the events that new nodes become informed. Consequently, all other particularities of the protocol can be ignored. The precise definition of phases, and also the bounds for the covariance numbers c_k are discussed in Chapter II, Section 2. Despite this generality, our method gives bounds for the expected rumor spreading time that are *tight apart from an additive constant number of rounds*. Such tight bounds so far have only been obtained once, namely for the basic push protocol [DK14]. We use our method to obtain the following particular results.

Classic protocols: For the three basic push, pull, and push-pull protocols, both in the fault-free setting and when assuming that calls fail independently with probability $1 - p$, our method easily yields the expected rumor spreading times given in Table 1. The comparison with the previous-best result (also given in the table) is not immediately obvious since most previous works obtained bounds that hold with high probability [4] (with probability $1 - o(1)$ or better). Conversely, we prove that the probability that the rumor spreading times deviates from its expectation by more than r rounds is exponentially small in r. Nevertheless, choosing $r = \omega(1)$, we see that our results also hold with high probability.

Note that for half of the settings regarded in Table 1 no previous result existed. In particular, we are the first to find that the double logarithmic shrinking phase observed by Karp et al. [KSSV00] for the pull and push-pull protocols disappears when messages fail with constant probability. This increases the message complexity of the push-pull protocol from the theoretically optimal $\Theta(n \log \log n)$ value to an order of magnitude of $\Theta(n \log n)$ as observed also for the push protocol (see the discussion following the proof of Theorem IV.6 for more details).

Multiple calls: The model of Panagiotou, Pourmiri, and Sauerwald [PPS15] makes sense when assuming that nodes have generally different communication capacities, i.e., the number of calls each node performs when active is defined at the beginning of the process. To model momentarily different capacities, e.g.,

[4] The meaning of "with high probability" is slightly different in different papers. Thus, it means with probability at least $1 - O(n^{-c})$, where c is an arbitrary constant for Karp et al. [KSSV00], and with probability $1 - O\left(n^{-h(n)}\right)$ with $h(n) = o(1)$ arbitrary slow for Doerr et al. [DHL13].

caused by loads with other tasks, we assume that the random variable is resampled for each node in each round. We also allow R to take the value 0. Again for the case $\mathbb{E}[R] = \Theta(1)$ and $\mathrm{Var}[R] = O(1)$, we show that the expected rumor spreading time of the push protocol is $\log_{1+\mathbb{E}[R]} n + \frac{1}{\mathbb{E}[R]} \ln n \pm O(1)$ (see Table 2). The rumor spreading time of the push-pull protocol depends critically on the smallest value ℓ which R takes with positive probability. If $\ell = 0$, that is, with constant probability nodes contact no other node, then there is no double exponential shrinking and the expected rumor spreading time is $\log_{1+2\mathbb{E}[R]} n + \frac{1}{\mathbb{E}[R]-\ln \mathbb{P}[R=0]} \ln n \pm O(1)$. If nodes surely perform at least one call, then we have a double exponential shrinking regime and an expected rumor spreading time of $\log_{1+2\mathbb{E}[R]} n + \log_{1+\ell} \ln n \pm O(1)$.

Dynamic networks: As a proof of concept, we also show that our method is capable of analyzing dynamic networks when the dynamic is memory-less. To see how our method copes with more dependent network structures, we regard the case that the network in each round is a newly sampled 2-regular[5] simple graph. For this scenario, we show that the push protocol has an expected rumor spreading time of $\log_2 n + \log_4 n \pm O(1)$, the pull protocol takes $\log_2 n + \log_2 \ln n \pm O(1)$ rounds, and the push-pull protocol finishes after $\log_{5/2} n + \log_2 \ln n \pm O(1)$ rounds. Interestingly, the push protocol profits from the dynamicity of the network (compared to the complete graph), whereas the push-pull protocol needs a longer time by a constant factor.

For the case when the network is a newly sampled $G(n,p)$ random graph, we sharpen the result of Clementi et al. [CCD+16] for the most interesting regime that $p = \frac{a}{n}$, a is a positive constant. For this case, we show that the expected rumor spreading time is $\log_{2-e^{-a}} n + \frac{1}{1-e^{-a}} \ln(n) + O(1)$. Our tail bound $\mathbb{P}[|T - \mathbb{E}[T]| \geq r] \leq A' \exp(-\alpha' r)$ for suitable constants $A', \alpha > 0$ implies also the large deviation statement of [CCD+16] (where for $\Theta(\log n)$ deviations in the lower tail the trivial $\log_2(n)$ lower bound holding with probability 1 should be used).

Answering single calls only: We finally use our method to discuss an aspect mostly ignored by previous research. While in all protocols above (apart from the one of [PPS15]) it is assumed that each node can call at most one other node per round, it is tacitly assumed in the pull and push-pull protocols that nodes can answer all incoming calls. For complete graphs on n vertices, the classic balls-into-bins theory immediately gives that in a typical round there is at least one node that receives $\Theta(\frac{\log n}{\log \log n})$ calls. So unlike for the outgoing traffic, nodes are implicitly assumed to be able to handle very different amounts of incoming traffic in one round.

[5] A regular graph is a graph where each vertex has the same degree.

With our generic method, we can easily analyze this aspect of rumor spreading on complete graphs. We consider the process proposed by Daum et al. [DKM15], when only one incoming call can be answered and this choice is taken randomly. While for the pull protocol only the growth phase mildly slows down, giving a total expected rumor spreading time of $\mathbb{E}[T] = \log_{2-1/e} n + \log_2 \ln n \pm O(1)$ (see the second row in Table 2), for the push-pull protocol also the double logarithmic shrinking phase breaks down and we observe a total runtime of $\mathbb{E}[T] = \log_{3-2/e} n + \frac{1}{2} \ln n \pm O(1)$ and, similarly as for the push-pull protocol with transmission failures, an increase of the message complexity to $\Theta(n \log n)$. The reason, as our proof reveals, is that when a large number of nodes are informed, then their push calls have little positive effect (as in the classic push-pull protocol), but they now also block other nodes' pull calls from being accepted. This problem can be overcome by changing the protocol so that informed nodes stop calling others when the rumor is $\log_{3-2/e} n$ rounds old. The rumor spreading time of this modified push-pull protocol is $\mathbb{E}[T] = \log_{3-2/e} n + \log_2 \ln n \pm O(1)$ and, when halted at the right moment, this process takes $\Theta(n \log \log n)$ messages.

5 Plan of the Work

We start by providing an extended state of the art for rumor spreading processes in Chapter II, Section 1. We make a brief overview of known facts about the rumor spreading. In particular, we discuss some subtle details of gossiping in the complete graph in Section 1.5 of the same chapter. Then, in Chapter II, Section 2, we provide the complete formulation of our method of analysis. Our main result is contained in Theorem II.14, II.15, and II.16. These theorems provide three criteria of the different behavior of the rumor spreading process depending only on *success probabilities* and *covariance numbers* which are the global characteristics depending only on the number of currently informed nodes (see Definition II.12).

In Chapter III, we do all necessary technical work to prove Theorems II.14, II.15, and II.16. For each theorem we separate upper and lower bounds to the corresponding sections.

In Chapter IV, we justify the strength of our method by analyzing different variants of push, pull and push-pull rumor spreading protocols and adjusting the existing runtime estimates in many cases (see Table 1 and 2). The classic versions of the protocols are discussed in Section 1, the protocols with possible transmission failures – in Section 2.1, the protocols allowing multiple outgoing call per node – in Section 2.2. We discuss the classic protocols running on the dynamic graphs in Section 2.3. Finally, we discuss the limitation on the number of incoming calls in Section 3.

In Chapter V, we discuss how is it possible to overcome the memorylessness limitation of our method. We show that our method might be generalized to the analysis of multi-parametric protocols that require a constant amount of memory in addition to the number of informed nodes for the description of their states. In particular, we propose several of protocols which cannot be analyzed by the current version of our method because the probability numbers depend not only on the number of informed nodes but also from one additional parameter. In Section 3, we provide the exponential growth conditions for the multi-parametric rumor spreading. Unlike the exponential growth and shrinking conditions from previous chapters, these ones are mostly a plan of the lemmas that should be proven to conclude the rumor spreading time.

Finally, in Chapter VI, we provide a brief outlook of our work and discuss the problems that are left open.

Chapter II

Overview of Gossip Protocols

1 Classic Results

In this section we briefly review what is known about the rumor spreading processes. We start by clarifying in which processes refer to rumor spreading and stating three basic examples – *push*, *pull*, and *push-pull* protocols. Then, we discuss the difference between synchronous and asynchronous rumor spreading. We provide some classic results for different network topologies and for both models of synchronization, that better illustrates the difference between them. Finally we consider the case of the rumor spreading in the fully connected network. We discuss the ideas of the proofs for the known estimates of the runtime explaining why the precise analysis of the rumor spreading time in the complete graph can is a challenging problem.

1.1 Definition of Rumor Spreading

Despite the intuitiveness, it is hard to define precisely which processes refer to the rumor spreading. We have not found any general definition in the literature, so let us propose the following scheme. The rumor spreading is a special class of dissemination processes in the network. By the dissemination process we mean the following.

Definition (dissemination process)**.** Let G denote a simple and connected graph with n vertices. Initially one vertex of G knows the rumor that has to be conveyed to every other vertex of G. Each node i is controlled by some algorithm A_i which determines how node i spreads the rumor to its neighbors in G. The *spread time* or runtime is the time elapsed before every node knows the rumor.

The set of algorithms A_i determines the concrete process. Very often A_i are the same, so that the behavior of any node determines the whole process. Intuitively, if the behavior of A_i is similar to the "real world gossiping", then they determine a rumor spreading process. More formally, the rumor spreading is a dissemination process which satisfies certain requirements.

Definition (Rumor spreading process)**.** We call dissemination process *gossip-based* (or *rumor spreading*) if the following properties are satisfied.

(i) All interactions between nodes are pairwise. Each node can participate in only limited number of simultaneous communications.

(ii) There is some form of randomness in the choice of communication partners.

(iii) The information exchanged during these interactions is of bounded size.

(iv) The interaction frequency is low compared to typical message latencies so that the protocol cost is negligible.

The nature of the requirements is the following. First, (i) forbids any node to instantly broadcast the rumor to all its neighbors. Thus, communications remind the phone *calls*, hence we name them so.[1] Second, (ii) requires the rumor spreading to be nondeterministic. Due to this condition the reasonable gossip protocols are simpler and better scalable than their deterministic competitors. In addition randomized gossip processes are robust, i.e., they work well even if reliable communication is not assumed or there are some node failures in the network (see [FPRU90]). By (iii) and (iv), the duration of each communication is negligible, i.e., no overflow is possible (see discussion in the Section 3).

In this work we are interested by the runtime analysis for the different rumor spreading protocols. By the runtime or *rumor spreading time* we mean the time elapsed since the rumor appeared in the network until all nodes know of the rumor (see Definition II.13 for the synchronous model). The rumor spreading time is a random variable, so we are mostly interesting in the following characteristics.

Definition (Guaranteed and average rumor spreading times)**.** *Guaranteed* spreading time is the smallest deterministic number t such that for any choice of initially informed vertex, the whole graph will be informed with high probability after the time t. *Worst average spreading time* is the smallest deterministic number t such that for every choice of initially informed vertex, the expected time until the whole graph is informed is at most t.

[1] In most of the existing works, while the number of outgoing calls performed by a single node is limited, it is supposed that any node can reply all incoming requests. We discuss this assumption in Chapter IV, Section 3.

Remark.

(i) By high probability we understand the probability at least $1 - o(1)$.

(ii) If it does not creates the ambiguity especially for the homogeneous rumor spreading studied so far, we will omit the word "worst" and say simply average or expected spreading time.

It is worth to mention that the definition above is suitable for both synchronous and asynchronous models. The main difference between them is the following.

Synchronous model: Suppose that all nodes share a clock which rings at the discrete time steps. When the clock rings, a new round begins and nodes communicate. All communications during the same round are considered simultaneous. Nevertheless, simultaneous communications are independent, i.e., if x calls y and y calls z in the same round but only x knows the rumor, then z will remain uninformed at the end round. If we assume that the time gap between rounds is 1, then the rumor spreading time is simply the number of rounds passed before all nodes are informed.

Asynchronous model: Each node decides to make a communication independently from the others according to its inner clock typically represented by a Poisson random variable of rate 1 [ACMW14] (see Section 1.3 of this chapter). Such model is also known as Richardson's model [Ric73] after the first who solved the problem proposed by Eden [Ede61] about the asynchronous rumor spreading on the square lattice.

1.2 Independent Random Phone-Call Model

Examples of Rumor Spreading Processes

In both synchronous and asynchronous models, the most classic gossip processes refer to the *independent random phone-call model* [DGH$^+$87], [BEPS14], [PPS15], where each node chooses neighbors to call uniformly at random and independently[2] from other nodes. Within each call the rumor spreads in a natural way, i.e., if the call is established between informed and uninformed node, then one of them forwards the rumor to another. If both nodes are of the same status, then nothing

[2]Another possible behavior is quasi-random rumor spreading: we assume that each node has a cyclic list of its neighbors. Nodes make calls with respect to the order of the list starting from a random position on it. Doerr et al. [DFS08] showed that that, irrespective of the orders of the lists, the quasi-random push protocol succeeds to inform all nodes of the complete graph of size n in time $O(\log n)$.

happens. To distinguish two different mechanisms of informing we say that outgoing call of an informed node is *push* call. Otherwise it is a *pull* call. Restricting the rumor spreading protocol to push-only or pull-only call mechanisms, we obtain push and pull protocol correspondingly. If both call mechanisms are allowed, we say that the rumor spreading follows the push-pull protocol.

Definition. *Push protocol* is the rumor spreading process in which only informed nodes can make calls. *Basic push protocol* is the synchronous push protocol in which each informed node makes exactly one call per round according the independent random phone-call model.

Definition. *Pull protocol* is the independent call gossip process in which only uninformed nodes can make calls. *Basic pull protocol* is the synchronous pull protocol in which each uninformed node makes exactly one call per round according the independent random phone-call model.

Definition. *Push-pull protocol* is the independent call gossip process in which each round every node is allowed to make calls. *Basic push-pull protocol* is the synchronous push-pull protocol in which each node makes exactly one call per round according the independent random phone-call model.

The basic versions of push, pull, and push-pull protocols are the simplest models of rumor spreading processes. However, the application of basic protocols is often challenging and leads to certain generalizations.

Transmission failures: Usually the communications are not reliable. This environment can be modeled by supposing that each call can be lost with respect to some independent Bernoulli random variable. Doerr et al.[DHL13] analyzed the push protocol where each call can be lost with the same probability $1 - p$. In Chapter IV, Section 2.1, we also discuss the pull and push-pull protocols.

Multiple calls: Different nodes may have different connection capacities and might participate in different number of communications per round. Panagiotou et al. [PPS15] proposed the gossip process where each node can call a random number of peers per round with respect to some distribution shared by all nodes. They considered two versions of this process. In the first one, these random numbers are sampled in the beginning and stay constant for each node. In the second one, the number of communications re-samples each round for each node. We discuss the second version in Chapter IV, Section 2.2.

Stoppage problem for push calls: In the basic push and push-pull protocols, informed nodes make calls eternally regardless if everybody knows the rumor or not. One of the possible solution is the median counter algorithm [KSSV00] – a variance of the push-pull protocol in which informed nodes terminates the rumor spreading since they are principally called by informed nodes rather than by uninformed ones.

Multiple incoming pull calls: For the basic pull and push-pull protocols we suppose that if informed node receives several pull requests in one round, then it has enough time to forward the rumor to all its interlocutors. Sometimes it is worth to limit the number of transactions per node in one round by constant. Daum et al. [DKM15] proposed the process where all extra pull request are dropped that is a natural idea for the phone-call model. In Chapter IV, Section 3, we refined their analysis for the complete graph.

1.3 Asynchronous Rumor Spreading

Now we discuss the relations between spreading time for the synchronous and asynchronous rumor speading. The fundamental question is what are the minimum and maximum spreading times in arbitrary n-vertex graph G? The answer was given by Acan, Collevecchio, Mehrabian, and Wormald [ACMW14]. Considering the basic push-pull protocol, they obtained the tight bounds, up to the constant factor cited in this section as Theorems II.1—II.4.

For the guaranteed and average spreading time we will follow the notation of Acan et al. Thus, $gst_a(G)$ and $gst_s(G)$ mean the guaranteed spreading time of graph G in asynchronous and synchronous modes, respectively. Correspondingly, $wast_a(G)$ and $wast_s(G)$ – the worst average spreading time of graph G in asynchronous and synchronous modes. Their first result compares the guaranteed and the worst average spreading time in the asynchronous setting.

Theorem II.1. *The following holds for any n-vertex graph G.*

$$\begin{aligned}(1-1/n)wast_a(G) \le gst_a(G) &\le e\, wast_a(G)\ln n,\\ wast_a(G) = \Omega(\ln n), &\quad wast_a(G) = O(n),\\ gst_a(G) = \Omega(\ln n), &\quad gst_a(G) = O(n\ln n).\end{aligned}$$

Moreover, these bounds are asymptotically best possible, up to the constant factor.

The same apart of the $\Omega(\ln n)$-lower bounds holds for the synchronous rumor spreading.

Theorem II.2. *The following holds for any n-vertex graph G.*

$$(1-1/n)wast_s(G) \le gst_s(G) \le e\, wast_s(G)\ln n,$$
$$wast_s(G) = O(n),$$
$$gst_s(G) = O(n\ln n).$$

Moreover, these bounds are asymptotically best possible, up to the constant factor.

The relationship between the asynchronous and synchronous variants is stated in two following theorem.

Theorem II.3. *For any G we have $gst_a(G) = O(gst_s(G)\ln n)$, and this bound is best possible.*

Theorem II.4. *For any $\alpha \in [0,1[$ we have*

$$gst_s(G) \le n^{1-\alpha} + O\left(gst_a(G)n^{(1+\alpha)/2}\right).$$

Corollary II.5. *We have*

$$\tfrac{gst_s(G)}{gst_a(G)} = \Omega(1/\log n), \qquad \tfrac{gst_s(G)}{gst_a(G)} = O\left(n^{2/3}\right),$$

and the left-hand bound is asymptotically best possible, up to the constant factor. Moreover, there exist infinitely many graphs for which this ratio is exactly $\Omega\left(n^{1/3}(\log n)^{-4/3}\right)$.

We can see that the spreading time for the synchronous and asynchronous environment may significantly differ (see Corollary II.5). In the next section we will provide some examples illustrating this phenomenon. From wide point of view, two main factors should be taken into account when we turn from synchronous to the asynchronous rumor spreading. First, calls can no longer be simultaneous, that can accelerate the rumor spreading. (Let x is informed and y, z are not. If x talks to y and after that y talks to z, then both y and z becomes informed. If the same event happened in one round of the synchronous rumor spreading, then only y would be informed.) Second, since all nodes decide to make call independently according to some random timer, it is very likely that one of the nodes waits for a long time before making the first call. For this reason, the asynchronous push-pull protocol is slow if G is a star graph (see the corresponding example below).

1.4 Path, Square Lattice, Star, Necklace

In this section we illustrate the difference between different modes of the rumor spreading providing some classic results for different graphs.

Before getting started with the concrete examples, we recall the most general upper bound for the spreading time proved by Feige et al. [FPRU90].

Theorem II.6. *Consider a synchronous basic push protocol in graph G with n vertices. With probability at least $1 - 1/n$, all nodes in G are informed after $O(\Delta(G)(\mathrm{diam}(G) + \ln n))$.* [3]

Sketch of the proof. The proof is based on the following simple idea. Consider two adjacent nodes x and y. If one of them is informed, then the probability that the rumor traverses the edge xy is at least $1/\Delta(G)$. Therefore, the probability that the rumor will not traverse the path of the length k in $3\Delta(G)(k + 2\ln n)$ rounds is bounded by the Chernoff bound by $1/n^2$ (using the same arguments as in the proof of Proposition II.7). Since the distance from the initially informed node to any other vertex is at most $\mathrm{diam}(G)$, the probability that the rumor does not reach all vertices in time $O(\Delta(G)(\mathrm{diam}(G) + \ln n))$ is bounded by $1/n$. □

The examples below show that this bound is tight. Also note that the same upper bound holds for the basic pull and push-pull protocols since we did not use in the proof the exact mechanism of calls. As for the lower bound, Feige et al. [FPRU90] proved that the guaranteed rumor spreading time at least $\ln n$, that is tight within a constant factor since $O(\ln n)$ iterations suffices for the complete graph.

Example 1. Path

The simplest example is when G is a path of length n. Theorem II.6 claims that the guaranteed rumor spreading time is $O(n)$. This bound is tight.

Proposition II.7. *Let G be a path of length n. For any of synchronous basic push, pull, and push-pull protocols, the guaranteed rumor spreading time is equal to $\Theta(n)$.*

Proof. For concreteness, consider the push protocol. Observe that the set of informed nodes is always a subpath of G, and at any round the path elongates by at most one node from each end. Therefore, the rumor spreading time is at least $n/2$.

Without loss of generality, suppose that the initially informed node is one of G's endpoints, so that the rumor spreads only in one direction. Let X_i be a random indicator variable for an event that at round i rumor moves forward. Therefore, the rumor spreading time is the smallest t such that $X_1 + \ldots + X_t \geq n - 1$. Let $X := X_1 + \ldots + X_{4n}$. Since $\mathbb{P}[X_i = 1] = 1/2$ for any i, we have $\mathbb{E}[X] = 2n$. Since all X_i are independent, we have $\mathrm{Var}[X] = O(n)$. Applying Chebyshev' inequality to X we obtain

$$\mathbb{P}[|X - 2n| \geq n] \leq \tfrac{\mathrm{Var}[X]}{n^2} = O\left(\tfrac{1}{n}\right).$$

[3] By $\Delta(G)$ we denote the maximum degree of the vertices in G. $\mathrm{diam}\, G$ is diameter of G.

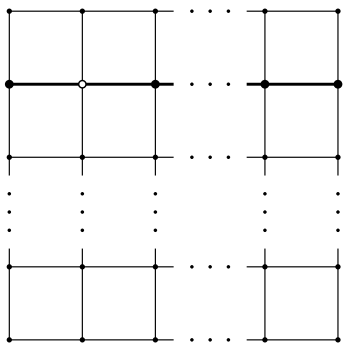

Figure 1: The regular square lattice. All nodes in the horizontal bold path containing the initial node get informed after $O(\sqrt{n})$ rounds with high probability. Then, each of these nodes acts an initial node for the vertical path containing this node, so that the rumor spreading time is $O(\sqrt{n})$ with high probability.

Therefore, after $4n$ rounds at least n nodes will be informed with probability $O\left(\frac{1}{n}\right)$. □

Acan et al. [ACMW14] showed that same bounds holds for the asynchronous rumor spreadings. The proof is similar to the one for the synchronous case. Again, we suppose that the initially informed node is on of endpoints of G. The basic idea is that the spreading time T is equal to $\sum_{e \in E} T(e)$, were $T(e)$ is the communication time via e edge e defined as follows. Let t_1 be the first time that one of e's endpoints gets the rumor, and t_2 is the first time after t_1, when a call goes through e. Then, $T(e) := t_2 - t_1$. Since the nodes make calls according to independent memoryless timers (the exponential random variables of rate 1 are used as timers), $T(e)$ is bounded by the minimum of two random exponential variables. Since the path of length n has $n - 1$ edges, both guaranteed and expected rumor spreading times are equal $\Theta(n)$.

Example 2. Square Lattice

Another classic example for the rumor spreading is square lattice (see Figure 1) with n vertices (we suppose that n is a square number). The diameter of such graph is $\sqrt{n}$ and the maximum degree is 4. So, Theorem II.6 claims that guaranteed rumor spreading time is $O(\sqrt{n})$. It is easy to see that the result is tight up to the constant factor.

Proposition II.8. *Let G be $\sqrt{n} \times \sqrt{n}$ square lattice. For any of synchronous basic push, pull, and push-pull protocols, the guaranteed rumor spreading time is equal to $O(\sqrt{n})$.*

The lower bound for the spreading time is trivial – to inform all nodes we need at least to forward the rumor through a horizontal path containing the initially informed vertex. The proof for the upper bound follows the same steps as the proof of Theorem II.6.

Proof. Any node can send the rumor to at most 4 directions. Consider a horizontal path containing an initially informed node. By the same arguments as for Proposition II.7, all nodes in any horizontal or vertical path will be informed not later than $O(\sqrt{n})$ rounds after that first node in this path has been informed. Since, after $O(\sqrt{n})$ rounds there will be at least one informed node in each of $\sqrt{n}$ vertical paths in graph, the rumor spreading time depends on the longest rumor spreading time in these paths. From the proof of Proposition II.7 follows that for some $c > 0$, the probability that one path is not informed after $c\sqrt{n}$ rounds is at most $1/n$. Since there are $\sqrt{n}$ vertical path which we can consider independently, the probability that after $O(\sqrt{n}) + c\sqrt{n}$ rounds at least one node stays uninformed is bounded by $O(1/n)$. □

The guaranteed rumor spreading time in the asynchronous model is also $O(\sqrt{n})$, the result was first proved by Richardson [Ric73]. The proof is based on the similar idea to the proof for the asynchronous rumor spreading in the path (see Example 1).

The most recent results concerning synchronous rumor spreading in the square lattice were obtained by Fatès and Gerin [FG09].They considered a different process, when a node gets involved with probability α if one of its neighbors is already involved[4]. They showed that the worst expected runtime for such process in n-nodes square lattice is bounded between $\frac{\sqrt{n}}{8\alpha}$ and $3\frac{\sqrt{n}}{\alpha}$.

Example 3. Star

Star is a perfect example to illustrate the difference between synchronous push, pull, and push-pull protocols. In addition, the reduction to the coupon collector problem (see Appendix C), which is also a powerful idea for the analysis of the rumor spreading in the complete graph, first appears in this example. The star G with n vertices has $n - 1$ leaves and a central vertex that is adjacent to every other vertex.

Consider the synchronous rumor spreading first. Suppose that the central vertex is informed initially. In the pull process, all leaves will be informed in one round, because each leaf can call only the central vertex. Unlike, in the push

[4] If $\alpha = 1/4$, then the process almost coincides with the synchronous pull protocol apart from the side-effects on the outer boundary of the graph. For other values of α it is similar to the pull protocol with transmission failures or multiple calls with the corresponding protocol parameters.

protocol, only the central node is active and makes one call per round, trying to reach all leaves. This is equivalent to the *coupon collector* problem with $n-1$ coupons [FGT92]. Therefore, the spreading time for the push process is equal to $\Theta(n \ln n)$ rounds.

Suppose now that initially informed node is a leaf. In the push protocol, the central vertex will be informed next round, then the rest of the graph will be informed after $\Theta(n \ln n)$ rounds with high probability. In the pull process, the average number of rounds until the central vertex gets informed is $\Theta(n)$. However, only after $\Omega(n)$ rounds the central vertex gets informed with high probability. The probability at least $1 - O(1/n)$ can be guaranteed only after $\Theta(n \log n)$ rounds. The pull calls make all the leaves informed in the next round after informing the central vertex.

For the push-pull protocol, the rumor spreading time is equal to 1, if the central vertex is informed initially, otherwise it is equal 2. Therefore, we have the following.

Proposition II.9. *Let G be a star graph with n vertices. Consider one of the basic synchronous push, pull, and push-pull protocols.*

The guaranteed rumor spreading time for the push protocol in G is equal to $\Theta(n \ln n)$ and the worst average one is equal to $\Theta(n \ln n)$.

The guaranteed rumor spreading time for the pull protocol in G is equal to $\Omega(n)$ and the worst average one is equal to $\Theta(n)$.

Both guaranteed and worst average rumor spreading times for the push-pull protocol in G are equal to 2.

Note that Acan et al. showed that both guaranteed and average rumor spreading times for the asynchronous push-pull protocol in G are equal to $\Theta(\ln n)$ using the argument that since the central node is informed, the remaining spreading time is equal to the time the last vertex makes its first call, that is equal to $\Theta(\ln n)$ with high probability.

Example 4. Necklace graph

The last example shows the difference between synchronous and asynchronous push-pull rumor spreading. The necklace graph is constructed as in Figure 2. It consists of m connected diamonds, each diamond consists of k edge-disjoint paths of length 2 with the same end vertices. Acan et al. [ACMW14] claim the following.

Proposition II.10. *Let G be a necklace graph with m diamonds of k paths each. Then the worst average rumor spreading time in G is equal to $O(\ln(km) + mk^{-1/2})$ for the asynchronous push-pull process.*

Sketch of the proof. Consider one of the diamonds in G and suppose that a vertex with degree $2k$ is informed. For an edge e, let $T(e)$ denote the communication time via e (see Example 1 for the definition). For any e, one of endpoints has degree k and another has degree 2. Then $T(e)$ is stochastically dominated by an independent exponential random variable with mean 2. Therefore, the time T_d to spread the rumor from one endpoint of the diamond to another is stochastically dominated by $\min\{X_1 + Y_1, \ldots, X_k + Y_k\}$, where X_i and Y_i are independent $Exp(1/2)$ random variables. Then, by some elementary computations one can obtain that $\mathbb{E}[T_d] = O\left(k^{-1/2}\right)$. Thus, the expected time until all cut vertices in G get the rumor is equal to $O\left(mk^{-1/2}\right)$.

Finally, since all cut vertices are informed, any degree 2 vertex gets the rumor in time $Exp(1)$, so that the expected time until the last node finds out the rumor is $O\left(\ln(km)\right)$. □

Now suppose $k = \Theta\left((n/\ln n)^{2/3}\right)$ and $m = \Theta(n^{1/3}\left(\ln n\right)^{2/3})$. Since the worst average rumor spreading time for the synchronous push-pull protocol is at least $2m$, we obtain that $wast_s(G)/wast_a(G) = \Omega\left((n/\ln n)^{1/3}\right)$. The last expression shows that the bound in Corollary II.5 is tight.

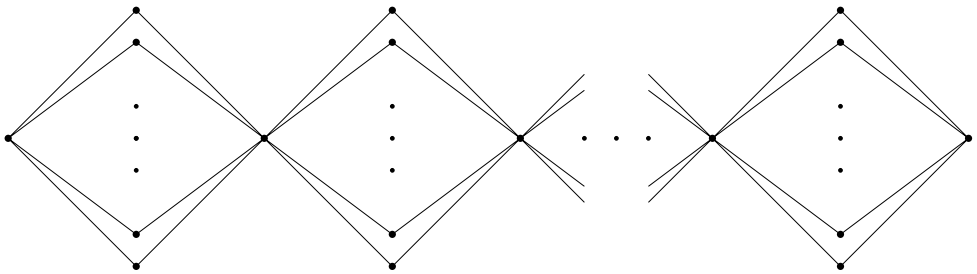

Figure 2: The necklace graph is a "path" of m diamonds, each of them consists of k edge-disjoint paths. The asynchronous push-pull protocol is much faster on this graph than its synchronous variant.

1.5 Complete Graph. Overview of the Proofs

In this section we discuss the important details of previous solutions related to the base problem of this memoir, that is the analysis of the rumor spreading time the in complete graph on n vertices. From Theorem II.6 it follows that the rumor spreading time is at most $O(n \ln n)$ (the same result as for the push protocol in the star graph). However, all reasonable rumor spreading techniques succeed to inform all nodes in time $O(\log n)$ that is the best time possible up to the constant factor for the synchronous phone-call model.

Proposition II.11. *Let G be a complete graph on n vertices. The average spreading time for the basic push protocol in G is $\Theta(\ln n)$.*

Sketch of the proof. Since each round the number of informed nodes at most doubles, the logarithmic lower bound is trivial.

For the upper bound we use the following argument. Suppose that all calls are made one by one and their targets are written in a row. This is equivalent to the coupon collector problem with n coupons[5]already seen in previous section. Thus, the total number of calls until all nodes are informed equals $O(n \ln n)$ with high probability. Therefore, since $\Theta(n)$ nodes are informed, the round-based process finishes up within $O(\ln n)$ rounds.

Now we show that at with high probability least $\Theta(n)$ nodes can be informed within a logarithmic number of rounds. Consider one round starting with $k \leq n/4$ informed nodes. As before, we assume that all calls in this round are made one by one in some arbitrary order. Then, the probability that the ith call reaches a node which was not called before is at least $3/4 - i/n \geq 1/2$. Therefore, the expectation of the number of newly informed nodes is at least $k/2$. Using one of the concentration inequalities[6], one can see that this number is strongly concentrated around its expectation, e.g., at least $k/3$ with the probability which is exponentially small in k. Therefore, the time elapsed until at least $n/4$ nodes are informed is less than $\log_{4/3} n$ that concludes the proof. □

This proof is a simplified version of different existing proofs. However, the proof above reveals a good intuition about the rumor spreading in the complete graph. As the bound of $\Theta(n)$ informed nodes is reached, the process suffers from the phase transition from the exponential growth of the number of informed nodes to the exponential shrinking[7] of the number of uninformed nodes. To illustrate the phenomenon we can see the plot on Figure 3: while the number of informed nodes is relatively small, it doubles each round, but the doubling weakens upon

[5]These two processes coincide if the target of each call is uniformly distributed among all nodes, i.e., if nodes are allowed to call themselves. If the loop calls are forbidden, we can use the similar arguments for the coupon collector process with $n - 1$ coupons.

[6] Classically, Chernoff bound is used that needs the independence between the events corresponding the actions of single nodes. In current proof it suffices to consider an event that node i calls a node different from k informed ones and from $i - 1$ nodes that contains all targets of first $i - 1$ calls in this round instead of ith calls an informed node. All these nodes are independent.

[7] It is slightly technical to show that exponential shrinking of number of uninformed nodes follows directly from the coupon collector reduction. The informal argument is that since $m \leq n/2$ nodes are uninformed, then it will take in expectation n/m calls to inform a new node, then $n/(m+1)$ for the next one, etc. Thus, if we expect to inform j nodes in current round, then the total number of calls, is approximatively equal to $n/m + \ldots + n/(m-j) \sim \ln m - \ln(m-j)$. Since the total number of calls is bounded between $n/2$ and n, one can see that $\frac{m-j}{m}$ is bounded between approximatively e^{-1} and $e^{-1/2}$.

the progress; when most of the nodes are informed, the process informs some non-negligible fraction of uninformed nodes which converges to $1/e$ upon the progress. Thus, the usual way to analyze the rumor spreading in the complete graph is to partition the execution of the process in phases and within each phase to uniformly estimate the progress. Nevertheless, to get a sharper result, one should overcome the following sources of the inaccuracy.

(i) Our estimate for the number of newly informed nodes in the beginning of the process is too rough. When the number of informed nodes is $O(\sqrt{n})$, the collisions between calls are negligible. Thus, the number of informed nodes doubles each round. However, this doubling becomes weaker when more than $O(\sqrt{n})$ nodes are informed. This weakening of the doubling process is one main difficulty in all previous works.

(ii) We consider the second part of the process from $n/2$ informed nodes arguing that the number of calls is at least $n/2$ each round. However, this number increases during the process. Together with (i), these two effects are related with the phase transition that happens when the number of informed and uninformed nodes are both $\Theta(n)$.

(iii) The analysis strongly depends on the concrete process and the arguments cannot be directly applied to, e.g., the pull protocol. Especially, the independence, which is necessary to provide concentration guarantees, is hard to observe in more sophisticated protocols (single-incoming-call protocol from Section 3 and push protocol in dynamic graphs from Section 2.3 discussed in Chapter IV).

Typically, the tighter bounds require more accurate analysis and more phases to consider. The first strong result appears in Frieze and Grimmett's paper [FG85]: by partitioning the execution of the push protocol into 5 phases (see Figure 4), they showed that the guaranteed rumor spreading time for the push protocol in the complete graph is $\log_2 n + \ln n + o(\log n)$. One of their main observations is that the middle phase, when the number of informed nodes is between ξn and $n - \eta n$, requires only $O(1)$ rounds with high probability. Other words, during the rumor spreading process, either the number of informed nodes is small or the number of uninformed nodes is small, with high probability. The rest of the proof was based on the argument that choosing ξ (resp. η) small enough, we can show that the number of rounds within the exponential growth (resp. shrinking) phase is arbitrary close to $\log_2 n$ (resp. $\ln n$), that decreases the inaccuracy term to $o(\log n)$.

Two years later, Pittel [Pit87] gains $O(1)$-precision by considering 7 phases. He argues that first, with high probability the number of informed nodes doubles

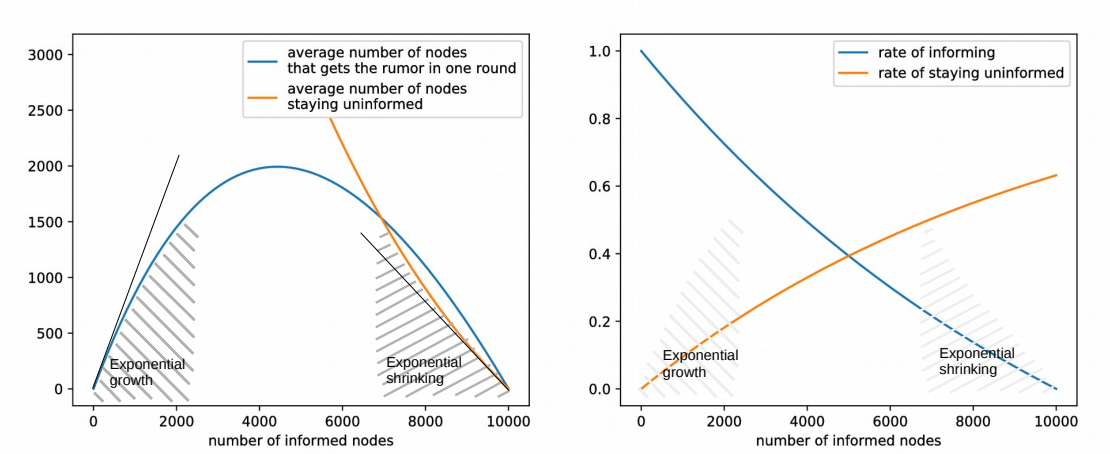

Figure 3: The phase transition for the push protocol. On the left, we plot the average number of newly informed nodes (resp. nodes that stay uninformed) as a function of the number of informed nodes k on the previous step.
On the right, we plot the rate of informing and the rate of staying uninformed. The first is defined as $\alpha := \mathbb{E}[X(k)]/k$ and shows the relative growth of the number of informed nodes. The rate of staying uninformed $\beta := \mathbb{E}[X]/(n-k)$ shows the proportion of newly informed node among all uninformed nodes at the beginning of the current round, i.e., the relative decrease of the number of uninformed nodes.

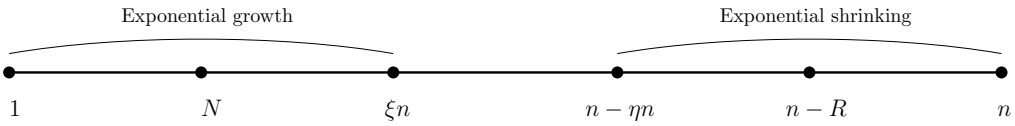

Figure 4: The phase structure of Frieze and Gimmet's proof [FG85]. N and R are fixed large numbers, ξ and η are small and positive.

each round until $n_1 = o(\sqrt{n})$ nodes are informed. Then, until $n_2 = n/\log^2(n)$ nodes are informed, with high probability the number of informed nodes increases by at least a factor of $2(1 - \frac{1}{\log^2(n)})$. Consequently, this second phase lasts at most $\log_{2(1-\frac{1}{\log^2(n)})}(n_2/n_1)$ rounds. The third phase of the exponential growth lasts at most $\log_2(1-\varepsilon_3)(\varepsilon_3 \log^2 n)$ rounds. This, together with two previous ones, implies that with high probability, at least $n\varepsilon_3$ nodes will be informed within $\log_2 n + O(1)$ rounds for sufficiently small ε_3.

The current best result was obtained by Doerr and Künnemann [DK14]. They use only three different phases and explicitly bound the error term between -2 and 3. Such a simplification could be reached using a different technique for the

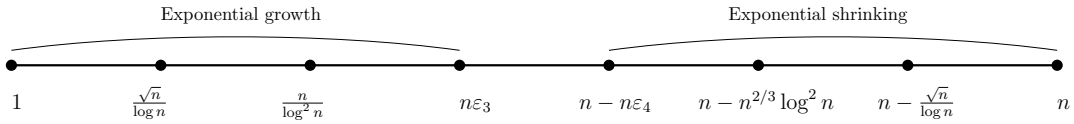

Figure 5: The phase structure of Pittel's proof [Pit87]. $\varepsilon_3 \to 0$ slowly such that $n\varepsilon_3 \geq \frac{n}{\log^2 n}$ and $\varepsilon_4 \to 0$ slowly such that $n\varepsilon_4 \geq n^{2/3}\log^2 n$.

exponential growth, covered by the first two phases. They cut the interval between 1 and $n/2$ informed nodes, in $\log_2 n + O(1)$ sub-phases so that with probability $1 - q_j$ the protocol spends at most one round in phase j. Until $O(\sqrt{n})$ nodes are informed they use the birthday paradox: within the start-up phase no two calls collide with high probability, so the number of informed nodes doubles each round and the phases are split by the numbers $1, 2, 4, \ldots$ of informed nodes. The important observation is that the failure probabilities q_j during the second phase are not necessarily small, but their sum is less than 1. Therefore, the number of rounds spent until $n/2$ nodes are informed is stochastically dominated by $\log_2 n + O(1) + \mathrm{Geom}\left(1 - \sum_j q_j\right)$. Together with coupon collector reduction, this yields the upper bound for the spreading time which is tight up to an additive constant. Thus, Doerr and Künnemann's proof also illustrates that artificially large number of phases is not necessary for precise estimating the spreading time.

Note that a phase-based proof has also been seen for other protocols. For example, Karp et al. [KSSV00] propose a 4-phase analysis of the push-pull protocol, the obtained estimate for the rumor spreading time is tight up to $O(\ln \ln n)$ additive term (see Figure 6). The main difference between push and push-pull protocols in the complete graph is that the last does not suffer from the coupon collector effects: since $\Theta(n)$ nodes are informed, the probability that a node stays uninformed in current round is proportional to the fraction of uninformed nodes. Therefore, the rest of the nodes can be informed within $O(\ln \ln n)$ rounds. Such behavior is common for most of the protocols with pull calls, in Section 2 we will call it the *double exponential shrinking phase*[8].

Note finally, that the phase structure is much less useful for the asynchronous rumor spreadings when the informing of any single node is a distinguish event. However, Janson [Jan99] implicitly proved that the guaranteed rumor spreading time for the same process is $3/2 \ln n + o(\ln n)$. More recently, Acan et al. [ACMW14] showed for the asynchronous push-pull protocol that the time until a new vertex is informed is distributed as the minimum of $k(n-k)$ independent exponential

[8] For the particular case of the basic push-pull protocol, Karp et al. used the term "quadratic shrinking" instead of the double exponential shrinking due to the fact that the fraction of informed nodes squares each round.

random variables each with rate $2/(n-1)$. This implies that the average rumor spreading time is $\ln n + O(1)$.

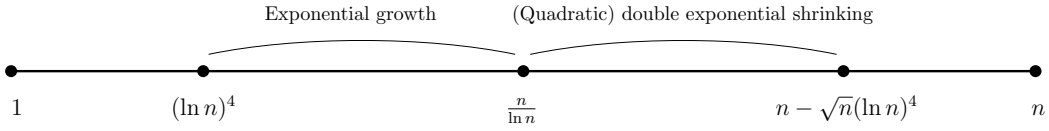

Figure 6: The phase structure of Karp's proof [KSSV00].

2 Precise Statements of Our Results

As just discussed, the main advantages of our approach are its universality and the very tight bounds it proves. We now briefly sketch the main new ideas that lead to this progress. Interestingly, they are rather simpler than the ones used in previous works.

2.1 Tight Bounds via a Target-Failure Calculus

We first describe how we obtain estimates for the rumor spreading time that are *tight apart from additive constants*. Our strategy is close to the one seen in [DK14] briefly discussed is previous section. It is easy to compute the expected number $E(k)$ of newly informed nodes in a round starting with k informed nodes, e.g., for the classic push protocol, $E(k) = k - \Theta(k^2/n)$. For each number k of informed nodes, we formulate a pessimistic round target $E_0(k)$ that is sufficiently below or above the expected number $E(k)$ of newly informed nodes. Here "sufficiently away" means that the probability $q(k)$ to fail the target is small, but not necessarily $o(1)$ as in all previous analyses. When proving upper bounds on rumor spreading times, we see that the expected time to go from k to at least $E_0(k)$ informed nodes is at most $1 + \frac{q(k)}{1-q(k)}$, simply with the argument that in the case of a failure, we can try again (this needs some elementary monotonicity statements for the $q(k)$ and $E_0(k)$).

The second, again elementary, key argument still in the language of the push protocol is that when we define a sequence of round targets by $k_0 := 1$, $k_1 := E_0(k_0)$, $k_2 := E_0(k_1), \dots$ with suitably defined $E_0(\cdot)$, then the k_i grow almost like 2^i. More precisely, there is a $T = \log_2 n \pm O(1)$ such that $k_T = \Theta(n)$. Hence together with the previous paragraph we obtain that the expected number of rounds to reach k_T informed nodes is $\sum_{i=0}^{T-1} 1 + \frac{q(k_i)}{1-q(k_i)} = T + O(\sum_{i=0}^{T-1} q(k_i))$. So it suffices that the sum of the failure probabilities $q(k_i)$ is a constant (unlike in previous works, where it needed to be $o(1)$).

We remark that this target-failure argument was used already in [DK14], there however only to give an upper bound for the runtime of the push protocol in the regime from n^s, s a small constant, to $\Theta(n)$ informed nodes, that is, the later part of the exponential growth regime of the push process, in which via Chernoff bounds very strong concentration results could be exploited. Hence the novelty of this work with respect to the target-failure argument is that this analysis method can be used (i) also from the very beginning of the process on, where we have no strong concentration, (ii) also for the exponential and double exponential shrinking regimes of rumor spreading processes, and (iii) also for lower bounds.

2.2 Uniform Treatment of Many Rumor Spreading Processes

As discussed earlier, the previous works regarding different rumor spreading processes on complete graphs all had to use different arguments. The reason is that the processes, even when looking similar, are intrinsically different when looking at detail. As an example, let us consider the first few rounds of the push and the pull protocol. In the push protocol, we just have just seen that while there are at most $o(\sqrt{n})$ nodes informed, then a birthday paradox type argument gives that with high probability we have perfect doubling in each round. For the pull process, in which each uninformed node calls a random node and becomes informed when the latter was informed, we also easily compute that a round starting with k informed nodes creates an expected number of $(n-k)\frac{k}{n} = k - \frac{k^2}{n}$ newly informed nodes. However, since these are binomially distributed, there is no hope for perfect doubling. In fact, for the first few rounds, we even have a constant probability that no single node becomes informed.

The only way to uniformly treat such different processes is by making the analysis which depends only on general parameters of the process as opposed to the precise definition. Our second main contribution is distilling a few simple conditions that (i) subsume essentially all symmetric and time-invariant rumor spreading processes on complete graphs and (ii) suffice to prove rumor spreading times via the above described target-failure method. All this is made possible by the observation that the target-failure method needs much less in terms of failure probabilities than previous approaches. Consequently, instead of using Chernoff and Azuma bounds for independent or negatively correlated random variables (which rely on the precise definition of the process), it suffices to estimate the number of newly informed nodes via computing the expectation and using Chebyshev's inequality as concentration result. Consequently, to apply our method we only need to (i) understand (with a certain precision) the probability p_k that an uninformed node becomes informed in a round starting with k informed nodes (recall that we

assumed symmetry, that is, this probability is the same for all uninformed nodes) and (ii) we need to have a mild upper bound on the covariance of the indicator random variables of the events that two nodes become informed. Note that for most of the known protocols these events rather tend to be negatively correlated. The probabilities p_k usually are easy to compute from the protocol definition. Also, we do not know them precisely. For example, for the growth phase of the push protocol discussed above, it suffices to know that there are constants $a < 2$ and a' such that for all $k < n/2$ we have $\frac{k}{n}(1 - a\frac{k}{n}) \leq p_k \leq \frac{k}{n}(1 + a'\frac{k}{n})$. This (together with the mild covariance condition to be detailed later) is enough to show that the rumor spreading process takes $\log_2 n \pm O(1)$ rounds to inform $n/2$ nodes or more.

We profit from the fact that seemingly all reasonable rumor spreading processes in complete networks can be described via three regimes:

Exponential growth: Up to a constant fraction fn of informed nodes, $p_k = \gamma\frac{k}{n}(1 \pm O(\frac{k}{n}))$. The number of informed nodes thus increases roughly by a factor of $(1+\gamma_n)$ in each round, hence the expected time to reach fn informed nodes or more is $\log_{1+\gamma_n} n \pm O(1)$.

Exponential shrinking: From a certain constant fraction $u = n - k = gn$ of uninformed nodes on, the probability of remaining uninformed satisfies $1 - p_{n-k} = e^{-\rho_n} \pm O(\frac{u}{n})$. This leads to a shrinking of the number of uninformed nodes by essentially a factor of $\frac{1}{\rho_n}$ per round. Hence when starting with gn informed nodes, it takes another $\frac{1}{\rho_n} \ln n \pm O(1)$ rounds in expectation until all are informed.

Double exponential shrinking: From a certain constant fraction $u = n - k = gn$ of uninformed nodes on, the probability of remaining uninformed satisfies $1 - p_{n-k} = \Theta((\frac{u}{n})^{\ell-1})$. Now the expected time to go from gn uninformed nodes to no uninformed node is $\log_\ell \ln n \pm O(1)$.

Due to their different nature, each of these three regimes should be discussed separately. However, all can be treated with the target-failure method. Hence the main differences between these regimes lie in defining the pessimistic estimates for the targets, computing the failure probabilities, and computing the number of intermediate targets until the goal is reached. All this only needs computing expectations, using Chebyshev's inequality, and a couple of elementary estimates.

2.3 Precise Statement of the Technical Results

In this work, we consider only *homogeneous rumor spreading processes* characterized as follows. We always assume that we have n nodes. Each node can be either

informed or *uninformed*. We assume that the process starts with exactly one node being informed. Uninformed nodes may become informed, but an informed node never becomes uninformed. We consider a discrete time process, so the process can be partitioned into *rounds*. In each round each uninformed node can become informed. Whenever a round starts with k nodes being informed, then the probability for each uninformed node to become informed is p_k, which only depends on the number k of informed nodes at the beginning of the round.

The main insight of this work is that for such homogeneous rumor spreading processes we can mostly ignore the particular structure of the process and only work with the *success probabilities* p_k defined above and the *covariance numbers* c_k defined as follows.

Definition II.12 (Covariance numbers)**.** For a given homogeneous rumor spreading process and $k \in [1..n-1]$ let c_k be the smallest number such that whenever a round starts with k informed nodes and for any two uninformed nodes x_1, x_2, the indicator random variables X_1, X_2 for the events that these nodes become informed in this round satisfy

$$\mathrm{Cov}[X_1, X_2] \leq c_k.$$

Our main interest is studying after how many round all nodes are informed.

Definition II.13 (Rumor spreading times)**.** Consider a homogeneous rumor spreading process. For all $t = 0, 1, \ldots$ denote by I_t the number of informed nodes at the end of the t-th round ($I_0 := 1$). Let $k \leq m \leq n$. By $T(k, m)$ we denote the time it takes to increase the number of informed nodes from k to m or more, that is,

$$T(k, m) = \min\{t - s | I_s = k \text{ and } I_t \geq m\}.$$

We call $T(1, n)$ the rumor spreading time of the process.

As it turns out, almost all homogeneous rumor spreading processes can be analyzed via three regimes.

Exponential Growth Regime

Let γ_n be bounded between two positive constants. Let $a, b, c \geq 0$ and $0 < f < 1$ with $af < 1$. We say that a homogeneous rumor spreading process satisfies the *upper (resp. lower) exponential growth conditions* in $[1, fn[$ if for any $n \in \mathbb{N}$ big enough the following properties are satisfied for any $k < fn$.

(i) $p_k \geq \gamma_n \frac{k}{n} \cdot \left(1 - a\frac{k}{n} - \frac{b}{\ln n}\right)$ resp. $p_k \leq \gamma_n \frac{k}{n} \cdot \left(1 + a\frac{k}{n} + \frac{b}{\ln n}\right)$.

(ii) $c_k \leq c\frac{k}{n^2}$.

In the case of the upper exponential growth condition, we also require $af < 1$.

Theorem II.14. *If a homogeneous rumor spreading process satisfies the upper (resp. lower) exponential growth conditions in* $[1, fn[$*, then*

$$\mathbb{E}[T(1, fn)] \leq \log_{1+\gamma_n} n + O(1) \text{ resp. } \mathbb{E}[T(1, fn)] \geq \log_{1+\gamma_n} n - O(1).$$

There exist $A' > 0, \alpha' > 0$ *such that for any* $r \in N$*,*

$$\mathbb{P}[T(1, fn) > \log_{1+\gamma_n} n + r] \leq A' e^{-\alpha' r} \text{ resp. } \mathbb{P}[T(1, fn) \leq \log_{1+\gamma_n} n - r] \leq A' e^{-\alpha' r}.$$

When the lower exponential growth conditions are satisfied, in addition, there exists an $f' \in]f, 1[$ *such that with probability* $1 - O\left(\frac{1}{n}\right)$ *there are at most* $f'n$ *nodes informed after the round which first reaches* fn *or more informed nodes.*

Exponential Shrinking Regime

Let ρ_n be bounded between two positive constants. Let $0 < g < 1$, and $a, c \in \mathbb{R}_{\geq 0}$. We say that a homogeneous rumor spreading process satisfies the *upper (lower) exponential shrinking conditions* if for any $n \in \mathbb{N}$ big enough, the following properties are satisfied for all $u = n - k \leq gn$.

(i) $1 - p_k = 1 - p_{n-u} \leq e^{-\rho_n} + a\frac{u}{n}$ resp. $1 - p_k = 1 - p_{n-u} \geq e^{-\rho_n} - a\frac{u}{n}$.

(ii) $c_k = c_{n-u} \leq \frac{c}{u}$.

For the upper exponential shrinking conditions, we also assume that $e^{-\rho_n} + ag < 1$.

Theorem II.15. *If a homogeneous rumor spreading process satisfies the upper (lower) exponential shrinking conditions, then*

$$\mathbb{E}[T(n - \lfloor gn \rfloor, n)] \leq \tfrac{1}{\rho_n} \ln n + O(1) \text{ resp. } \mathbb{E}[T(n - \lfloor gn \rfloor, n)] \geq \tfrac{1}{\rho_n} \ln n - O(1).$$

There exist $A' > 0, \alpha' > 0$ *such that for any* $r \in N$*,*

$$\mathbb{P}[T(n - \lfloor gn \rfloor, n) > \tfrac{1}{\rho_n} \ln n + r] \leq A' e^{-\alpha r} \text{ resp. } \mathbb{P}[T(n - \lfloor gn \rfloor, n) \leq \tfrac{1}{\rho_n} \ln n - r] \leq A' e^{-\alpha r}.$$

Double Exponential Shrinking

Let $g \in]0, 1]$, $\ell > 1$, and $a, c \in \mathbb{R}_{\geq 0}$ such that $ag^{\ell-1} < 1$. We say that a homogeneous rumor spreading process satisfies the *upper (lower) double exponential shrinking conditions* if for any n big enough the following properties are satisfied for all $u = n - k \in [1, gn]$.

(i) $1 - p_{n-u} \leq a\left(\frac{u}{n}\right)^{\ell-1}$ resp. $1 - p_{n-u} \geq a\left(\frac{u}{n}\right)^{\ell-1}$.

(ii) $c_{n-u} \leq c\frac{n}{u^2}$.

Theorem II.16. *If a homogeneous rumor spreading process satisfies the upper (lower) double exponential shrinking conditions, then*

$$\mathbb{E}[T(\lceil(1-g)n\rceil, n)] \leq \log_\ell \ln n + O(1) \quad \textit{resp.} \quad \mathbb{E}[T(\lceil(1-g)n\rceil, n)] \geq \log_\ell \ln n - O(1).$$

If the process satisfies the upper double exponential shrinking conditions, then there exist $A' > 0, \alpha' > 0$ such that for any $r \in \mathbb{N}$ we have

$$\mathbb{P}[T(\lceil(1-g)n\rceil, n) \geq \log_\ell \ln n + r] \leq O(n^{-\alpha' r + A'}).$$

If the process satisfies the lower double exponential shrinking conditions, then for any $\alpha \in]0,1[$ there exists $r = O(1)$ such that

$$\mathbb{P}[T(n - \lceil gn\rceil, n - \lfloor n^{1-\alpha}\rfloor) \leq \log_\ell \ln n - r] \leq O(n^{-1+2\alpha\ell}).$$

Remark. Note that the large deviation statements for the double exponential shrinking regime are different. Beyond this regime the process follows the round targets so accurately that at each round the failure probability that the rumor spreading over-progress current target is negligible. Hence, with high probability the runtime of the process differs by at most $O(1)$ from the number of round targets between gn and $\lfloor n^{1-\alpha}\rfloor$ that equals $\log_\ell \ln n + O(1)$ for any $\alpha \in]0,1[$.

While mostly these three regimes suffice, occasionally it will be necessary or convenient to separately regard a small, constant time segment between the growth and the shrinking regime. This is achieved by the following two lemmas.

Lemma II.17. *Consider a homogeneous rumor spreading process. Let $0 < \ell < m < n$ and $0 < p < 1$. Suppose for any number $\ell \leq k < m$, we have $p_k \geq p$. Then*

$$\mathbb{E}[T(\ell, m)] \leq \frac{n-\ell}{n-m} \cdot \frac{1}{p}.$$

Lemma II.18. *Let $f, p \in]0,1[$ and $c > 0$. Suppose that for any $k < fn$ we have $p_k \leq p$ and $c_k \leq \frac{c}{n}$. Then there exists $f' \in]f,1[$ such that with probability $1 - O\left(\frac{1}{n}\right)$ at the end of some round the number of informed nodes will be between fn and $f'n$.*

2.4 Applying the Above Technical Results

In this section, we sketch how to use the above tools to obtain some of the results described in Chapter IV. Since it does not make a difference, to ease the notation we always assume that nodes call random nodes, that is, including themselves.

The main observation is that computing the p_k is usually very elementary. For the covariance conditions, often we observe a negative or zero covariance, but when this is not true, then things can become technical.

For the *basic push, pull, and push-pull protocols*, we easily observe that all covariances to be regarded are negative or zero. Knowing that one uninformed node x_1 becomes informed in the current round has no influence on the pull call of another uninformed node x_2. When the protocol has push calls and x_1 was informed via a push call, then this event makes it slightly less likely that x_2 becomes informed via a push call, simply because at least one informed node is occupied with calling x_1.

The success probabilities p_k are easy to compute right from the protocol definition. When k nodes are informed, then for the three basic protocols the probability that an uninformed node becomes informed are the following.

$$\begin{aligned} p_k &= 1-(1-1/n)^k && \text{(push protocol);}\\ p_k &= k/n && \text{(pull protocol);}\\ p_k &= 1-(1-1/n)^k \cdot \tfrac{n-k}{n} && \text{(push-pull protocol).} \end{aligned}$$

Using elementary estimates like $1-k/n \le (1-1/n)^k \le 1-k/n+k^2/2n^2$, we see that the push and pull protocols satisfy the exponential growth conditions with $\gamma_n = 1$, whereas the push-pull protocol does the same with $\gamma_n = 2$. The push protocol satisfies the exponential shrinking conditions with $\rho_n = 1$. The pull and push-pull protocols satisfy the double exponential shrinking conditions with $\ell = 2$. All growth conditions are satisfied at least up to $k = n/2$ informed nodes and all shrinking conditions are satisfied at least for $u \le n/2$ uninformed nodes, so we do not need the intermediate lemmas.

Faulty communication: The same arguments suffice to analyze these protocols when messages get lost independently with probability $1-p$. The only difference in terms of the results is that now for the pull and push-pull protocols uninformed nodes remain uninformed with at least constant probability (namely when their pull call fails). For this reason, now all three protocols have an exponential shrinking phase.

The push-pull protocol with the restriction that nodes *answer only a single incoming call* randomly chosen among the incoming calls is an example where the exponential growth and shrinking conditions are harder to prove. To compute the p_k we assume that all n calls have a random unique priority in $[1..n]$ and that the call with lowest priority number is accepted. For fixed priority, the probability of being accepted is easy to compute, and this leads to the success probability of a pull call. For the probability to become informed via a push call, the simple argument that the first incoming call is from an informed node with probability k/n solves the problem. When showing the covariance conditions, we face the problem that

it is indeed not clear if we have negative or zero covariance. The event that some node becomes informed increases the chance that this node received a push call. This push call cannot interfere with another node's pull call to an informed node. So it does have some positive influence on the probability of another uninformed node to become informed. Fortunately, for our covariance conditions allow some positive correlation. Because of this, very generally speaking, we can ignore certain difficult to handle situations when they occur rarely enough.

2.5 Limitations of the Phase Method

It is worth mentioning the main limitations of our method related to the considering only of the homogeneous rumor spreading processes (see Definition III.1 in Chapter III) that must satisfy the following properties.

(i) Any node can be in one of the two states: either informed or uninformed. Thus, we cannot directly apply our method to the multistate processes such as median counter algorithm [KSSV00] or any other *independent stop protocol* discussed in Chapter VI which is a variation of the basic push-pull protocol when each informed node may decide to stop making the push calls. This decision is made by each node independently unlike the protocol with the transition time discussed in Section 3.3 of Chapter IV. However, this limitation seems to be the weakest one, we discuss in Chapter VI how we can overcome it.

(ii) The studied process should be *symmetric*, i.e., the probability that an uninformed node learns the rumor does not depends on the choice of the node, but only on the number of informed nodes. Typically, this requirement is satisfied when we consider rumor spreading on the complete graph (or the random graph resampled each round). One of the simple example of the asymmetric protocol is pull protocol when different nodes has different outgoing call capacities [PPS15].

(iii) The strongest limitation is that we consider only *memoryless* processes. This means that even for the multistate process discussed in Chapter IV, the result of the current round should depend only on the macroparameters describing the number of nodes in each state. For the two state rumor spreading discussed in the main part of this work, the result of the current round cannot depend on anything except the number k of informed nodes. As an example of non-memoryless protocol, we can propose the quasi-random rumor spreading [DFS08] when any node cannot call twice the same neighbor.

Chapter III

Main Analysis Technique

Contents

1	What Is Rumor Spreading	**1**
2	Motivation	**3**
3	State of the Art	**6**
4	Our Contribution	**9**
5	Plan of the Work	**12**

As outlined earlier, in this work we attempt to develop a general analysis technique that covers a large class of rumor spreading problems in perfectly connected networks (complete graphs). To this aim, we define a general class of rumor spreading processes and then distill three regimes such that most rumor spreading processes regarded in the literature are covered by these regimes. For each regime, we prove rumor spreading times sharp apart from additive constants. We shall treat upper and lower bounds separately, so that in cases where only estimates in one direction are known, we still obtain this type of bounds.

1 Homogeneous Rumor Spreading Processes

We now characterize the class of rumor spreading processes we aim at analyzing.

Definition III.1 (Homogeneous rumor spreading process)**.** We always assume that we have n nodes. Each node can be either *informed* or *uninformed*. We assume that the process starts with exactly one node being informed. Uninformed nodes may become informed, but an informed node never can become uninformed. We consider a discrete time process, so the process can be partitioned into *rounds*.

In each round each uninformed node can become informed. Whenever a round starts with k nodes being informed, then the probability for each uninformed node to become informed is some number p_k, which only depends on the number k of the informed nodes at the beginning of the round.

The above definition is relatively abstract and, in principle, could be simply phrased as a Markov process on the number $k \in [1..n]$ of informed nodes. We still find it natural to use the language of rumor spreading. We will discuss many rumor spreading processes covered by this definition in Sections 1, 2, and 3, so let us for the moment only remark that the definition covers all processes regarded in the literature as long as they are memoryless (the events in the current round depend only on which nodes are informed) and symmetric (only the numbers of informed and uninformed nodes is relevant, but not which nodes these are). We remark that our methods can be applied to suitable processes that are not memoryless, see Section 3.3 for an example that is not memoryless due to the use of a time counter.

The main insight of this work is that we can mostly ignore the particular structure of a rumor spreading process and only work with the *success probabilities* p_k and the *covariance numbers* c_k defines as follows.

Definition III.2 (Covariance numbers)**.** For a given homogeneous rumor spreading process and $k \in [1..n-1]$ let c_k be the smallest number such that whenever a round starts with k informed nodes and for any two uninformed nodes x_1, x_2, the indicator random variables X_1, X_2 for the events that these nodes become informed in this round satisfy

$$\mathrm{Cov}[X_1, X_2] \leq c_k.$$

It turns out that essentially all homogeneous rumor spreading processes have an *exponential growth phase*, which is roughly characterized by the fact that for suitable constants $f \in (0,1]$, $c \in \mathbb{R}$ and $\gamma_n > 0$ we have for all $k \in [1..fn-1]$ both $p_k = \gamma_n \frac{k}{n}(1 \pm O(\frac{k}{n}))$ and $c_k \leq c\frac{k}{n^2}$.

This growth phase is followed by one of the following two shrinking regimes. (i) Exponential shrinking regime: For suitable constants $g > 0$, $c > 0$, and $\rho_n > 0$, we have for all $u \leq gn$ that $1 - p_{n-u} = e^{-\rho_n} \pm \Theta(\frac{u}{n})$ and $c_{n-u} \leq \frac{c}{u}$. In particular, in a round starting with $u \leq gn$ uninformed nodes, we expect the number of uninformed nodes to shrink by a factor of roughly $e^{-\rho_n}$. (ii) Double exponential shrinking regime: For suitable constants $g > 0$ and $\ell > 1$, we have that for all $u \leq gn$ both $1 - p_{n-u} = \Theta((\frac{u}{n})^{\ell-1})$ and $c_{n-u} \leq c\frac{n}{u^2}$. In particular, we expect the fraction of uninformed nodes to be raised to some positive power $\ell - 1$.

In the following subsections, we shall analyze each of these regimes, treating separately upper and lower bound guarantees. The very rough analysis idea is

the same in each subsection, so we present and discuss it in more detail in the following subsection and then are more brief in the remaining ones.

Before doing so, we define the rumor spreading time and show an elementary fact that will be convenient several times in the following.

Definition III.3 (Rumor spreading times)**.** Consider a homogeneous rumor spreading process. For all $t = 0, 1, \dots$ denote by I_t the number of informed nodes at the end of the t-th round ($I_0 := 1$). Let $k \leq m \leq n$. By $T(k, m)$ we denote the time it takes to increase the number of informed nodes from k to m or more, that is,

$$T(k, m) = \min\{t - s | I_s = k \text{ and } I_t \geq m\}.$$

We call $T(1, n)$ the rumor spreading time of the process.

Most homogeneous rumor spreading processes have the property that when a constant fraction of the nodes is informed, then each uninformed node has a constant positive probability of becoming informed in one round. In this situation, the following lemma allows us to argue that an expected constant number of rounds suffices to go from any constant fraction of informed nodes to any constant fraction of uninformed nodes. This will be convenient in some the following proofs of upper bounds for rumor spreading times, namely when the growth or shrinking conditions are not strong enough near to the middle point of $n/2$ informed nodes.

Lemma III.4. *Consider a homogeneous rumor spreading process. Let* $0 < \ell < m < n$ *and* $0 < p < 1$*. Suppose for any number* $\ell \leq k < m$*, we have* $p_k \geq p$*. Then*

$$\mathbb{E}[T(\ell, m)] \leq \frac{n - \ell}{n - m} \cdot \frac{1}{p}.$$

Proof. Let $q := 1 - p$. We regard a dummy process which coincides with the given process until the number of informed nodes is at least m. If there are at least m nodes informed, then the dummy process shall be such that each uniformed node in each round independently becomes informed with probability p. Obviously, $T(\ell, m)$ is the same for both processes, so we consider the dummy process in the following.

In this dummy process, by the memorylessness of our rumor spreading process, an uninformed node remains uninformed for r rounds with probability at most q^r. Hence the expected number U_r of uninformed nodes after r rounds is $\mathbb{E}[U_r] \leq (n - \ell)q^r$ and Markov's inequality gives

$$\mathbb{P}[T(\ell, m) > r] = \mathbb{P}[U_r > (n - m)] < \frac{n - \ell}{n - m} \cdot q^r.$$

Hence

$$\mathbb{E}[T(\ell,m)] = \sum_{r=0}^{\infty} \mathbb{P}[T(\ell,m) > r] < \frac{n-\ell}{n-m} \sum_{r=0}^{\infty} q^r = \frac{n-\ell}{n-m} \cdot \frac{1}{1-q} \ .$$

□

Similarly to the lemma above, the following lemma will be convenient in some of the proofs of lower bounds for rumor spreading times, again when the growth and shrinking conditions do not cover the whole process. In this case, the following lemma allows us to argue that an arbitrarily small, but still constant fraction of uninformed nodes will be reached at some time.

Lemma III.5. *Let $f, p \in]0,1[$ and $c > 0$. Suppose that for any $k < fn$ we have $p_k \leq p$ and $c_k \leq \frac{c}{n}$. Then there exists $f' \in]f,1[$ such that with probability $1 - O\left(\frac{1}{n}\right)$ at the end of some round the number of informed nodes will be between fn and $f'n$.*

Proof. Suppose $k < fn$. Denote by $X(k)$ the number of newly informed nodes in a round starting with k informed nodes. Since $p_k \leq p$, we have $\mathbb{E}[X(k)] \leq pn(1-f) \leq pn$. Then by Lemma A.3 we have $\mathrm{Var}[X(k)] \leq (p+c)n$. Let $f' \in]f + p(1-f), 1[$. Applying Chebyshev's inequality, we compute

$$\begin{aligned}
&\mathbb{P}[k + X(k) \geq f'n] \leq \mathbb{P}[X(k) \geq (f'-f)n] \\
&\leq \mathbb{P}[X(k) \geq \mathbb{E}[X(k)] + n(f'-f-p(1-f))] \\
&\leq \tfrac{\mathrm{Var}[X(k)]}{n^2(f'-f-p(1-f))^2} = \tfrac{p+c}{n(f'-f-p(1-f))^2} = O\left(\tfrac{1}{n}\right).
\end{aligned}$$

Therefore, the probability that the process "jumps over" the interval $[fn, f'n]$ is $O\left(\frac{1}{n}\right)$. □

2 Exponential Growth Regime

2.1 Upper Bound

In this section and the following, we analyze the runtime of a homogeneous rumor spreading process in the regime where the number of informed nodes roughly grows by a constant factor until a linear number fn of nodes is informed. Not surprisingly, this implies that the process takes a logarithmic time to inform a linear number of nodes.

The challenge in the following analysis, which was also faced by previous works ([FG85], [Pit87], [KSSV00], etc.), is that in most rumor spreading processes the dissemination speed reduces when more nodes are informed. So it is not true

that for all $k \in [1, fn]$, a round starting with k informed nodes ends with an expected number of $k + \gamma k$ nodes, where γ is some constant, but rather that we only expect $E_k = \gamma k(1 - \Theta(k/n))$ newly informed nodes. This non-linearity also implies that a round starting with an *expected* number of k nodes does not end with an expected number of $k + E_k$ informed nodes, but less. So we also need to argue that the number of newly informed nodes a round ends with is strongly concentrated around its expectation, and that thus, we can assume that with sufficiently high probability we end up not too far below the expectation (which gives another small loss over the idealized multiplicative increase of the number of informed nodes).

We overcome these difficulties as follows. (i) We formulate an *exponential growth condition* that is satisfied by essentially all homogeneous rumor spreading processes showing an exponential growth regime. The key observation, which allows us to treat many protocols with this single analysis is that it is not necessary that the actions of the nodes show particular independences. It suffices that a relatively mild covariance condition is satisfied. (ii) We then use (throughout the whole regime from the first informed node to a linear number of informed nodes) a simple phase-target argument. (a) We define for each number k of initially informed nodes a *round target* $E_0(k)$ such that a round starting with k informed nodes with (sufficiently high) probability $1 - q_k$ ends with $E_0(k)$ informed nodes. Hence the expected time to go from k to $E_0(k)$ or more informed nodes is $t_k = 1 + \frac{q_k}{1-q_k}$. (b) From this, we define a sequence of target $k_0 = 1, k_1 = E_0(k_0), k_2 = E_0(k_1), \ldots, k_J = \Theta(n)$ and argue that the time to reach k_J informed nodes is just the sum of the expected times t_{k_j}. By defining the round targets in a suitable manner, we ensure that $J = \log_{1+\gamma}(n)+\Theta(1)$ and that the sum of the t_{k_j} is $J+\Theta(1)$. We note that the phase-target argument was also used in [DK14], there however only for the push-protocol and only in the regime from n^s, s a small constant, to $\Theta(n)$ informed nodes. Consequently, due to the large number of active nodes acting independently, the phase failure probabilities where ignorable small.

In principle, all the arguments outlined above are very elementary and use nothing more advanced than expectations and Chebyshev's inequality. Hence the main technical progress of this work is formulating an exponential growth condition (including the covariance condition) that allows these elementary arguments in a way that the deviations from the idealized "multiply-by-γ" world in the end all disappear in the $\Theta(1)$ term of the dissemination time. These technicalities also appear in some of the following calculations, which therefore, while all not difficult, are at times slightly lengthy. Since arguments similar to the ones in this section are used throughout this work, we give all details in this section and will be more brief in the following ones.

We start in this section with proving an upper bound for the runtime given that

we have suitable lower bounds for the probability that an uninformed node becomes informed. In the following section, we prove a lower bound for the runtime given that we have suitable upper bounds on the speed of the progress. These bounds will match apart from additive constants if the growth factor γ is identical.

Exponential Growth Conditions

Throughout this section, we assume that we regard a homogeneous epidemic protocol which satisfies the following *upper exponential growth conditions* including a covariance condition.

Definition III.6 (upper exponential growth conditions)**.** Let γ_n be bounded between two positive constants. Let $a, b, c \geq 0$ and $0 < f < 1$ with $af < 1$. We say that a homogeneous epidemic protocol satisfies the *upper exponential growth conditions* in $[1, fn[$ if for any $n \in \mathbb{N}$ big enough the following properties are satisfied for any $k < fn$.

(i). $p_k \geq \gamma_n \frac{k}{n} \cdot \left(1 - a\frac{k}{n} - \frac{b}{\ln n}\right)$.

(ii). $c_k \leq c\frac{k}{n^2}$.

The main result of this section is that the upper exponential growth conditions imply that the number of informed nodes multiplies by, essentially, $1 + \gamma_n$ in each round, and that the expected number of rounds until fn nodes are informed, is at most $\log_{1+\gamma_n} n + O(1)$.

Theorem III.7 (upper bound for the spreading time)**.** *Consider a homogeneous epidemic protocol satisfying the upper exponential growth conditions in* $[1, fn[$. *Then there exist constant* A', α' *such that*

$$\mathbb{E}[T(1, fn)] \leq \log_{1+\gamma_n} n + O(1),$$
$$\mathbb{P}[T(1, fn) > \log_{1+\gamma_n} n + r] \leq A' e^{-\alpha' r} \text{ for any } r \in N.$$

Below we will first introduce all preliminary lemmas, and then we will prove Theorem III.7 at the end of the section.

Round Targets and Failure Probabilities

Let us introduce the random variable $X(k)$ being equal to the number of newly informed nodes in a round having k informed nodes at the beginning. Since $\mathbb{E}[X(k)] = p_k(n - k)$, the exponential growth conditions imply $\mathbb{E}[X(k)] \geq E(k)$, where

$$E(k) := \gamma_n k \left(1 - (a+1)\frac{k}{n} - \frac{b}{\ln n}\right). \tag{III.1}$$

Using Chebyshev's inequality we can show that the value of $X(k)$ is concentrated around its expected value. Lemma III.9 hence claims that with good probability, $X(k)$ attains at least the *target value*

$$E_0(k) := E(k) - Ak^B, \tag{III.2}$$

where $A > 0$ and $B \in]0.5, 1[$ are some constants chosen uniformly for all values of k and n. There are no special conditions on B, so we suppose that B is fixed from now on, e.g., to 3/4. We will, in the following, choose A small enough to ensure that the $-Ak^B$ term has a sufficiently small influence on the general bevahior of $E_0(k)$.

Lemma III.8. *There exist $f' > 0$ and $A' > 0$ such that for n big enough, the following conditions are satisfied.*

- *$E(\cdot)$ is increasing up to $f'n$, that is, for all $i < j \leq f'n$ we have $E(i) < E(j)$;*
- *When A in equation* (III.2) *satisfies $0 < A < A'$, then also $E_0(\cdot)$ is increasing up to $f'n$;*
- *$E_0(k) > 0$ for all $k \in [1, f'n[$.*

Proof. The first claim follows from the second, so let us regard the derivative of $E_0(k)$,

$$E_0'(k) = \gamma_n - 2\gamma_n(a+1)\tfrac{k}{n} - \gamma_n \tfrac{b}{\ln n} - ABk^{-1+B}.$$

We see that, for any $f' < \frac{1}{2(a+1)}$, any $A > 0$ small enough, and any n large enough, $E_0'(k)$ is positive for all $k \in [1, f'n[$. Therefore, to satisfy the first two parts of the claim, we pick any $f' \in]0, \frac{1}{2(a+1)}[$ and then any $A' < \frac{1}{B}\gamma_n(1 - 2(a+1)f')$.

To show that $E_0(k) > 0$ for all $k \in [1, f'n[$, it suffices to check this for $k = 1$. By possibly lowering A' further, we obtain for n large enough that

$$E_0(1) = \gamma_n \left(1 - \tfrac{a+1}{n} - \tfrac{b}{\ln n}\right) - A > 0.$$

□

We assume in the following that f in Definition III.6 satisfies $f < f'$ and that A in (III.2) was chosen in $]0, A'[$.

Lemma III.9. *For any $k < fn$,*

$$\mathbb{P}[X(k) \leq E_0(k)] \leq \min\left\{q(k), \tfrac{1}{1+1/q(1)}\right\},$$

where $q(k) := \frac{\gamma_n + c}{A^2} \cdot k^{-2B+1}$.

Proof. By the exponential growth conditions, $\mathbb{E}[X(k)] \geq E(k)$. Applying Chebyshev's inequality, we compute

$$\begin{aligned}
\mathbb{P}[X(k) \leq E_0(k)] &= \mathbb{P}\left[X(k) \leq E(k) \cdot \left(1 - \tfrac{Ak^B}{E(k)}\right)\right] \\
&\leq \mathbb{P}\left[X(k) \leq \mathbb{E}[X(k)] \cdot \left(1 - \tfrac{Ak^B}{E(k)}\right)\right] \\
&= \mathbb{P}\left[X(k) \leq \mathbb{E}[X(k)] - Ak^B \cdot \tfrac{\mathbb{E}[X(k)]}{E(k)}\right] \\
&\leq \tfrac{\mathrm{Var}[X]}{(Ak^B)^2} \cdot \tfrac{E(k)^2}{\mathbb{E}[X(k)]^2}.
\end{aligned}$$

From the covariance condition, it follows that $\mathrm{Var}[X(k)] \leq \mathbb{E}[X(k)] + ck$. Using $E(k)/\mathbb{E}[X(k)] \leq 1$ once, we obtain

$$\begin{aligned}
\mathbb{P}[X(k) \leq E_0(k)] &\leq \left(1 + \tfrac{ck}{\mathbb{E}[X(k)]}\right) \cdot \tfrac{E(k)}{\mathbb{E}[X(k)]} \cdot \tfrac{E(k)}{A^2k^{2B}} \\
&\leq \left(1 + \tfrac{ck}{E(k)}\right) \cdot \tfrac{E(k)}{A^2k^{2B}} \\
&= \tfrac{E(k)+ck}{A^2k^{2B}} \leq \tfrac{\gamma_n k + ck}{A^2k^{2B}}.
\end{aligned}$$

One can see that for small values of k, $q(k)$ might be more than one. To avoid such a trivial bound for the failure probability, it suffices to replace Chebyshev's inequality in the proof by the Cantelli's inequality (see Lemma A.5) and bound the probability by $\frac{1}{1+1/q(k)}$. To finish the proof we note that $q(k)$ is decreasing in k, so $\mathbb{P}[X(k) \leq E_0(k)] \leq \min\left\{q(k), \frac{1}{1+1/q(1)}\right\}$. □

The Phase Calculus

Having just defined round targets for all numbers k of initially informed nodes and the probabilities that these targets are not achieved within a round, we now proceed to define the sequence k_j of round targets which we aim at satisfying one after the other, ideally within one round per target.

We define recursively

$$k_0 = 1, \quad k_{j+1} := k_j + E_0(k_j). \tag{III.3}$$

Lemma III.10. *After possibly lowering A' from Lemma III.8, there exist $\alpha > 0$ and $J = \log_{1+\gamma_n} n + O(1)$ such that*

$$fn > k_j \geq \alpha(1+\gamma_n)^j,$$

for all $j \leq J$. In particuar, $k_J = \Theta(n)$.

Proof. By definition of k_j,

$$k_j = k_{j-1} + E_0(k_{j-1}) = k_{j-1}\left(1 + \gamma_n - \gamma_n(a+1)\tfrac{k_{j-1}}{n} - \gamma_n\tfrac{b}{\ln n} - Ak_{j-1}^{-1+B}\right).$$

Let $\Gamma_n := 1 + \gamma_n - \gamma_n\frac{b}{\ln n}$. Then,

$$k_j = \Gamma_n k_{j-1}\left(1 - \gamma_n\tfrac{a+1}{\Gamma_n}\cdot\tfrac{k_{j-1}}{n} - \tfrac{A}{\Gamma_n}\cdot k_{j-1}^{-1+B}\right).$$

Clearly, $\Gamma_n \geq (1+\gamma_n)(1-\frac{b}{\ln n})$. By our assumption on γ_n, Γ_n is bounded from above by a constant and is at least $1+\gamma_n/2$ for n big enough. Let hence $\tilde{a} := \gamma_n\frac{a+1}{1+\gamma_n/2}$ and $\tilde{A} := \frac{A}{1+\gamma_n/2}$. Then, for any big n,

$$k_j \geq (1+\gamma_n)\left(1-\tfrac{b}{\ln n}\right)k_{j-1}\left(1 - \tilde{a}\tfrac{k_j-1}{n} - \tilde{A}k_{j-1}^{-1+B}\right).$$

We assume that A (resp. $\tilde{A}$) and f are small enough such that the expression in the brackets is positive. Since $k_0 = 1$, by induction we obtain for all j that

$$k_j \geq (1+\gamma_n)^j(1-\tfrac{b}{\ln n})^j\prod_{i=0}^{j-1}\left(1 - \tilde{a}\tfrac{k_i}{n} - \tilde{A}k_i^{-1+B}\right).$$

By choosing f and A small enough, we can assume that $k_i > 0$ for all $i < j$.

$$k_j \geq (1+\gamma_n)^j(1-\tfrac{b}{\ln n})^j\left(1 - \tilde{a}\sum_{i=0}^{j-1}\tfrac{k_i}{n} - \tilde{A}\sum_{i=0}^{j-1}k_i^{-1+B}\right).$$

Let $J := \log_{1+\gamma_n}(fn) - \Delta r$ for some positive $\Delta r = O(1)$ determined later. For $j \leq J$ we have $k_j \leq (1+\gamma_n)^j$ by construction, and thus $k_j \leq fn$. Also we have $(1-\frac{b}{\ln n})^j = \Theta(1)$. In particular this term is at least 2α for some $\alpha > 0$ and all n big enough.

We show by induction on j that $k_j \geq \alpha(1+\gamma_n)^j$ for all $j \leq J$. The base for $j = 0$ and $k_0 = 1$ is obvious. Let $1 \leq j \leq J$ and let $k_i \geq \alpha(1+\gamma_n)^i$ for all $i < j$. By construction, we have $k_i \leq (1+\gamma_n)^i$. Therefore,

$$\begin{aligned} k_j &\geq 2\alpha(1+\gamma_n)^j\left(1 - \tfrac{\tilde{a}}{n}\sum_{i=0}^{j-1}(1+\gamma_n)^i - \tilde{A}\alpha^{-1+B}\sum_{i=0}^{j-1}(1+\gamma_n)^{i(-1+B)}\right) \\ &\geq 2\alpha(1+\gamma_n)^j\left(1 - \tilde{a}\cdot\tfrac{(1+\gamma_n)^{-\Delta r}}{\gamma_n} - \tilde{A}\alpha^{-1+B}\cdot\tfrac{1}{1-(1+\gamma_n)^{B-1}}\right). \end{aligned}$$

By choosing Δr large enough and $\tilde{A}$ (resp. A) small enough, we can bound the last two expressions by $1/4$, and obtain

$$k_j \geq 2\alpha(1+\gamma_n)^j(1-1/4-1/4) = \alpha(1+\gamma_n)^j.$$

□

By Lemma III.8, the k_j form a non-decreasing sequence. We say that our homogeneous rumor spreading process is in phase j for $j \in \{0, \dots, J-1\}$, if the number of informed nodes is in $[k_j, k_{j+1}[$.

Lemma III.11. *If our process is in phase $j < J$, then the number of rounds to leave phase j is stochastically dominated by $1 + \mathrm{Geom}(1 - Q_j)$, where $Q_j := \min\left\{q(k_j), \frac{1}{1+1/q(1)}\right\}$.*

Proof. By Lemma III.8 we have $k + E_0(k) \geq k_j + E_0(k_j) = k_{j+1}$ for any $k_j \leq k < fn$. By Lemma III.9,

$$\mathbb{P}[k + X(k) \leq k_{j+1}] < \mathbb{P}[k + X(k) \leq k + E_0(k)] < \min\left\{q(k), \tfrac{1}{1+1/q(1)}\right\}.$$

Since $q(k)$ is decreasing,

$$\max_{k_{j+1} > k \geq k_j} \mathbb{P}[k + X(k) < k_{j+1}] \leq Q_j,$$

and this is an upper bound for the probability to stay in phase j for one round. We can thus bound the number of rounds taken to leave phase j by a random variable with geometric distribution $\mathrm{Geom}(1 - Q_j)$. □

Lemma III.12. $\sum_{j=0}^{J-1} Q_j = O(1)$.

Proof. We apply the estimate for $q(k_j)$ from Lemma III.9 and the bounds for k_j from Lemma III.10. Therefore,

$$\begin{aligned}\sum_{j=0}^{J-1} Q_j \leq \sum_{j=0}^{J-1} q(k_j) &\leq \tfrac{\gamma_n + c}{A^2} \cdot \sum_{j=0}^{J-1} k_j^{-2B+1} \\ &\leq \tfrac{\gamma_n + c}{A^2} \cdot \alpha^{-2B+1} \cdot \sum_{j=0}^{J-1} (1+\gamma_n)^{j(-2B+1)}.\end{aligned}$$

The last sum is a decreasing geometric series as $B > 0.5$. So, $\sum_j Q_j = O(1)$. □

Now we can prove the main result of this section.

Proof of Theorem III.7. By Lemma III.10, there exists $J = \log_{1+\gamma_n} n + O(1)$ such that $k_J = \Theta(n)$. In the following we assume that $J \leq \log_{1+\gamma_n} + \tau$ for some constant τ. The phase method allows us to bound the number of rounds until at least k_J nodes are informed. We denote by the random variable T_j the number of

rounds spent in the jth phase. By Lemma III.11, T_j is stochastically dominated by $1 + \text{Geom}(1 - Q_j)$. With Lemma III.12, we compute

$$\begin{aligned}\mathbb{E}[T(1, k_J)] &\leq \sum_{j=0}^{J-1} \mathbb{E}[T_j] \leq \sum_{j=0}^{J-1} (1 + \tfrac{Q_j}{1-Q_j}) \\ &= J + \sum_{j=0}^{J-1} \tfrac{Q_j}{1-Q_j} \leq J + \tfrac{1}{1-Q_0} \sum_{j=0}^{J-1} Q_j \\ &= J + O(1).\end{aligned}$$

Since Q_j is bounded by a geometric sequence, Lemma A.2 claims that there exist A'_1, α'_1 such that

$$\mathbb{P}[T(1, k_J) > J + r/2] \leq A'_1 e^{-\alpha'_1 r}.$$

If $k_J < fn$, then we observe that for all $k \in [k_J, fn[$, p_k satisfies the conditions of Lemma III.4. Therefore, $T(k_J, fn) = O(1)$ and there exist A'_2, α'_2 such that $\mathbb{P}[T(k_J, fn) > r/2] \leq A'_2 e^{-\alpha'_2 r}$. Combining bounds for $T(1, k_J)$ and $T(k_J, fn)$ we obtain the following.

$$\begin{aligned}&\mathbb{E}[T(1, fn)] \leq \mathbb{E}[T(1, k_J)] + \mathbb{E}[T(k_J, fn)] \leq \log_{1+\gamma_n} n + O(1), \\ &\mathbb{P}[T(1, fn) > \log_{1+\gamma_n} n + r] \leq A' e^{-\alpha' r},\end{aligned}$$

where $A' := (A'_1 + A'_2) e^{\alpha' \tau}$ and $\alpha' := \min\{\alpha'_1, \alpha'_2\}$. □

2.2 Lower Bound

In this section, we prove a lower bound for an exponential growth regime. We formulate a condition matching the upper bound condition and show that this leads to a lower bound on the rumor spreading time that matches the upper bound apart from a constant number of rounds. We use again the target-phase method.

This is the first time that the target-phase argument is used to prove a lower bound. In the work closest to ours, [DK14], only the classic push protocol was regarded. Consequently, there, the simple argument that the number of nodes can at most double each round was sufficient to obtain a lower bound for the growth regime. Such an argument, e.g., is not possible for the classic pull protocol.

The main difference to the upper bound proof lies in the final argument. In the upper bound proof, the failure to reach a round target simply resulted in that we had to try again to reach this target. For the lower bound, a failure is that the process gains more than one phase in one round, resulting in that the time usually spent in these now skipped phases is spared. Arguing that the total time spared by such events is only $O(1)$ needs a slightly more complicated book-keeping of the failure events and a slightly more complicated final argument.

Exponential Growth Conditions

We formulate the lower exponential growth condition in an analoguous way as the upper one. In particular, the covariance condition is identical.

Definition III.13 (lower exponential growth conditions)**.** Let γ_n be bounded between two positive constants and let $a, b, c \geq 0$ and $0 < f < 1$. We say that a homogeneous epidemic protocol satisfies the *lower exponential growth conditions* in $[1, fn[$ if for any $n \in \mathbb{N}$ big enough, the following properties are satisfied for any $k < fn$.

(i). $p_k \leq \gamma_n \frac{k}{n} \cdot \left(1 + a\frac{k}{n} + \frac{b}{\ln n}\right)$.

(ii). $c_k \leq c\frac{k}{n^2}$.

These conditions imply the following lower bounds on the rumor spreading time.

Theorem III.14. *Consider a homogeneous epidemic protocol satisfying the lower exponential growth conditions in* $[1, fn[$*. Then there are constant* $A', \alpha' > 0$ *such that*

$$\mathbb{E}[T(1, fn)] \geq \log_{1+\gamma_n} n - O(1),$$
$$\mathbb{P}[T(1, fn) \leq \log_{1+\gamma_n} n - r] \leq A' \exp(-\alpha' r) \text{ for all } r \in \mathbb{N}.$$

In addition there exists $f' \in]f, 1[$ *such that with probability* $1 - O\left(\frac{1}{n}\right)$ *there are at most* $f'n$ *informed nodes after* $T(1, fn)$ *rounds.*

Below we will first introduce all preliminary lemmas, and then we will prove Theorem III.14 at the end of the section.

Round Targets and Failure Probabilities

As above, we consider a round with k informed nodes initially. We define $X(k)$ to be the number of newly informed nodes in this round. Since $\mathbb{E}[X(k)] = p_k(n-k)$, the exponential growth conditions give $\mathbb{E}[X(k)] \leq E(k)$ with

$$E(k) := \gamma_n k \left(1 + a\frac{k}{n} + \frac{b}{\ln n}\right).$$

Note that we could replace the a above by $a-1$, giving an expression closer resembling the corresponding one from the previous section. Since all these constants do not matter, we preferred the simpler version without the extra -1.

Like in the previous section we introduce

$$E_0(k) := E(k) + Ak^B, \tag{III.4}$$

where $A > 0$ and $B \in]0.5, 1[$ are some constants chosen uniformly for all values of k and n. Unlike in Section 2.1, it is obvious that $E(k)$ and $E_0(k)$ are increasing.

Note that we can freely replace f in the definition of the lower exponential growth conditions by a smaller constant f', since showing $\mathbb{E}[T(1, f'n)] \geq \log_{1+\gamma_n}(n) - O(1)$ in Theorem III.14 would immediately imply $\mathbb{E}[T(1, fn)] \geq \log_{1+\gamma_n}(n) - O(1)$. Consequently, let us assume that f is small enough such that for any n sufficiently large and $k < fn$,

$$E(k) \leq 2\gamma_n k. \tag{III.5}$$

The following lemma will later be used to argue that an unexpectedly fast progress is unlikely. Different from the upper bound analysis in the previous section, we now need a failure probability for different excessive progresses (quantified by the parameter h below).

Lemma III.15. *For any $k < fn$ and $h = 0, 1, 2, \ldots$,*

$$\mathbb{P}[X(k) \geq E(k) + Ak^B(1+\gamma_n)^h] \leq q_h(k) := \tfrac{2\gamma_n + c}{A^2} \cdot \tfrac{k^{-2B+1}}{(1+\gamma_n)^{2h}}.$$

Proof. By the exponential growth conditions, $\mathbb{E}[X(k)] \leq E(k)$. By the covariance condition and (III.5),

$$\mathrm{Var}[X(k)] \leq E(k) + n^2 c_k \leq k(2\gamma_n + c).$$

Applying Chebyshev's inequality, we obtain

$$\begin{aligned}
&\mathbb{P}[X(k) \geq E(k) + Ak^B(1+\gamma_n)^h] \\
&\quad \leq \mathbb{P}[X(k) \geq \mathbb{E}[X(k)] + Ak^B(1+\gamma_n)^h] \\
&\quad \leq \tfrac{\mathrm{Var}[X(k)]}{(Ak^B)^2(1+\gamma_n)^{2h}} \\
&\quad \leq \tfrac{2\gamma_n + c}{A^2} \cdot k^{-2B+1} \cdot \tfrac{1}{(1+\gamma_n)^{2h}}.
\end{aligned}$$

□

The Phase Calculus

Like in Section 2.1, we define the sequence k_j recursively by

$$k_0 = 1, \quad k_{j+1} := k_j + E_0(k_j),$$

and obtain the following exponential growth behavior.

Lemma III.16. *By taking A small enough in* (III.4)*, there exist $\alpha > 0$ and $J = \log_{1+\gamma_n} n - O(1)$ such that for all $j < J$*

$$(1+\gamma_n)^j \leq k_j \leq \alpha(1+\gamma_n)^j \text{ and } k_j < fn.$$

Proof. Note that $k_j \geq (1+\gamma_n)^j$ is immediate from the definitions and a simple induction. So it remains to show the upper bound on the k_j. Clearly, by definition of k_j,

$$k_j \leq (1+\gamma_n)(1+\tfrac{b}{\ln n})k_{j-1}\left(1+a\tfrac{k_{j-1}}{n}\right)\left(1+Ak_{j-1}^{-1+B}\right).$$

Since $k_0 = 1$, by induction we obtain

$$k_j \leq (1+\gamma_n)^j(1+\tfrac{b}{\ln n})^j \prod_{i=0}^{j-1}\left(1+a\tfrac{k_i}{n}\right)\prod_{i=0}^{j-1}\left(1+Ak_i^{-1+B}\right).$$

Let $J := \log_{1+\gamma_n} n - \Delta r$ for some $\Delta r = O(1)$ determined later. If $j < J$, then $(1-\frac{b}{\ln n})^j = \Theta(1)$. In particular, it is at most $\frac{\alpha}{4}$ for some $\alpha > 0$ and any n big enough. By the fact that $1+x \leq e^x$ for any $x > 0$, we have

$$k_j \leq \tfrac{\alpha}{4}(1+\gamma_n)^j \exp\left(\sum_{i=0}^{j-1} a\tfrac{k_i}{n}\right)\cdot\exp\left(\sum_{i=0}^{j-1} Ak_i^{-1+B}\right). \tag{III.6}$$

We prove the claim of lemma by induction on j. Assume that for some $j < J$ we have $k_i \leq \alpha(1+\gamma_n)^i$ for any $i < j$. Since $k_i \geq (1+\gamma_n)^i$ for all i, both sums in (III.6) can be bounded by geometric series. Therefore,

$$k_j \leq \tfrac{\alpha}{4}(1+\gamma_n)^j \exp\left(\sum_{i=0}^{j-1}\tfrac{a}{n}\cdot\alpha(1+\gamma_n)^i\right)\cdot\exp\left(\sum_{i=0}^{j-1} A(1+\gamma_n)^{i(-1+B)}\right).$$

Since $j < J$, by choosing Δr large enough and A small enough, we can bound both sums by any positive constant, in particular by $\ln 2$. Therefore, for any $j < J$,

$$k_j \leq \tfrac{\alpha}{4}(1+\gamma_n)^j \exp(\ln 2)\cdot\exp(\ln 2) = \alpha(1+\gamma_n)^j.$$

□

By definition, the k_j form a non-decreasing sequence. Like in Section 2.1, we say that the rumor spreading process is in phase j for $j = 0,\ldots,J-1$, if the number of informed nodes is in $[k_j, k_{j+1}[$.

Lemma III.17. *Let $h \geq 2$. If the process is in phase $j < J$ at the beginning of one round, then the probability that the number of informed nodes is at least k_{j+h} at the end of the round, is at most $q_{h-2}(k_j)$.*

Proof. For $1 \leq k \leq k_{j+1}$, we have

$$k + E(k) + Ak^B \leq k_{j+1} + E(k_{j+1}) + Ak_{j+1}^B = k_{j+2}.$$

Since $k_{j+h} \geq (1+\gamma_n)^{h-2} k_{j+2}$, we have

$$k_{j+h} \geq (1+\gamma_n)^{h-2} \left(E(k) + Ak^B + k\right) \geq k + E(k) + Ak^B (1+\gamma_n)^{h-2}.$$

By Lemma III.15, the maximum probability to have at least k_{j+h} informed nodes at the end of the round is

$$\begin{aligned}
&\max_{k \in [k_j, k_{j+1}[} \mathbb{P}[k + X(k) \geq k_{j+h}] \\
&\leq \max_{k \in [k_j, k_{j+1}[} \mathbb{P}[k + X(k) \geq k + E(k) + Ak^B (1+\gamma_n)^{h-2}] \\
&\leq \max_{k \in [k_j, k_{j+1}[} q_{h-2}(k) \leq q_{h-2}(k_j).
\end{aligned}$$

The last inequality follows from the fact that since $B > 1/2$, $q_{h-2}(\cdot)$ decreases. □

With Lemma III.16 and III.17, we can now prove Theorem III.14.

Proof of Theorem III.14. Let S be the set of visited phases, e.g., if the process does not jump over any phase, then $S = \{0, \ldots, J-1\}$. By τ_j we denote the number of rounds spent in the jth phase. So the spreading time $T(k_0, k_J) = \sum_{j \in S} \tau_j$. We do not know the size of S, so in order to bound the spreading time below, let us introduce the random variable Δ_j which is equal to the length of the jump from the jth phase when the process leaves it. Let also $d_j := \Delta_j - \tau_j$. Since $\sum_{j \in S} \Delta_j = J$, we have $T(k_0, k_J) = J - \sum_{j \in S} d_j$. By definition, for $j \in S$ and $h > 0$, we have $\mathbb{P}[d_j \geq h] \leq \mathbb{P}[\Delta_j \geq h+1]$. Then, by Lemma III.15 and III.17,

$$\mathbb{P}[d_j \geq h] \leq q_{h-1}(k_j) \leq \tfrac{2\gamma_n + c}{A^2} \tfrac{k_j^{-2B+1}}{(1+\gamma_n)^{2h-2}}.$$

The above argument shows that $T(k_0, k_J)$ stochastically dominates $J - D$, where D is the sum of independent non-negative integer random variables $D = \sum_{j=0}^{J-1} D_j$ satisfying $\mathbb{P}[D_j \geq h] \leq \tfrac{2\gamma_n + c}{A^2} \tfrac{k_j^{-2B+1}}{(1+\gamma_n)^{2h-2}}$ for all $h \geq 1$. Let $R_h := \{(r_0, \ldots, r_{J-1}) \in$

$\mathbb{Z}^J_{\geq 0} \mid \sum_{j=0}^{J-1} r_i = h\}$ for all $h \geq 1$. We compute

$$\begin{aligned}
\mathbb{P}[D \geq h] &\leq \sum_{r \in R_h} \prod_{j=0}^{J-1} \mathbb{P}[D_j \geq r_j] \\
&\leq (1+\gamma_n)^{-2h} \sum_{r \in R_h} \prod_{j \in [0..J-1], r_j > 0} \tfrac{2\gamma_n + c}{A^2} \tfrac{k_j^{-2B+1}}{(1+\gamma_n)^{-2}} \\
&\leq (1+\gamma_n)^{-2h} \sum_{M \subseteq [0..J-1]} \prod_{j \in M} \tfrac{2\gamma_n + c}{A^2} \tfrac{k_j^{-2B+1}}{(1+\gamma_n)^{-2}} \\
&\leq (1+\gamma_n)^{-2h} \prod_{j \in [0..J-1]} \left(1 + \tfrac{2\gamma_n + c}{A^2} \tfrac{k_j^{-2B+1}}{(1+\gamma_n)^{-2}}\right) \\
&\leq (1+\gamma_n)^{-2h} \exp\left(\sum_{j \in [0..J-1]} \tfrac{2\gamma_n + c}{A^2} \tfrac{k_j^{-2B+1}}{(1+\gamma_n)^{-2}}\right) \\
&\leq (1+\gamma_n)^{-2h} O(1),
\end{aligned}$$

where the last estimate uses Lemma III.16. This proves that tail bound statement. For the claim on the expected rumor spreading time, we compute

$$\mathbb{E}[D] \leq \sum_{h \geq 1} \mathbb{P}[D \geq h] \leq \sum_{h \geq 1} (1+\gamma_n)^{-2h} O(1) = O(1).$$

Finally, by Lemma III.5, there exists $f' \in]f, 1[$ such that with probability $1 - O\left(\frac{1}{n}\right)$ there are at most $f'n$ informed nodes after $T(1, fn)$ rounds. □

3 Exponential Shrinking Regime

3.1 Upper Bound

We now regard the regime that at most gn, g a small constant, nodes are not informed, and that in each round each of these nodes has an approximately constant chance of becoming informed. From a very distant point of view, this part of the process vaguely resembles the exponential growth regime with time running backwards, but the details are too different to simply transfer our previous results to this setting (e.g., the failure probabilities increase during the process in the exponential shrinking regime and decrease in the exponential growth).

We start in this section with the upper bound on the runtime. Throughout this section, we assume that our homogeneous epidemic protocol satisfies the following *upper exponential shrinking conditions* including the covariance condition.

Definition III.18 (upper exponential shrinking conditions)**.** Let ρ_n be bounded between two positive constants. Let $0 < g < 1$ and $a, c \in \mathbb{R}_{\geq 0}$ such that $e^{-\rho_n} + ag < 1$. We say that a homogeneous epidemic protocol satisfies *the upper exponential shrinking conditions* if for any $n \in \mathbb{N}$ big enough, the following properties are satisfied, for all $u = n - k \leq gn$.

(i). $1 - p_k = 1 - p_{n-u} \leq e^{-\rho_n} + a\frac{u}{n}$;

(ii). $c_k = c_{n-u} \leq \frac{c}{u}$.

Let us note that in this section we study the number of uninformed nodes $u := n - k$ instead of k, i.e., the number of informed ones. We will show that u shrinks by almost a constant factor each round. So the main result of the section is the following theorem (as above, we first introduce all preliminary lemmas and then we prove the theorem at the end of this section).

Theorem III.19 (upper bound for spreading time)**.** *Consider a homogeneous epidemic protocol satisfying the upper exponential shrinking conditions. Then there are constant $A', \alpha' > 0$ such that*

$$\mathbb{E}[T(n - \lfloor gn \rfloor, n)] \leq \tfrac{1}{\rho_n} \ln n + O(1),$$
$$\mathbb{P}[T(n - \lfloor gn \rfloor, n) > \tfrac{1}{\rho_n} \ln n + r] \leq A' e^{-\alpha r} \text{ for all } r \in \mathbb{N}.$$

We first note that the upper exponential shrinking conditions imply that nodes remain uninformed with at most a constant probability. Hence Lemma III.4 shows that we reach any constant fraction of uninformed nodes in expected constant time. For this reason, we may conveniently assume that *g is an arbitrarily small constant* in the following. We shall also always assume that *n is large enough.*

The proof below follows the general principle established in this work, that is, we define for each number u of uninformed nodes a suitable target $E_0(u)$ such that with sufficiently high probability $1-q(u)$ (following from the covariance condition and Chebyshev's inequality), one round started with at most u uninformed nodes ends with at most $E_0(u)$ uninformed nodes. The choice of $E_0(u)$ is such that the sequence $u_0 = gn, u_1 = E_0(u_0), u_2 = E_0(u_1), \dots$ within $J = \frac{1}{\rho_n}\ln(n) + O(1)$ steps reaches a constant u_J and such that failure probabilities $q(u_i)$, $i = 0, \dots, J-1$, imply that only an expected constant number of rounds in addition to J are needed to reach at most u_J nodes. For the constant number of u_J or less remaining uninformed nodes, we use the simple waiting time argument that each of them needs an expected constant number of rounds to be informed, adding another constant number of rounds to the expected spreading time.

Round Targets and Failure Probabilities

Let us introduce the random variable $Y(u)$ being equal to the number of uninformed nodes at the end of a round started with u uninformed ones. Since $\mathbb{E}[Y(u)] = u(1-p_{n-u})$, the exponential shrinking conditions imply that

$$\mathbb{E}[Y(u)] \le E(u) := u\left(e^{-\rho_n} + a\tfrac{u}{n}\right).$$

As before, the Lemma III.21 shows that with "good" probability, $Y(u)$ is less than the *target value*

$$E_0(u) := E(u) + Au^{1-B}, \tag{III.7}$$

where $A > 0$ and $0 < B < 1/2$ are some constants chosen uniformly for all values of u and n. In addition we will choose g and A small enough (relative to g) to ensure that for all $u \le gn$, the target value $E_0(u)$ is less than u (see Lemma III.20) and that the "chain" of consequent target values forms an exponentially decreasing sequence (see Lemma III.23).

Lemma III.20. *Assume that g and A are sufficiently small constants. Then for all $u \in [1, gn]$, we have $E_0(u) < u$.*

Proof. Indeed, it suffices to show that

$$\tfrac{E_0(u)}{u} = e^{-\rho_n} + a\tfrac{u}{n} + Au^{-B} < 1.$$

Since $u \in [1, gn]$, we have

$$\tfrac{E_0(u)}{u} \le e^{-\rho_n} + ag + A.$$

Clearly there exist positive A and g small enough such that the expression above is less than 1. □

We assume in the following that g and A are small enough to make the assertion of the lemma above true. We compute the target failure probabilities as follows.

Lemma III.21. *For any* $1 \le u < gn$,

$$\mathbb{P}[Y(u) \ge E_0(u)] \le q(u) := \tfrac{(1+a)e^{-\rho_n}+c}{A^2} \cdot \tfrac{1}{u^{1-2B}}.$$

Proof. Like in the proofs of Lemma III.9 and III.15, using Chebyshev's inequality and taking into account $E(u) \ge \mathbb{E}[Y(u)]$, we compute

$$\mathbb{P}[Y(u) \ge E_0(u)] \le \mathbb{P}\left[Y(u) \ge \mathbb{E}[Y(u)] + Au^{1-B}\right] \le \tfrac{\mathrm{Var}[Y(u)]}{(Au^{1-B})^2}.$$

From Lemma A.3 and the covariance condition it follows that

$$\mathrm{Var}[Y(u)] \le \mathbb{E}[Y(u)] + cu \le \mathbb{E}[Y(u)] + cu.$$

Therefore,

$$\mathbb{P}[Y(u) \ge E_0(u)] \le \tfrac{E(u)+cu}{A^2u^{2-2B}} \le \tfrac{(1+a)e^{-\rho_n}+c}{A^2} \cdot \tfrac{1}{u^{1-2B}}.$$

□

The Phase Calculus

Let us define the sequence u_j recursively by

$$u_0 = gn, \quad u_{j+1} := E_0(u_j).$$

The next observation follows from the definition.

Observation III.22. *For any* $j \ge 1$ *we have* $u_j \ge u_0 e^{-j\rho_n}$. *In particular, for any* $j \le \frac{1}{\rho_n} \ln n$ *we have* $u_j \ge \frac{u_0}{n}$.

Lemma III.23. *By choosing* A *in* (III.7) *and* g *sufficiently small, we can assume that for all* $j \le \frac{1}{\rho_n} \ln n$, *we have* $u_j \le 2u_0 e^{-j\rho_n}$.

Proof. For $j = 0$, there is nothing to prove. Consider $1 \le j \le \frac{1}{\rho_n} \ln n$ and assume that for all $i < j$ we have $u_i \le 2u_0 e^{-i\rho_n}$. We will show that $u_j \le 2u_0 e^{-j\rho_n}$. By definition,

$$\begin{aligned} u_j &= u_0 e^{-j\rho_n} \prod_{i=0}^{j-1} \left(1 + ae^{\rho_n} \tfrac{u_i}{n} + Ae^{\rho_n} u_i^{-B}\right) \\ &\le u_0 e^{-j\rho_n} \prod_{i=0}^{j-1} \exp\left(ae^{\rho_n} \tfrac{u_i}{n} + Ae^{\rho_n} u_i^{-B}\right) \\ &\le u_0 e^{-j\rho_n} \exp\left(\sum_{i=0}^{j-1} ae^{\rho_n} \tfrac{u_i}{n} + \sum_{i=0}^{j-1} Ae^{\rho_n} u_i^{-B}\right). \end{aligned} \tag{III.8}$$

We estimate separately the two sums. Since $u_i \leq 2u_0 e^{-i\rho_n}$ for $i < j$, the first sum can be bounded by a geometric series:

$$\sum_{i=0}^{j-1} ae^{\rho_n} \tfrac{u_i}{n} \leq \tfrac{ae^{\rho_n}}{n} \sum_{i=0}^{j-1} 2u_0 e^{-i\rho_n} \leq ae^{\rho_n} \cdot \tfrac{2u_0}{n} \cdot \tfrac{1}{1-e^{-\rho_n}}.$$

This expression is proportional to $\frac{u_0}{n} = g$, so by choosing g small enough, we can bound it by $\frac{\ln 2}{2}$. For the second sum we use Observation III.22 and obtain

$$\begin{aligned}\sum_{i=0}^{j-1} Ae^{\rho_n} u_i^{-B} &\leq Ae^{\rho_n} \sum_{i=0}^{j-1} u_0^{-B} e^{i\rho_n B} \leq Ae^{\rho_n} u_0^{-B} \frac{e^{j\rho_n B}}{e^{\rho_n B} - 1} \\ &\leq Ae^{\rho_n} \left(\tfrac{n}{u_0}\right)^B \tfrac{1}{e^{\rho_n B}-1} \leq Ae^{\rho_n} g^{-B} \tfrac{1}{e^{\rho_n B}-1}. \qquad \text{(III.9)}\end{aligned}$$

By taking A small enough, the result is also at most $\frac{\ln 2}{2}$. Substituting the sums in (III.8) by their bounds of $\frac{\ln 2}{2}$, we obtain

$$u_j \leq u_0 e^{-j\rho_n} \exp\left(\tfrac{\ln 2}{2} + \tfrac{\ln 2}{2}\right) = 2u_0 e^{-j\rho_n}.$$

□

We assume in the following that A and g are as in Lemma III.23. Combining the lemma above with the definition of $q(u)$ in Lemma III.21, one can easily see the following.

Corollary III.24. *There exists $J \leq \frac{1}{\rho_n} \ln n$ such that (i) $q(u_J) < \frac{1}{2}$ and (ii) $u_J = O(1)$.*

By Lemma III.20, u_j form a decreasing sequence. We say that the rumor spreading process is in *phase* j, $j \in \{0, \ldots, J-1\}$, if the number of informed nodes is in $[u_{j+1}, u_j[$.

Lemma III.25. *If the process is in phase $j < J$, then the number of rounds to leave phase j is stochastically dominated by $1+\mathrm{Geom}(1-Q_j)$, where $Q_j := q(u_{j+1})$.*

Proof. Consider a round with u uninformed nodes. By definition, the process leaves the phase j if $Y(u) < u_{j+1} = E_0(u_j)$. Since $E_0(u)$ is an increasing function, the upper bound for the probability to stay in phase j in current round is the following.

$$\max_{u \in [u_{j+1}, u_j[} \mathbb{P}[Y(u) \geq E_0(u_j)] \leq \max_{u \in [u_{j+1}, u_j[} \mathbb{P}[Y(u) \geq E_0(u)] \leq q(u_{j+1}).$$

So the number of rounds to leave phase j is stochastically dominated by $1 + \mathrm{Geom}(1 - Q_j)$. □

Lemma III.26. $\sum_{j=0}^{J-1} Q_j = O(1)$.

Proof. By Lemma III.21, we have

$$\sum_{j=0}^{J-1} Q_j \leq O(1) \cdot \sum_{j=1}^{J} \tfrac{1}{u_j^{1-2B}} = O(1),$$

where the last equality follows as in (III.9), using that $J \leq \frac{1}{\rho_n} \ln n$. □

Now we can proof the main result of this section, i.e., Theorem III.19.

Proof of Theorem III.19. First, let $g' > 0$ be smaller than g. Then,

$$\mathbb{E}[T(n - \lfloor gn \rfloor, n)] \leq \mathbb{E}[T(n - \lfloor gn \rfloor, n - \lceil g'n \rceil)] + \mathbb{E}[T(n - \lfloor g'n \rfloor, n)].$$

By Lemma III.4, the exponential shrinking conditions imply that $\mathbb{E}[T(n - \lfloor gn \rfloor, n - \lceil g'n \rceil)]$ is at most a constant. In addition there exist $A'_0, \alpha'_0 > 0$ such that $\mathbb{P}[T(n - \lfloor gn \rfloor, n - \lceil g'n \rceil) > r/3] \leq A'_0 e^{-\alpha'_0 r}$. We can hence assume that g is small enough so that all Lemma III.20 and III.23 are satisfied.

We denote by the random variable T_j the number of rounds spent in phase j. With Corollary III.24 and Lemma III.26, we compute

$$\begin{aligned}
\mathbb{E}[T(n - \lfloor gn \rfloor, n - \lceil u_J \rceil)] &\leq \sum_{j=0}^{J-1} \mathbb{E}[T_j] \leq \sum_{j=0}^{J-1} \left(1 + \tfrac{Q_j}{1-Q_j}\right) \\
&= J + \sum_{j=0}^{J-1} \tfrac{Q_j}{1-Q_j} \leq J + \tfrac{1}{1-Q_J} \cdot \sum_{j=0}^{J-1} Q_j \\
&= J + O(1).
\end{aligned}$$

Since Q_j form a geometrical sequence, it follows from Lemma A.2 that there exist $A', \alpha' > 0$ such that

$$\mathbb{P}[T(n - \lfloor gn \rfloor, \lceil u_J \rceil) > J + r/2] \leq A' e^{-\alpha' r}. \tag{III.10}$$

For the last at most u_J uninformed nodes, we argue as follows. Consider one uninformed node. From the exponential shrinking conditions it follows that the expected number of rounds until this node is informed is at most $O(1)$. So, $\mathbb{E}[T(n - \lfloor u_J \rfloor, n)] \leq u_J \cdot O(1) = O(1)$. Finally,

$$\mathbb{E}[T(n - \lfloor gn \rfloor, n)] \leq \mathbb{E}[T(n - \lfloor gn \rfloor, n - \lceil u_J \rceil)] + \mathbb{E}[T(n - \lfloor u_J \rfloor, n)] \leq \tfrac{1}{\rho_n} \ln n + O(1).$$

To prove the tail bound statement, let $q = 1 - \min_{k \in [n-u_J, n]} p_k$. Now we consider the epidemic protocol with $m = O(1)$ uninformed nodes. Since an uninformed

node stays uninformed for $r/2$ rounds with probability at most $q^{r/2}$, we have $\mathbb{P}[T(n-m,n) > r/2] \leq m \cdot q^{r/2}$. Combining the last inequation with (III.10), we obtain

$$\mathbb{P}[T(n - \lfloor gn \rfloor, n) > J + r] \leq (u_J + A') \exp\left(-r \cdot \min\{\, \alpha', \tfrac{\ln q}{2} \,\}\right) .$$

Since $u_J = O(1)$, the tail bound statement directly follows as in the proof of Theorem III.7. □

3.2 Lower Bound

Exponential Shrinking Conditions

Definition III.27 (lower exponential shrinking conditions)**.** Let ρ_n be bounded between two positive constants. Let $0 < g < 1$ and $a, c \in \mathbb{R}_{\geq 0}$. We say that a homogeneous epidemic protocol satisfies *the lower exponential shrinking conditions* if for any $n \in \mathbb{N}$ big enough, the following properties are satisfied, for all $u \leq gn$ (resp. $k \in [n - \lfloor gn \rfloor, n]$).

(i). $1 - p_k = 1 - p_{n-u} \geq e^{-\rho_n} - a\frac{u}{n}$;

(ii). $c_k = c_{n-u} \leq \frac{c}{u}$.

Theorem III.28 (lower bound of spreading time)**.** *Consider a homogeneous epidemic protocol satisfying the lower exponential shrinking conditions (see definition above). There is a constant $g' \in]0,1[$ and further constants $A', \alpha' > 0$ such that for any positive $g < g'$,*

$$\mathbb{E}[T(n - \lfloor gn \rfloor, n)] \geq \tfrac{1}{\rho_n} \ln n + O(1),$$
$$\mathbb{P}[T(n - \lfloor gn \rfloor, n) \leq \tfrac{1}{\rho_n} \ln n - r] \leq A' \exp(-\alpha' r) \text{ for all } r \in \mathbb{N}.$$

Below we will first introduce all preliminary lemmas, and then we will prove Theorem III.28 at the end of the section.

Round Targets and Failure Probabilities

Let $Y(u)$ be the number of uninformed nodes at the end of the round with u uninformed ones. From the exponential shrinking conditions it follows that

$$\mathbb{E}[Y(u)] \geq E(u) := u\left(e^{-\rho_n} - a\tfrac{u}{n}\right).$$

We define the *target value* in the same way as for the upper bound.

$$E_0(u) := E(u) - Au^{1-B}, \tag{III.11}$$

where $A > 0$ and $B \in]0, 1/2[$ are some constants chosen uniformly for all values of u and n. In addition A is required to be small enough to satisfy Lemma III.31.

Lemma III.29. *For any $u > gn$ and $u \in \mathbb{N}$,*

$$\mathbb{P}[Y(u) \leq E_0(u)] \leq q(u) := \tfrac{e^{-\rho_n}+c}{A^2} \cdot \tfrac{1}{u^{1-2B}}.$$

Proof. As before, using Chebyshev's inequality and taking into account that $E(u) \leq \mathbb{E}[Y(u)]$, we compute

$$\begin{aligned}\mathbb{P}[Y(u) \leq E_0(u)] &= \mathbb{P}\left[Y(u) \leq E(u) \cdot \left(1 - \tfrac{Au^{1-B}}{E(u)}\right)\right] \\ &\leq \mathbb{P}\left[Y(u) \leq \mathbb{E}[Y(u)] - Au^{1-B} \cdot \tfrac{\mathbb{E}[Y(u)]}{E(u)}\right] \\ &\leq \tfrac{\mathrm{Var}[Y(u)]}{(Au^{1-B})^2} \cdot \tfrac{E(u)^2}{\mathbb{E}[Y(u)]^2}.\end{aligned}$$

From covariance condition, it follows that $\mathrm{Var}[Y(u)] \leq \mathbb{E}[Y(u)] + cu$. Therefore,

$$\begin{aligned}\mathbb{P}[Y(u) \leq E_0(u)] &\leq \left(1 + \tfrac{cu}{\mathbb{E}[Y(u)]}\right) \cdot \tfrac{E(u)}{\mathbb{E}[Y(u)]} \cdot \tfrac{E(u)}{(Au^{1-B})^2} \\ &\leq \left(1 + \tfrac{cu}{E(u)}\right) \cdot \tfrac{E(u)}{(Au^{1-B})^2} \\ &= (E(u) + cu) \cdot \tfrac{1}{(Au^{1-B})^2} \leq \tfrac{e^{-\rho_n}+c}{A^2} \cdot \tfrac{1}{u^{1-2B}}.\end{aligned}$$

□

The Phase Calculus

We define the sequence u_j recursively by

$$u_0 := gn, \qquad u_{j+1} := E_0(u_j).$$

The next observation follows from the definition.

Observation III.30. *For any $j \geq 0$ we have $u_j \leq u_0 e^{-j\rho_n}$.*

Lemma III.31. *By choosing A in* (III.11) *and g sufficiently small, we can assume that for all $j \leq \frac{1}{\rho_n} n$, we have $u_j \leq \frac{1}{2} u_0 e^{-j\rho_n}$.*

Proof. For $j = 0$, there is nothing to prove. Consider $1 \leq j \leq \frac{1}{\rho_n} \ln n$ and assume that for all $i < j$ we have $u_i \geq \frac{1}{2} u_0 e^{-i\rho_n}$. We will show that $u_j \geq \frac{1}{2} u_0 e^{-j\rho_n}$. By definition,

$$\begin{aligned}u_j &= u_0 e^{-j\rho_n} \prod_{i=0}^{j-1} \left(1 - e^{\rho_n} a \tfrac{u_i}{n} - A \tfrac{1}{u_i^B}\right) \\ &\geq u_0 e^{-j\rho_n} \left(1 - \tfrac{e^{\rho_n} a}{n} \sum_{i=0}^{j-1} u_i - A \sum_{i=0}^{j-1} \tfrac{1}{u_i^B}\right)\end{aligned}$$

Like in the proof of Lemma III.23, we estimate separately the two sums. Using Observation III.30, we obtain for the first sum that

$$\frac{e^{\rho_n} a}{n} \sum_{i=0}^{j-1} u_i \leq e^{\rho_n} a \frac{u_0}{n} \sum_{i \geq 0} e^{-i\rho_n} = \frac{e^{\rho_n} a}{1-e^{-\rho_n}} \cdot \frac{u_0}{n} = g \cdot O(1).$$

By the hypothesis of induction, for any $i < j$, $u_i \geq \frac{1}{2} u_0 e^{-i\rho_n}$. Since $j < \frac{1}{\rho_n} \ln n$,

$$\begin{aligned} A \sum_{i=0}^{j-1} \frac{1}{u_i^B} &\leq \frac{A}{2^B u_0^B} \sum_{i=0}^{j-1} e^{-i\rho_n B} \leq \frac{A}{2^B u_0^B} \cdot \frac{e^{j\rho_n B}}{e^{\rho_n B}-1} \\ &= \frac{A}{2^B (e^{\rho_n B}-1)} \cdot \frac{n^B}{u_0^B} = \frac{A}{2^B (e^{\rho_n B}-1)} \cdot g^{-B} = A g^{-B} \cdot O(1). \end{aligned}$$

Then, by choosing A and g small enough, we can bound both sums by $1/4$, so that

$$u_j \geq u_0 e^{-j\rho_n} \left(1 - \tfrac{1}{4} - \tfrac{1}{4}\right) \geq \tfrac{1}{2} u_0 e^{j-\rho_n}.$$

□

Having u_j bounded from above and below, one can easily see the following.

Corollary III.32. *There exists $J = \frac{1}{\rho_n} \ln n + O(1)$ such that $u_J > 1$ for any n big enough.*

By definition, the u_j form a non-decreasing sequence. We say that the rumor spreading process is in phase j, $j \in \{0, \ldots, J-1\}$, if the number of informed nodes is in $[u_{j+1}, u_j[$.

Lemma III.33. *If the process is in phase $j < J-1$, then the probability that it "leapfrogs" phase $j+1$ (i.e., proceeds to phase $j+2$ or further in current round) is at most $q(u_j)$.*

Proof. Consider a round with $u \in [u_{j+1}, u_j[$ uninformed nodes. The protocol jumps over the phase $j+1$, if at the end of current round $Y(u) < u_{j+2} = E_0(u_{j+1})$. Since E_0 is increasing,

$$\mathbb{P}[u < u_{j+2}] \leq Pr[u < E_0(u)] \leq q(u).$$

Since $q(u)$ is a decreasing function, the upper bound for the probability to jump over phase $j+1$ is the following.

$$\max_{u \in [u_{j+1}, u_j[} \mathbb{P}[u < u_{j+2}] \leq q(u_{j+1}).$$

□

Now we can proof the main result of this section, i.e., Theorem III.28.

Proof of Theorem III.28. Let τ be the first round t (of this shrinking phase) in which the process leapfrogs a phase. Let $\tau = \infty$ if such an event does not occur. By Corollary III.32, the interval $[1, gn]$ is cut into at least $J = \frac{1}{\rho_n} \ln n + O(1)$ phases. Clearly, if $\tau < J$, then $T(n - \lfloor gn \rfloor, n) \geq \tau$, and if $\tau \geq J$, then $T(n - \lfloor gn \rfloor, n) \geq J$.

If $\tau = J - t$, then the process in phase $J - t$, that is, from some number u of uninformed nodes belonging to phase $J - t$, makes an exceptionally large progress from. Since $q(u)$ is a decreasing function, we have $\mathbb{P}[\tau = J - t] \leq q(u_{J-t})$. Consequently, using the fact that $q(u_j)$ forms a decreasing geometric sequence, we obtain

$$\begin{aligned}\mathbb{P}[T(n - \lfloor gn \rfloor, n) \leq J - t] &\leq \mathbb{P}[\tau \leq J - t] \\ &\leq q(u_0) + q(u_1) + \ldots + q(u_{J-t}) = O(q(u_{J-t})).\end{aligned}$$

Then, using $u_{J-t} \geq O(1) \cdot u_J \cdot e^{\rho_n t}$, we compute

$$\begin{aligned}\mathbb{P}[T(n - \lfloor gn \rfloor, n) \leq J - t] &\leq O(q(u_{J-t})) \leq O(1) u_{J-t}^{-2B+1} \\ &\leq O(1)(u_J e^{\rho_n t})^{-2B+1} \leq O(1) \exp(-\Omega(t)).\end{aligned}$$

Applying Lemma III.29, we obtain

$$\begin{aligned}\mathbb{E}[T(n - \lfloor gn \rfloor, n)] &\geq J\mathbb{P}[\tau > J] + \sum_{t=1}^{J-1} t \cdot \mathbb{P}[\tau = t] = J - \sum_{t=1}^{J-1} t\mathbb{P}[\tau = J - t] \\ &\geq J - \sum_{t=1}^{J-1} t q(u_{J-t}) \geq J - \tfrac{e^{\rho_n} + a + c}{A^2} \cdot \sum_{t=1}^{J-1} \tfrac{t}{u_{J-t}^{1-2B}}.\end{aligned}$$

Since $B < 1/2$ and $u_{J-t} \geq O(1) \cdot u_J \cdot e^{\rho_n t}$, the sum above converges. Therefore,

$$\mathbb{E}[T(n - \lfloor gn \rfloor, n)] \geq J + O(1).$$

□

4 Double Exponential Shrinking Regime

4.1 Upper Bound.

In the following two sections we consider the regime in which uninformed nodes remain uninformed with probability proportional to the fraction uninformed nodes, or, more generally, some positive power $\ell - 1$ there of. Such a regime typically occurs in protocols using pull operations. We show that the *fraction* of uninformed

nodes is raised to the ℓ-th power each round and that such a regime informs the last gn nodes (g a small constant) in a double logarithmic number of rounds.

We discuss the upper bound on the runtime first. Throughout this section, we assume that our homogeneous epidemic protocol satisfies the following *upper double exponential shrinking conditions* including the covariance condition.

Definition III.34 (upper double exponential shrinking conditions)**.** Let $g, \alpha \in [0,1]$, $\ell > 1$, and $a, c \in \mathbb{R}_{\geq 0}$ such that $ag^{\ell-1} < 1$. We say that a homogeneous epidemic protocol satisfies *the upper double exponential shrinking conditions* if for any n big enough, the following properties are satisfied for all $u \in [n^{1-\alpha}, gn]$.

(i). $1 - p_{n-u} \leq a\left(\frac{u}{n}\right)^{\ell-1}$.

(ii). $c_{n-u} \leq c\frac{n}{u^2}$.

Similarly to the exponential shrinking regime we argue with the number u of uninformed nodes rather than the number k of informed ones. To ease the notation in the double exponential shrinking regime we use the *fraction* $\varepsilon := \frac{u}{n}$ of uninformed nodes instead of the absolute number u. Thus, the double exponential shrinking conditions turns into the following bounds, valid for all $\varepsilon \in [n^{-\alpha}, g]$ with $\varepsilon n \in \mathbb{N}$.

(i). $1 - p_{n(1-\varepsilon)} \leq a\varepsilon^{\ell-1}$.

(ii). $c_{n(1-\varepsilon)} \leq \varepsilon^{-2}\frac{c}{n}$.

In the definition above, we cover the rounds starting with a number of uninformed nodes between $n^{1-\alpha}$ and gn. While, by taking $\alpha = 1$ this would allow to analyze the process until all nodes are informed, it turns out that the crucial part is reduce the number of uninformed nodes from $\Theta(n)$ to $n^{1-\alpha}$ for an arbitrarily small constant α. For $u \in [1, n^{1-\alpha}]$, the double exponential shrinking conditions can be relaxed: the covariance condition is no longer needed and it is sufficient to bound uniformly the probability of a node to stay uninformed by $n^{-\tau}$, for some $\tau < 1$.

The main result of the section is the following theorem.

Theorem III.35. *Consider a homogeneous epidemic protocol satisfying the upper double exponential shrinking conditions in $[n^{-\alpha}, g]$. Suppose further that there exists $\tau > 0$ such that $1 - p_{n-u} \leq n^{-\tau}$ for all $u \leq n^{1-\alpha}$.*

Then there exist constant $A', \alpha' > 0$ such that

$$\mathbb{E}[T(\lceil(1-g)n\rceil, n)] \leq \log_\ell \ln n + O(1),$$
$$\mathbb{P}[T(\lceil(1-g)n\rceil, n) \geq \log_\ell \ln n + r] \leq O(n^{-\alpha' r + A'}) \text{ for all } r \in \mathbb{N}.$$

Below we will first introduce all preliminary lemmas, and then we will prove Theorem III.35 at the end of the section.

Round Targets and Failure Probabilities

Let the random variable $y(\varepsilon)$ denote to the fraction of uninformed nodes at the end of a round started with εn uninformed ones. The double exponential shrinking conditions state that

$$\mathbb{E}[y(\varepsilon)] \leq E(\varepsilon) := a\varepsilon^{\ell}.$$

Lemma III.36. $\mathrm{Var}[y(\varepsilon)] \leq \frac{1+c}{n}$.

Proof. Indeed, $\mathrm{Var}[y(\varepsilon)] = \frac{1}{n^2}\mathrm{Var}[Y(\varepsilon)]$, where $Y(\varepsilon) := ny(\varepsilon)$ is the number of uninformed nodes at the end of the round. By Lemma A.3,

$$\mathrm{Var}[Y(\varepsilon)] \leq \mathbb{E}[Y(\varepsilon)] + (n\varepsilon)^2 c_{n(1-\varepsilon)} \leq n + cn.$$

□

The next lemma states that with good probability, $y(\varepsilon)$ is less than the *target value* $2E(\varepsilon)$.

Lemma III.37. *For any fraction of uninformed nodes $\varepsilon \in [n^{-\alpha}, g]$,*

$$\mathbb{P}[y(\varepsilon) \geq 2E(\varepsilon)] \leq q := \tfrac{(1+c)}{a^2} n^{2\alpha\ell-1}.$$

Proof. Applying Chebyshev's inequality and taking into account that $E(\varepsilon) \geq \mathbb{E}[y(\varepsilon)]$, we compute

$$\mathbb{P}[y(\varepsilon) \geq 2E(\varepsilon)] \leq \mathbb{P}[y(\varepsilon) \geq \mathbb{E}[y(\varepsilon)] + E(\varepsilon)] \leq \tfrac{\mathrm{Var}[y(\varepsilon)]}{E(\varepsilon)^2}.$$

By Lemma III.36 and since $\varepsilon \geq n^{-\alpha}$,

$$\mathbb{P}[y(\varepsilon) \geq 2E(\varepsilon)] \leq \tfrac{1+c}{n} \cdot \tfrac{1}{(a\varepsilon^{\ell})^2} \leq \tfrac{1+c}{a^2} n^{2\alpha\ell-1}.$$

□

Our choice to analyze the double exponential shrinking regime only up to $n^{1-\alpha}$ uninformed nodes allows us to define q independent of ε. Since the double exponential shrinking conditions imply the second assumption of Theorem III.35, without loss of generality we may assume that $\alpha < \frac{1}{2\ell}$, and that consequently $q = n^{-\Theta(1)}$.

The Phase Calculus

Let us define the sequence ε_j recursively by

$$\varepsilon_0 := g, \quad \varepsilon_{j+1} := 2E(\varepsilon_j).$$

The following observation can be obtained by a simple induction.

Observation III.38. *For all $j \geq 0$, $\varepsilon_j = (2a)^{\frac{\ell^j-1}{\ell-1}} g^{\ell^j}$. In particular, the ε_j form a decreasing sequence if $g < (2a)^{-\frac{1}{\ell-1}}$.*

In the following we assume that g is small enough to ensure that the ε_j decrease. Applying logarithm twice to the previous equation one can also see the following.

Corollary III.39. *There exists $J = \log_\ell \ln n + O(1)$ such that for any n big enough*

$$n^{-\alpha} < \varepsilon_J \leq \left(\tfrac{n^{-\alpha}}{2a}\right)^{1/\ell}.$$

Proof. From Observation III.38 we see that the biggest J such that $\varepsilon_J > n^{-\alpha}$ is equal to $\log_\ell \ln n + O(1)$. Since $\varepsilon_{J+1} < n^{-\alpha}$, we have $\varepsilon_J < \left(\tfrac{n^{-\alpha}}{2a}\right)^{1/\ell}$. □

We say that the process is in phase j if the fraction ε of uninformed nodes is in $]\varepsilon_{j+1}, \varepsilon_j]$.

Lemma III.40. *If the process is in phase j, $j < J$, then the number of rounds to leave phase j is stochastically dominated by $1 + \mathrm{Geom}(1-q)$.*

Proof. Consider a round starting with εn uninformed nodes. By construction, the process leaves the phase j if $y(\varepsilon) \leq \varepsilon_{j+1} = 2E(\varepsilon_j)$. Since $E(\cdot)$ is an increasing function, an upper bound for the probability to stay in phase j in the current round is

$$\max_{\varepsilon \in]\varepsilon_{j+1}, \varepsilon_j]} \mathbb{P}[y(\varepsilon) > 2E(\varepsilon_j)] \leq \max_{\varepsilon \in]\varepsilon_{j+1}, \varepsilon_j]} \mathbb{P}[y(\varepsilon) \geq 2E(\varepsilon)] \leq q.$$

Hence, the number of rounds the process spends in phase j is stochastically dominated by a random variable with distribution $1 + \mathrm{Geom}(1-q)$. □

Let us now prove the main theorem of the section.

Proof of Theorem III.35. From Lemma III.4 it follows that for any $g' < g$ we have $\mathbb{E}[T(n - \lfloor gn \rfloor, n - \lceil g'n \rceil)] = O(1)$. So without loss of generality we can assume that $g < (2a)^{-\frac{1}{\ell-1}}$ that is required by Observation III.38 and, thus, by Corollary III.39.

Let the random variable T_j denote the number of rounds spent in phase j. With Corollary III.39 as well as Lemma III.37 and III.40, we compute

$$\mathbb{E}[T(n-\lfloor gn\rfloor, n-\lceil \varepsilon_J n\rceil)] \leq \sum_{j=0}^{J-1}\mathbb{E}[T_j]$$
$$\leq J\left(1+\tfrac{q}{1-q}\right) = \log_\ell \ln n + O(1) \qquad \text{(III.12)}$$
$$\mathbb{P}\left[T(n-\lfloor gn\rfloor, n-\lceil \varepsilon_J n\rceil) > J+r\right] \leq Jq^{-r} = n^{-\Theta(r)}. \qquad \text{(III.13)}$$

By Corollary III.39, $\varepsilon_J < \left(\frac{n^{-\alpha}}{2a}\right)^{1/\ell}$. Consequently, there exists $\alpha' \in]0,\alpha[$ such that $\varepsilon_J < n^{-\alpha'}$ for any n large enough. Without loss of generality we can assume that for any $u \leq n^{1-\alpha'}$ we have $1-p_{n-u} \leq n^{-\tau}$ (for $u \in [n^{1-\alpha}, n^{1-\alpha'}]$ it follows from the double exponential shrinking condition). Now suppose $u_0 \leq n^{1-\alpha'}$ and consider $T(n-u_0,n)$. By the argument above, any of the u_0 uninformed nodes stays uninformed for $r \geq 1$ rounds with probability at most $n^{-\tau r}$. Then by the union bound, we have $\mathbb{P}[T(n-u_0,n) > r] \leq P_r := \min\{1, n^{-\tau r+1-\alpha'}\}$, that together with (III.13) proves the tail bound statement.

Finally, $\mathbb{E}[T(n-u_0,n)] \leq 1 + \sum_{r\geq 1} P_r = O(1)$, for any $u_0 \leq n^{1-\alpha}$. Then, together with (III.12) it proves that $\mathbb{E}[T(n-\lfloor gn\rfloor, n)] \leq \log_\ell \ln n + O(1)$. □

4.2 Lower Bound.

We now prove that under lower bound conditions comparable to the upper bound conditions of the previous section, we obtain a lower bound on the runtime equaling our upper bound apart from an additive constant.

Double Exponential Shrinking Conditions

Throughout this section, we assume that the following *lower double exponential shrinking conditions* are satisfied.

Definition III.41 (lower double exponential shrinking conditions)**.** Let $g, \alpha \in]0,1]$ and $\ell > 1$. Let $a, c \in \mathbb{R}_{\geq 0}$. We say that a homogeneous epidemic protocol satisfies *the lower double exponential shrinking conditions* if for any n big enough, the following properties are satisfied for all $u \in [n^{1-\alpha}, gn]$.

(i). $1-p_{n-u} \geq a\left(\frac{u}{n}\right)^{\ell-1}$.

(ii). $c_{n-u} \leq c\frac{n}{u^2}$.

Similarly to the upper double exponential shrinking conditions, we work mostly with the fraction $\varepsilon := \frac{u}{n}$ of uninformed nodes instead of the absolute number u. Thus, the double exponential shrinking conditions turns into the following bounds, valid for all $\varepsilon \in [n^{-\alpha}, g]$ with $\varepsilon n \in \mathbb{N}$.

(i). $1 - p_{n(1-\varepsilon)} \geq a\varepsilon^{\ell-1}$.

(ii). $c_{n(1-\varepsilon)} \leq \varepsilon^{-2}\frac{c}{n}$.

The main result of this section is the following theorem.

Theorem III.42. *Consider a homogeneous epidemic protocol satisfying the lower double exponential shrinking conditions in the interval $[n^{1-\alpha}, gn]$. Let r be a sufficiently large constant (possibly depending on α). Then,*

$$\mathbb{E}[T(n - \lceil gn \rceil, n - \lfloor n^{1-\alpha} \rfloor)] \geq \log_\ell \ln n + O(1),$$
$$\mathbb{P}[T(n - \lceil gn \rceil, n - \lfloor n^{1-\alpha} \rfloor) \leq \log_\ell \ln n - r] \leq O(n^{-1+2\alpha\ell}),$$

Below we will first introduce all preliminary lemmas, and then we will prove Theorem III.42 at the end of the section.

Round Targets and Failure Probabilities

Let again $y(\varepsilon)$ denote the fraction of uninformed nodes at the end of a round started with εn uninformed ones. The double exponential shrinking conditions state that

$$\mathbb{E}[y(\varepsilon)] \geq E(\varepsilon) := a\varepsilon^\ell.$$

The next lemma gives that with good probability, $y(\varepsilon)$ is at least the *target* value $E(\varepsilon)/2$.

Lemma III.43. *For any fraction of uninformed nodes $\varepsilon \in [n^{-\alpha}, g]$,*

$$\mathbb{P}\left[y(\varepsilon) \leq \tfrac{1}{2}E(\varepsilon)\right] \leq \tfrac{4+4c}{a^2\varepsilon^2 n} \leq q := \tfrac{4+4c}{a^2} n^{2\alpha\ell-1}.$$

Proof. Applying Chebyshev's inequality and taking into account that $\mathbb{E}[y(\varepsilon)] \geq E(\varepsilon)$, we compute

$$\mathbb{P}[y(\varepsilon) \leq \tfrac{1}{2}E(\varepsilon)] \leq \mathbb{P}\left[y(\varepsilon) \leq \mathbb{E}[y(\varepsilon)] - \tfrac{1}{2}E(\varepsilon)\right] \leq 4 \cdot \tfrac{\mathrm{Var}[y(\varepsilon)]}{E(\varepsilon)^2}.$$

By the same arguments like in Lemma III.36, $\mathrm{Var}[y(\varepsilon)] \leq \frac{1+c}{n}$. Since $\varepsilon \geq n^{-\alpha}$, we have $E(\varepsilon) \geq an^{-\alpha\ell}$, and the claim of the lemma directly follows. □

Similarly to the upper bound, our choice to analyze the double exponential shrinking regime only up to $n^{1-\alpha}$ uninformed nodes allows us to define q independent of ε. We also assume that $\alpha < \frac{1}{2\ell}$ so that $q = n^{-\Theta(1)}$.

The Phase Calculus

Let us define the sequence ε_j recursively by

$$\varepsilon_0 := g, \quad \varepsilon_{j+1} := \tfrac{1}{2}E(\varepsilon_j).$$

The next observation follows from the definition by a simple induction. The ε_j are decreasing simply because $\varepsilon_{j+1} = \frac{1}{2}E(\varepsilon_j) < \mathbb{E}[y(\varepsilon_j)] \leq \varepsilon_j$. Note that $y(\varepsilon) \leq \varepsilon$ with probability one for any homogeneous protocol.

Observation III.44. *For all $j \geq 1$, $\varepsilon_j = (a/2)^{\frac{\ell^j - 1}{\ell - 1}} g^{\ell^j}$. The ε_j form a decreasing sequence.*

In the rest of the section we assume that $g < (a/2)^{-\frac{1}{\ell-1}}$. Applying logarithm twice to the previous equation one can also see the following.

Observation III.45. *There exists $J = \log_\ell \ln n + O(1)$ such that $\varepsilon_J > n^{-\alpha}$.*

As before, we say that the process is in phase j if the fraction ε of uninformed nodes is in $]\varepsilon_{j+1}, \varepsilon_j]$.

Lemma III.46. *If the process starts in phase j, $j < J$, then the probability that after one round it is in phase $j + 2$ or higher is at most q.*

Proof. Consider a round starting with εn uninformed nodes, where $\varepsilon \in]\varepsilon_{j+1}, \varepsilon_j]$. By construction, the process leapfrogs phase $j+1$ if $y(\varepsilon) \leq \varepsilon_{j+2} = \frac{1}{2}E(\varepsilon_{j+1})$. Since $E(\cdot)$ is an increasing function, an upper bound for the probability to jump over phase $j + 1$ is

$$\max_{\varepsilon \in]\varepsilon_{j+1}, \varepsilon_j]} \mathbb{P}[y(\varepsilon) \leq \tfrac{1}{2}E(\varepsilon_{j+1})] \leq \max_{\varepsilon \in]\varepsilon_{j+1}, \varepsilon_j]} \mathbb{P}[y(\varepsilon) \leq \tfrac{1}{2}E(\varepsilon)] \leq q.$$

□

Proof of Theorem III.42. Consider the rumor spreading process starting with $\varepsilon_0 n = gn$ uninformed nodes. By Lemma III.46, with probability at least $(1-q)^J \geq 1 - Jq$, the process visits each phase $j \in [0..J-1]$, which naturally takes at least $J - 1$ rounds. Consequently, by definition of J in Observation III.45, we have

$$\begin{aligned}\mathbb{E}[T(n - \lceil gn \rceil, n - \lfloor n^{1-\alpha} \rfloor)] &\geq \mathbb{E}[T(n - \lceil n\varepsilon_0 \rceil, n - \lfloor n\varepsilon_J \rfloor)] \\ &\geq (J-1)(1 - Jq) = \log_\ell \ln n + O(1).\end{aligned}$$

The large-deviation statement follows immediately from adding the failure probabilities $\frac{4+4c}{a^2\varepsilon_j^2 n}$, $j = 0, \ldots, J-1$, from Lemma III.43. □

Chapter IV

Applications of our Method

Contents

1	Classic Results		**15**
	1.1	Definition of Rumor Spreading	15
	1.2	Independent Random Phone-Call Model	17
	1.3	Asynchronous Rumor Spreading	19
	1.4	Path, Square Lattice, Star, Necklace	20
	1.5	Complete Graph. Overview of the Proofs	25
2	Precise Statements of Our Results		**30**
	2.1	Tight Bounds via a Target-Failure Calculus	30
	2.2	Uniform Treatment of Many Rumor Spreading Processes	31
	2.3	Precise Statement of the Technical Results	32
	2.4	Applying the Above Technical Results	35
	2.5	Limitations of the Phase Method	37

1 Classic Protocols

In this section, we define the classic push, pull, and push-pull protocols, give some background information on them, and show how the methods developed above easily give very sharp (tight apart from additive constants) rumor spreading times. For this, we easily convince ourselves that all three protocols satisfy the exponential growth conditions. The push protocol satisfies the exponential shrinking

conditions, whereas the pull and push-pull protocols both satisfy the double exponential shrinking conditions. For all these conditions, we can show for the upper and lower bound part of the conditions the same value for the critical parameter γ_n, ρ_n, and ℓ), which is why we then obtain sharp estimates for the rumor spreading times.

We stick to the usual convention that for rumor spreading in complete graphs we allow that nodes call themselves, that is, the random communication partner is chosen uniformly at random from all nodes. By replacing all $(1 - \frac{1}{n})$ terms with $(1 - \frac{1}{n-1})$, the elementary proofs below can easily be transformed to the setting where nodes only call random neighbors in the complete graph.

1.1 Push Protocol

The push protocol appeared in the computer science literature first in the works of Frieze and Grimmett [FG85] (as a technical tool to analyze the all-pairs shortest path problem on complete digraphs with random edge weights) and, under the name *rumor mongering*, Demers et al. [DGH+87] which is the first work that proposed rumor spreading as a robust and scalable method to maintain consistency in replicated databases. In the push protocol, in each round each node knowing the rumor calls a random neighbor and gossips the rumor to it.

The push protocol is the most intensively studied rumor spreading process. It has been proven that with high probability it disseminates a rumor known to a single node to all others in time logarithmic in the number n of nodes when the communication networks is a complete graph (see below), a random graph in the $G(n,p)$ model with $p \geq (1+\varepsilon)\ln(n)/n$, that is, only very slightly above the connectivity threshold, or a hypercube [FPRU90], or a random regular graph [FP10] (and this list is not complete).

For the complete graph, Frieze and Grimmett [FG85] show (among other results) that with high probability, the rumor spreading time is $\log_2 n + \ln n \pm o(\log n)$. This estimate was sharpened by Pittel [Pit87], who proved that for any $h = \omega(1)$, the rumor spreading time with high probability is $\log_2 n + \ln n \pm h(n)$. The first explicit bound for the expected runtime, $\lfloor \log_2 n \rfloor + \ln n - 1.116 \leq E[S_n] \leq \lceil \log_2 n \rceil + \ln n + 2.765 + o(1)$ was shown in [DK14]. All these works are relatively technical (see, e.g., the 9-pages proof of [Pit87]) and heavily exploit particular properties of the push process (e.g., a birthday paradox argument for the first $\log_2(o(\sqrt{n}))$ calls and a reduction to the coupon collector process for the last roughly $\ln n$ rounds in [DK14]).

With the methods developed in this work, we only need to show that the push protocol satisfies the exponential growth and shrinking conditions (with $\gamma_n = 1$ and $\rho_n = 1$), which is very easy. This confirms the bound of [DK14] cited above apart from the additive constants, but with a, as we believe, much simpler proof.

Theorem IV.1. *The expected rumor spreading time of the push protocol on the complete graph with n vertices is $\log_2 n + \ln n \pm O(1)$. There also exist constants A', α' such that $\mathbb{P}[|T - \mathbb{E}[T]| > r] \leq A' e^{-\alpha' r}$.*

Proof. Consider one round of the protocol. Let x_1, x_2 be two different uninformed nodes. Let X_1 and X_2 be the indicator random variables for events that x_1 resp. x_2 become informed. Clearly, if we condition on that x_1 becomes informed, then it is slightly less likely that x_2 becomes informed. Consequently, $\mathrm{Cov}[X_1, X_2] < 0$ and the covariance part of the exponential growth and shrinking conditions is satisfied.

Therefore, it remains to analyze the probability p_k of an uninformed node to become informed.

For the exponential growth regime, suppose that k nodes are informed. An uninformed node remains uninformed when all informed nodes fail to call it. Consequently, it becomes informed with probability $p_k = 1 - \left(1 - \frac{1}{n}\right)^k$. With the estimates

$$\tfrac{k}{n} - \tfrac{k^2}{2n^2} \leq p_k \leq \tfrac{k}{n}$$

we see that the protocol satisfies the exponential growth conditions with parameter $\gamma_n = 1$. More precisely, we can take $\gamma_n = 1$, $f = 1$, $b = 0$ and $c = 0$ is both the upper and lower bound exponential growth condition. Taking $a = 1$ satisfies the upper exponential growth condition, taking $a = 0$ suffices for the lower exponential growth condition.

For the exponential shrinking conditions, suppose that there are u uninformed nodes. Again, the probability for a node to stay uninformed is $1 - p_{n-u} = \left(1 - \frac{1}{n}\right)^{n-u}$. By Corollary B.5, for any $u < n$ we have the following estimate.

$$\tfrac{1}{e} \leq 1 - p_{n-u} \leq \tfrac{1}{e} + \tfrac{2}{e} \cdot \tfrac{u}{n}$$

The push protocol hence satisfies the exponential shrinking conditions (from $gn := \frac{1}{2}n$ uninformed nodes on) with parameter $\rho_n = 1$.

By Theorems III.7, III.14, III.19, and III.28, the expected rumor spreading time of the push protocol is $\log_2 n + \ln n \pm O(1)$. □

1.2 Pull Protocol

The pull protocol is dual to the push protocol in the sense that now in each round, each uninformed node calls a random neighbor and becomes informed if the latter was informed. We are not aware of a convincing practical motivation for this protocol, however, it has been very helpful in proving performance guarantees for other protocols, e.g., in [Gia11]. Note that the duality between the two protocols immediately shows that the probability that the push protocol in t rounds moves a rumor initially present at a node u to a node v equals the probability that the pull

protocol gets the rumor from v to u in t rounds, but this does not imply that both protocols have the same rumor spreading times (as also Theorems IV.1 and IV.2 show).

We are not aware of any performance guarantees proven for the pull protocol. Some existing results for the push protocol obviously can be transformed into results for the pull protocol via the duality and union bounds. For complete graphs, we do not see how this would give bounds stronger than $\Theta(\log n)$.

Interestingly, the expansion phase of the pull protocol (when viewed from a distance) resembles the expansion phase of the push protocol—the probability that an uninformed node becomes informed in a round starting with k informed nodes is $p_k = \frac{k}{n}$ and thus, for small k, very close to the $\frac{k}{n} - \Theta(\frac{k^2}{n^2})$ probability of the push protocol. Nevertheless, the precise processes are very different. For example, in the push protocol we almost surely observe a perfect doubling of the number of informed nodes as long as $o(\sqrt{n})$ nodes are informed. For the pull protocol, the number of newly informed nodes in the first round is binomially distributed with parameters $n-1$ and $\frac{1}{n}$, so the probability for a perfect doubling is asymptotically equal to $\frac{1}{e}$. For this reason, the existing analyses of the push protocol cannot easily be transferred to the pull protocol. This is different for our method, which ignored many details of the process and only relies on the rough characteristics p_k and c_k of the process. We show below that the similar values of p_k lead to the same $\log_2 n \pm O(1)$ time it takes to inform a constant fraction of the nodes. From that point on, the double exponential shrinking conditions are obvious, leading to a double logarithmic remaining time.

Theorem IV.2. *The expected rumor spreading time of the pull protocol on the complete graph with n vertices is $\log_2 n + \log_2 \ln n \pm O(1)$. There also exist constants A', α' such that $\mathbb{P}[|T - \mathbb{E}[T]| > r] \leq A' e^{-\alpha' r}$.*

Proof. Clearly, the events that uniformed nodes become informed are mutually independent. Hence the covariance conditions are exponential growth and double exponential shrinking regimes are satisfied.

An uninformed node becomes informed if its call reaches an informed node. Hence for all $k \in [1..n-1]$, we have $p_k = k/n$. This shows that both the upper and lower exponential growth conditions are satisfied with parameter $\gamma_n = 1$ (and $f = 1$, $a = 0$, $b = 0$, $c = 0$).

For the same reason, the probability $1 - p_{n-u}$ that an uninformed node remains uninformed when u nodes are uninformed, is $1 - p_{n-u} = 1 - \frac{n-u}{n} = \frac{u}{n}$. Consequently, the upper and lower double exponential shrinking conditions are satisfied with $\ell = 2$ (and $g = 1$, $\alpha = 0$, $a = 1$, and $c = 0$).

By Theorems III.7, III.14, III.35, and III.42, the expected rumor spreading time is $\log_2 n + \log_2 \ln n \pm O(1)$. □

1.3 Push-Pull Protocol

In the push-pull protocol, both informed and uninformed nodes contact a random neighbor in each round. If one of the two partners of such a conversation is informed, then also the other one becomes informed. The push-pull protocol is popular for a number of reasons.

The push-pull protocol (called *anti-entropy* there) was found to be very reliable in the first experimental work on epidemic algorithms [DGH+87]. The seminal paper by Karp et al. [KSSV00] proved that the push-pull protocol disseminates a rumor in a complete graph in $\log_3 n \pm O(\log \log n)$ rounds with high probability. This not only is faster than the push and pull protocols, but it allows implementations using only few messages per node. The just mentioned rumor spreading time stems from an exponential growth phase of length roughly $\log_3 n$ and a double exponential shrinking phase. Hence by making informed nodes stop their activity after the exponential growth phase, the total number of messages can be reduced massively.

The push-pull protocol was also investigated in models for social networks. Clearly, when modeling human communication, say people randomly meeting at parties and chatting, a push-pull spreading mechanism makes sense. However, also from the algorithmic viewpoint, it was observed that in graphs with a non-concentrated degree distribution the push-pull protocol greatly outperforms the push and pull protocols. This was first made precise by Chierichetti, Latanzi, and Panconesi [CLP09], who showed that the push-pull protocol spreads a rumor in a preferential attachment graph [BA99, BR03] in time $O(\log^2 n)$, whereas both the push and the pull protocols need time $\Omega(n^\alpha)$ for some constant $\alpha > 0$ to inform all nodes. The precise rumor spreading time of $\Theta(\log n)$ of the push-pull protocol was shown in [DFF11] (see also [DFF12c]). There is was also proven that the rumor spreading time reduces to $\Theta(\frac{\log n}{\log \log n})$ when the communication partners are chosen randomly but with the previous partner excluded. This first sublogarithmic rumor spreading time was quickly followed up by other fast rumor spreading times in networks modeling social networks, e.g., [FPS12, DFF12a, MP14].

The push-pull protocol also performs well and admits strong theoretical analyses when the network has certain general expansion properties like a good vertex expansion [GS12, Gia14] or a low conductance [MS06, CLP10, Gia11].

Theorem IV.3. *The expected rumor spreading time of the push-pull protocol on the complete graph with n vertices is $\log_3 n + \log_2 \ln n \pm O(1)$. There also exist constants A', α' such that $\mathbb{P}[|T - \mathbb{E}[T]| > r] \leq A' e^{-\alpha' r}$.*

Proof. We again discuss the covariance condition first. Consider one round of the protocol. Let x_1, x_2 be two different uninformed nodes. For $i = 1, 2$, let X_i be the indicator random variable for the event that x_i becomes informed in this round, Y_i

the indicator random variable for the event that x_i is called by an informed node, and Z_i the indicator random variable for event that x_i calls an informed node. Clearly, $X_i = \max\{Z_i, Y_i\}$.

We show $\mathrm{Cov}[X_1, X_2] \leq 0$, and thus all covariance conditions, by showing that $\mathbb{P}[X_1 = 1 \mid X_2 = 1] \leq \mathbb{P}[X_1 = 1]$. We have

$$\begin{aligned}\mathbb{P}[X_1 =1 \mid X_2 = 1] = \mathbb{P}[X_1 = 1 \mid X_2 = 1 \wedge Z_2 = 1] \cdot \mathbb{P}[Z_2 = 1 \mid X_2 = 1] \\ + \mathbb{P}[X_1 = 1 \mid X_2 = 1 \wedge Z_2 = 0] \cdot \mathbb{P}[Z_2 = 0 \mid X_2 = 1]. \qquad \text{(IV.1)}\end{aligned}$$

Since the intersection of events $Z_2 = 1 \wedge X_2 = 1$ is equivalent to the single event $Z_2 = 1$ and the outgoing call of the uninformed node cannot inform any node, we have

$$\mathbb{P}[X_1 = 1 \mid X_2 = 1 \wedge Z_2 = 1] = \mathbb{P}[X_1 = 1 \mid Z_2 = 1] = \mathbb{P}[X_1 = 1]. \qquad \text{(IV.2)}$$

When $Z_2 = 0 \wedge X_2 = 1$ holds, then x_2 becomes informed via a push call, which is not available anymore to inform x_1. Hence

$$\mathbb{P}[X_1 = 1 \mid Z_2 = 0 \wedge X_2 = 1] \leq \mathbb{P}[X_1 = 1]. \qquad \text{(IV.3)}$$

From (IV.1) to (IV.3) we obtain $\mathbb{P}[X_1 = 1 \mid X_2 = 1] \leq \mathbb{P}[X_1 = 1]$.

An uninformed node remains uninformed if it is not called by any informed node and it calls an uninformed node itself. Hence $p_k = 1 - \left(1 - \frac{1}{n}\right)^k \cdot \frac{n-k}{n}$. Using the estimates from Lemma B.2 we obtain

$$2\tfrac{k}{n} - \tfrac{3k^2}{2n^2} \leq p_k \leq 2\tfrac{k}{n}$$

and see that the protocol satisfies the exponential growth conditions with $\gamma_n = 2$.

Likewise, the probability $1 - p_{n-u}$ that an uninformed node stays uninformed in a round starting with u uninformed nodes is equal to $\frac{u}{n}\left(1 - \frac{1}{n}\right)^{n-u}$. With Corollary B.5, we estimate

$$\tfrac{1}{e} \cdot \tfrac{u}{n} \leq 1 - p_{n-u} \leq \tfrac{u}{n}.$$

Therefore, the protocol satisfies the double exponential shrinking conditions with $\ell = 2$.

By Theorems III.7, III.14, III.35, and III.42, the expected rumor spreading time is $\log_3 n + \log_2 \ln n \pm O(1)$. □

2 Robustness, Multiple Calls, and Dynamic Graphs

In this section, we apply our analysis method to settings (i) in which calls fail independently with constant probability, (ii) in which nodes are allowed to call a random number of other nodes instead of one as proposed in [PPS15], and (iii) to a simple dynamic graph setting.

2.1 Transmission Failures

One key selling point for randomized rumor spreading, and more generally gossip-based algorithms, is that all these algorithms due to the intensive use of independent randomness are highly robust against all types of failures. In this subsection, we analyze the performance of the three classic protocols in the presence of independent transmission failures, that is, when calls are successful only with probability $p < 1$. Not unexpectedly, we can show that the rumor spreading times only increase by constant factors. However, we also observe a structural change, namely that the extremely fast double exponential shrinking previously seen with the pull and push-pull protocols is replaces by the slower single exponential shrinking regime. This has the important implication that the message complexity of the simple push-pull protocol (where messages are counted as in [KSSV00] and the protocol is assumed to stop when a suitable time limit is reached) increases from the theoretically optimal value of $\Theta(n \log \log n)$ to $\Theta(n \log n)$, see the remark following the proof of Theorem IV.6.

While the robustness of randomized rumor spreading is consistently emphasized in the literature, only relatively few proven guarantees for this phenomenon exist. All results model communication failures by assuming that each call independently with probability $1 - p$ fails to reach its target. The usual assumption is that the protocol does not take notice of such events. Elsässer and Sauerwald [ES09] show for any graph G that if the push protocol spreads a rumor with probability $1 - O(1/n)$ to all nodes in time T, then the push protocol with failures succeeds in informing all nodes with probability $1 - O(1/n)$ in time $\frac{6}{p}T$. This was made more precise for complete graphs in [DHL13], for which a rumor spreading time of $\log_{1+p} n + \frac{1}{p} \ln n \pm o(\log n)$ was shown to hold with high probability. The same result also holds for random graphs in the $G(n, p')$ model when the edge probability p' is $\omega(\log(n)/n)$, that is, asymptotically larger than the connectivity threshold [FHP10]. To the best of our knowledge, these few results are all that is known in terms of proven guarantees for the classic rumor spreading protocols in the presence of failures.

We now use the methods developed in this work to obtain very sharp estimates for the runtimes of the classic protocols on complete graphs when calls fail independently with probability $1 - p$, $p < 1$. As in Sections 1, the growth or shrinking conditions valid in each case are easily proven, showing again the versatility of our approach.

Theorem IV.4. *The expected rumor spreading time for the push protocol with success probability p on the complete graph of size n is equal to*

$$\log_{1+p} n + \tfrac{1}{p} \ln n \pm O(1).$$

There also exist constants A', α' such that $\mathbb{P}[|T - \mathbb{E}[T]| > r] \leq A'e^{-\alpha' r}$.

Proof. With the same argument as in the proof of Theorem IV.1, we see that the covariances regarded in the covariance conditions are all negative.

Consider an uninformed node in a round started with k informed nodes. The probability that it becomes informed in this round is $p_k = 1 - (1 - \frac{p}{n})^k$. By Lemma B.2, we estimate

$$\tfrac{pk}{n} - \tfrac{p^2k^2}{2n^2} \leq p_k \leq \tfrac{pk}{n}$$

for all $k < n$ and see that the protocol satisfies the exponential growth conditions in $[1, n[$ with $\gamma_n = p$.

Similarly, the probability that an uninformed node in a round starting with $u := n - k$ uninformed nodes stays uninformed, is $1 - p_{n-u} = \left(1 - \frac{p}{n}\right)^{n-u}$. By Corollary B.6, we estimate

$$e^{-p} \leq 1 - p_{n-u} \leq e^{-p}(1 + \tfrac{2pu}{n})$$

for all $u < n$ and thus have the exponential shrinking conditions with $\rho_n = p$ for all $u \leq n/2$.

By Theorems III.7, III.14, III.19, and III.28, the expected rumor spreading time is $\log_{1+p} n + \frac{1}{p} \log n \pm O(1)$. □

The result above and its proof are valid for $p = 1$ and then coincide with Theorem IV.1. For the pull protocol and the push-pull protocol, we observe a substantial change of the process when transmission errors occur. In this case, an uninformed node stays uninformed with probability at least $1 - p$, so the double exponential shrinking conditions cannot be satisfied. Instead, we observe that the single exponential shrinking conditions are satisfied.

Theorem IV.5. *The expected rumor spreading time of the pull protocol with success probability $p < 1$ on the complete graph of size n is equal to*

$$\log_{1+p} n + \tfrac{1}{\ln \frac{1}{1-p}} \ln n \pm O(1).$$

There also exist constants A', α' such that $\mathbb{P}[|T - \mathbb{E}[T]| > r] \leq A'e^{-\alpha' r}$.

Proof. As in the proof of Theorem IV.2, the events that uninformed nodes become informed are mutually independent. Hence all covariance conditions are satisfied with $c = 0$. The probability that an uninformed node becomes informed in a round starting with k informed nodes is $p_k = p\frac{k}{n}$, hence the protocol satisfies the exponential growth conditions in $[1, n[$ with $\gamma_n = p$.

Similarly, the probability that an uninformed node remains uninformed in a round starting with u uninformed nodes is

$$1 - p_{n-u} = 1 - p\tfrac{n-u}{n} = 1 - p + p\tfrac{u}{n} = \exp\Big(-\ln \tfrac{1}{1-p}\Big) + p\tfrac{u}{n}.$$

Consequently, the protocol satisfies the exponential shrinking conditions with $\rho_n = \ln \frac{1}{1-p}$ for all $u \leq gn$, g any constant smaller than 1.

By Theorems III.7, III.14, III.19, and III.28, the expected rumor spreading time is $\log_{1+p} n + \frac{1}{\ln(1/(1-p))} \ln n \pm O(1)$. □

Theorem IV.6. *The expected rumor spreading time for the push-pull protocol with success probability $p < 1$ on the complete graph of size n is equal to*

$$\log_{2p+1} n + \frac{1}{p+\ln \frac{1}{1-p}} \ln n \pm O(1).$$

There also exist constants A', α' such that $\mathbb{P}[|T - \mathbb{E}[T]| > r] \leq A' e^{-\alpha' r}$.

Proof. Using the same arguments as for the push-pull protocol without failures, we observe that the covariances are at most zero, so all covariance conditions are satisfied. Consider an uninformed node in a round starting with k informed nodes. The probability that this node does not inform itself via its pull call is $1 - p\frac{k}{n}$. The probability that it is not successfully called by an informed node is $\left(1 - \frac{p}{n}\right)^k$. Hence $p_k = 1 - \left(1 - p\frac{k}{n}\right)\left(1 - \frac{p}{n}\right)^k$ and Corollary B.3 gives

$$2p\tfrac{k}{n} - \tfrac{3p^2k^2}{2n^2} \leq p_k \leq 2p\tfrac{k}{n}.$$

Thus the protocol satisfies the exponential growth conditions in $[1, \frac{2}{3}n[$ with $\gamma_n = 2p$.

Likewise, the probability $1 - p_{n-u}$ that an uninformed node stays uninformed in a round starting with u uninformed nodes is equal to $\left(1 - p\frac{n-u}{n}\right)\left(1 - \frac{p}{n}\right)^{n-u}$. With Corollary B.6 we estimate

$$(1-p)e^{-p} + pe^{-p} \cdot \tfrac{u}{n} \leq 1 - p_{n-u} \leq (1-p)e^{-p} + 3pe^{-p} \cdot \tfrac{u}{n}.$$

Therefore, the protocol satisfies the exponential growth conditions with $\rho_n = p + \ln \frac{1}{1-p}$. Thus by Theorems III.7, III.14, III.19, and III.28, the expected spreading time is equal to $\log_{p+1} n + \frac{1}{p+\ln(1/(1-p))} \ln n \pm O(1)$. □

The fact that in the presence of transmission failures the double exponential shrinking regime ceases to exist has an important implication on the message complexity. In their seminal paper [KSSV00], Karp et al. show that any address-oblivious rumor spreading algorithm that informs all nodes of the complete graph with at least constant probability needs $\Omega(n \log \log n)$ message transmissions in expectation (we refer to that paper for a discussion of the tricky question how to count messages in algorithms performing pull calls).

This optimal order of magnitude is attained by the push-pull protocol when nodes stop sending a rumor that is older than $\log_3 n + O(\log \log n)$ rounds. As

Karp et al. remark, relying on such a time stamp is risky. A mild underestimate of the true rumor spreading time leaves a constant fraction of the nodes uninformed. A mild overestimate of the rumor spreading time by $\varepsilon \log n$ rounds leads to the situation that for $\varepsilon \log n$ rounds a constant fraction of the nodes knows and pushes the rumor, which implies a message complexity of $\Omega(n \log n)$. For this reason, Karp et al. propose the more complicated median-counter algorithms which is robust against a moderate number of adversarial node failures and against moderate deviations from the uniform choice of the nodes to contact.

Our above analysis of the push-pull protocol in the presences of transmission faults shows that not only an unexpected deviation from the ideal fault-free push-pull protocol leads to an increased message complexity, but even a perfectly anticipated faulty behavior. While we know the expected rumor spreading time very precisely (and we could with the same arguments also show a tail bound stating that our upper bound for the expectation is exceeded by λ with probability $\exp(-\Omega(\lambda))$ only), the "transmit until time limit reached" approach still leads to a message complexity of $\Omega(n \log n)$ due to the missing double exponential shrinking phase. As our analysis shows, after an expected number of $\log_{2p+1} n$ iterations, a constant fraction of the nodes are informed. However, it takes another $\frac{1}{p+\ln \frac{1}{1-p}} \ln n + O(1)$ rounds in the exponential shrinking regime until all nodes are informed. Hence when using the simple "transmit until time limit reached" approach to limit the number of messages, the exponential shrinking regime alone would see $\Omega(n \log n)$ push calls by the $\Omega(n)$ informed nodes.

It is not clear how to overcome this difficulty. The median-counter algorithm of Karp et al. for constant-probability transmission failures also seems to require $\Omega(n \log n)$ messages (see the comment right before Theorem 3.1 in [KSSV00]).

2.2 Multiple Calls

In this section, we analyze rumor spreading protocols in which in each round each node when active calls a random number R of nodes. This was proposed by [PPS15] to model different data processing speeds of nodes. Unlike in [PPS15], we assume that each node in each round resamples the number of nodes it may call. This allows to model changing data processing speed as opposed to nodes having generally different speeds.

Consider a random integer variable R taking values in $[0, n[$. We say that a rumor spreading protocol is an R-protocol if in each round it respects the following call procedure. Each node which can make calls in current round samples independently a new value r from R. Then it calls r different neighbors chosen uniformly at random.

In this section we consider the R-push protocol and the R-push-pull protocol

and prove the statements similar to Theorem 1.1, 1.2, and 1.3 from [PPS15]. Note that by putting $R \equiv 1$, we obtain the classic push and push-pull protocols.

Theorem IV.7. *Assume that R is a distribution with $\mathbb{E}[R] = \Theta(1)$ and $\mathrm{Var}[R] = O(1)$. Then the expected spreading time for the R-push protocol on the complete graph of size n is equal to*

$$\log_{1+\mathbb{E}[R]} n + \tfrac{1}{\mathbb{E}[R]} \ln n \pm O(1).$$

There also exist constants A', α' such that $\mathbb{P}[|T - \mathbb{E}[T]| > r] \le A' e^{-\alpha' r}$.

Proof. Consider a round of the protocol started from k informed nodes. Let x_1 and x_2 be two different uninformed nodes and let X_1 and X_2 be the indicator random variables for events that x_1 resp. x_2 become informed. Suppose that node y is informed. The probability that x_1 and x_2 are both called by y is at most

$$\sum_{j\ge 2} \mathbb{P}[R=j] \cdot \binom{j}{2} \cdot \tfrac{1}{n(n-1)} \le \tfrac{1}{n^2} \sum_{j\ge 2} j^2 \cdot \mathbb{P}[R=j] \le (\mathrm{Var}[R] + \mathbb{E}[R]^2) \cdot \tfrac{1}{n^2} = O\left(\tfrac{1}{n^2}\right).$$

Since there are k informed nodes, the probability that x_1, x_2 are both called by the same node (not necessary y) is $k \cdot O\left(\frac{1}{n^2}\right)$. In addition, if we condition on the event that x_1 and x_2 are not called by the same node, then the probability that they both get informed is slightly less than $p_k^2 = \mathbb{P}[X_1 = 1]^2$. Therefore, $\mathrm{Cov}[X_1, X_2] \le k \cdot O\left(\frac{1}{n^2}\right)$ for any $k < n$ which corresponds to the covariance condition for both exponential growth and exponential shrinking.

Now let us study the probability p_k. Since the probability that x does not belong to a random set of j nodes is equal to

$$\left(1 - \tfrac{1}{n}\right)\left(1 - \tfrac{1}{n-1}\right) \dots \left(1 - \tfrac{1}{n-j+1}\right) = \tfrac{n-j}{n},$$

the probability that y does not call x is equal to $\sum_{j\ge 0} \mathbb{P}[R=j] \cdot \frac{n-j}{n} = 1 - \frac{\mathbb{E}[R]}{n}$. Therefore the probability p_k that x gets informed in current round is equal to

$$1 - \left(1 - \tfrac{\mathbb{E}[R]}{n}\right)^k. \tag{IV.4}$$

With Corollary B.3 we estimate

$$\mathbb{E}[R] \cdot \tfrac{k}{n} - \mathbb{E}[R]^2 \cdot \tfrac{k^2}{2n^2} \le p_k \le \mathbb{E}[R] \cdot \tfrac{k}{n}, \tag{IV.5}$$

for any $k \le n/\mathbb{E}[R]$. Therefore, the protocol satisfies the exponential growth conditions in $[1, n/\mathbb{E}[R]]$ with $\gamma_n = \mathbb{E}[R]$.

Similarly, the probability that an uninformed node stays uninformed in a round starting with $u := n - k$ uninformed nodes, is $1 - p_{n-u} = \left(1 - \frac{\mathbb{E}[R]}{n}\right)^{n-u}$. By Corollary B.6, for all $u \leq n/\mathbb{E}[R]$ we estimate

$$e^{-\mathbb{E}[R]} \leq 1 - p_{n-u} \leq e^{-\mathbb{E}[R]}\left(1 + 2\mathbb{E}[R]\tfrac{u}{n}\right). \tag{IV.6}$$

Therefore, the protocol satisfies the exponential shrinking conditions in $[n(1 - 1/\mathbb{E}[R]), n]$ with $\rho_n = \mathbb{E}[R]$.

We note that the intervals for the exponential growth and shrinking regime does not intersect if $\mathbb{E}[R] > 2$. However, we still be able to bound the expected spreading time. From (IV.5) it follows that $p_{n/\mathbb{E}[R]} = 1 - \frac{1}{e} + o(1)$ and $p_{n(1-1/\mathbb{E}[R])} = 1 - e^{1-\mathbb{E}[R]} + o(1)$. Since p_k increases, it is bounded uniformly for any $k \in \left[\frac{n}{\mathbb{E}[R]}, n - \frac{n}{\mathbb{E}[R]}\right]$. Hence, by Lemma III.4, we have $\mathbb{E}\left[T\left(\frac{\mathbb{E}[R]}{n}, n - \frac{\mathbb{E}[R]}{n}\right)\right] = O(1)$. So by Theorems III.7 and III.19, the expected rumor spreading time is at most $\log_{1+\mathbb{E}[R]} n + \frac{1}{\mathbb{E}[R]} \log n \pm O(1)$.

Similarly, by Lemma III.5, there exists some $f' \in \left]1 - \frac{1}{\mathbb{E}[R]}, 1\right[$ such that with probability $1 - O\left(\frac{1}{n}\right)$ the number of informed nodes after some round will belong to $\left[n - \frac{n}{\mathbb{E}[R]}, f'n\right]$. Then by Theorems III.14 and III.28, the expected rumor spreading time is at least $\log_{1+\mathbb{E}[R]} n + \frac{1}{\mathbb{E}[R]} \log n \pm O(1)$. □

Theorem IV.8. *Assume that R is a distribution with $\mathbb{E}[R] = \Theta(1)$ and $\mathrm{Var}[R] = O(1)$. Let ℓ be the smallest nonnegative integer such that $\mathbb{P}[R = \ell] > 0$ and we suppose that $\mathbb{P}[R = \ell] = \Theta(1)$. Then the expected spreading time for the R-push-pull protocol on the complete graph of size n is at most*

$$\begin{aligned} &\log_{1+2\mathbb{E}[R]} n + \tfrac{1}{\mathbb{E}[R] - \ln \mathbb{P}[R=0]} \cdot \ln n \pm O(1), && \ell = 0; \\ &\log_{1+2\mathbb{E}[R]} n + \log_{1+\ell} \ln n \pm O(1), && \ell > 0. \end{aligned}$$

There also exist constants A', α' such that $\mathbb{P}[|T - \mathbb{E}[T]| > r] \leq A' e^{-\alpha' r}$.

Proof. As usual, we discuss the covariance condition first. Consider one round of the protocol started from k informed nodes. Let x_1, x_2 be two different uninformed nodes. For $i = 1, 2$, let X_i be the indicator random variables for event that x_i becomes informed in this round, Y_i the indicator random variable for the event that x_i is called by an informed node, and Z_i the indicator random variable for event that x_i calls an informed node. Since Y_i coincides with X_i for the push protocol from the proof of Theorem IV.7, we have $\mathrm{Cov}[Y_1, Y_2] \leq k \cdot O\left(\frac{1}{n^2}\right)$. In addition Z_i are pairwise independent and also independent from Y_i. Since $X_i = \max\{Z_i, Y_i\}$ we have also $\mathrm{Cov}[X_1, X_2] \leq k \cdot O\left(\frac{1}{n^2}\right)$ for any $k < n$. Therefore, the covariance

condition is satisfied for exponential growth and both exponential and double exponential shrinking conditions.

Let us study $\mathbb{P}[Z_1 = 0]$. If node x_1 calls j different nodes in current round, then the probability that it does not hit informed node is $\left(1 - \frac{k}{n}\right) \dots \left(1 - \frac{k}{n-j+1}\right)$. Summing over all possible values of j we obtain the following.

$$\mathbb{P}[Z_1 = 0] = \sum_{j=0}^{n-k} \mathbb{P}[R = j] \cdot \left(1 - \tfrac{k}{n}\right) \dots \left(1 - \tfrac{k}{n-j+1}\right). \tag{IV.7}$$

Recall that that $\sum_{j=0}^{n} j \cdot \mathbb{P}[R = j] = \mathbb{E}[R]$ and $\sum_{j=0}^{n} j^2 \cdot \mathbb{P}[R = j] = \mathrm{Var}[R] + \mathbb{E}[R]^2 = O(1)$. Using estimate from Corollary B.3, we compute for any $k \le \frac{n}{2}$

$$\begin{aligned}
\mathbb{P}[Z_1 = 0] &\le \sum_{j=0}^{n-k} \mathbb{P}[R = j] \cdot \left(1 - \tfrac{k}{n}\right)^j \\
&\le \sum_{j=0}^{n/k} \mathbb{P}[R = j] \cdot \left(1 - j\tfrac{k}{n} + j^2 \tfrac{k^2}{2n^2}\right) + \sum_{j=n/k+1}^{n-k} \mathbb{P}[R = j] \\
&= \sum_{j=0}^{n/k} \mathbb{P}[R = j] - \tfrac{k}{n} \sum_{j=0}^{n/k} j \cdot \mathbb{P}[R = j] + \tfrac{k^2}{2n^2} \sum_{j=0}^{n/k} j^2 \cdot \mathbb{P}[R = j] + \sum_{j=n/k-1}^{n-k} \mathbb{P}[R = j] \\
&\le 1 - \tfrac{k}{n} \left(\mathbb{E}[R] - \sum_{j=n/k-1}^{n} j \cdot \mathbb{P}[R = j] \right) + \tfrac{k^2}{n^2} \sum_{j=0}^{n-k} j^2 \cdot \mathbb{P}[R = j] \\
&\le 1 - \mathbb{E}[R] \cdot \tfrac{k}{n} + \tfrac{k^2}{n^2} \sum_{j=n/k-1}^{n} j^2 \cdot \mathbb{P}[R = j] + \tfrac{k^2}{n^2} \sum_{j=0}^{n-k} j^2 \cdot \mathbb{P}[R = j] \\
&\le 1 - \mathbb{E}[R] \cdot \tfrac{k}{n} + 2(\mathrm{Var}[R] + \mathbb{E}[R]^2) \cdot \tfrac{k^2}{n^2}.
\end{aligned}$$

For any $k \le \frac{n}{2}$ we can similarly bound $\mathbb{P}[Z_i = 0]$ from below using Bernoulli's inequality.

$$\begin{aligned}
\mathbb{P}[Z_1 = 0] &\ge \sum_{j=0}^{n-k} \mathbb{P}[R = j] \left(1 - k \cdot \tfrac{j}{n-j}\right) \\
&\ge \sum_{j=0}^{n-k} \mathbb{P}[R = j] \left(1 - \tfrac{jk}{n}\left(1 + 2\tfrac{j}{n}\right)\right) \\
&= 1 - \mathbb{E}[R] \cdot \tfrac{k}{n} + O(1) \cdot \tfrac{k^2}{n^2}
\end{aligned}$$

By (IV.5), we estimate $\mathbb{P}[Y_1 = 0] = 1 - \mathbb{E}[R] \cdot \frac{k}{n} \pm O(1) \cdot \frac{k^2}{n^2}$. Since Y_1 and Z_1 are independent, we have

$$\mathbb{P}[X_1 = 1] = 1 - \mathbb{P}[Y_1 = 0] \cdot \mathbb{P}[Z_1 = 0].$$

Therefore, $p_k = 2\mathbb{E}[R] \cdot \frac{k}{n} \pm O(1) \cdot \frac{k^2}{n^2}$ for any $k \leq \min\left\{\frac{n}{2}, \frac{n}{\mathbb{E}[R]}\right\}$. Hence the protocol satisfies the exponential growth conditions with $\gamma_n = 2\mathbb{E}[R]$ for any $k \leq \min\left\{\frac{n}{2}, \frac{n}{\mathbb{E}[R]}\right\}$.

Now we discuss the shrinking conditions. We consider a round started from $u := n - k$ uninformed nodes. Similarly to (IV.7), we have

$$\mathbb{P}[Z_1 = 0] = \sum_{j \geq 0} \mathbb{P}[R = j] \cdot \tfrac{u}{n} \cdot \tfrac{u-1}{n-1} \cdot \ldots \cdot \tfrac{u-j+1}{n-j+1}.$$

Assume first that $\mathbb{P}[R = 0] > 0$, i.e., $\ell = 0$. Since x_1 might not call in current round, there is at least a constant probability, that it stays uninformed. With (IV.6) and estimate

$$\mathbb{P}[R = 0] \leq \mathbb{P}[Z_1 = 0] \leq \mathbb{P}[R = 0] + \mathbb{P}[R \geq 1] \cdot \tfrac{u}{n},$$

we see that $\mathbb{P}[X_1 = 0] = \mathbb{P}[R = 0] \cdot e^{-\mathbb{E}[R]} \pm O(1) \cdot \frac{u}{n}$ for any $u \leq \min\left\{\frac{n}{2}, \frac{n}{\mathbb{E}[R]}\right\}$. In this case the protocol satisfies the exponential shrinking conditions with $\rho_n = \mathbb{E}[R] - \ln \mathbb{P}[R = 0]$. Applying Lemma III.4 and III.5 in the similar way as in the proof of Theorem IV.7, one can see that by Theorems III.7, III.14, III.19, and III.28, the expected rumor spreading time is $\log_{1+2\mathbb{E}[R]} n + \frac{1}{\mathbb{E}[R] - \ln \mathbb{P}[R=0]} \ln n \pm O(1)$.

Finally, suppose that $\mathbb{P}[R = 0] = 0$, and let ℓ be the smallest integer such that $\mathbb{P}[R = \ell] > 0$. In this case we can easily estimate the probability that x_1 stays uninformed. From below we have

$$\mathbb{P}[X_1 = 0] \geq \mathbb{P}[Y_1 = 0] \cdot \mathbb{P}[R = \ell] \cdot \tfrac{u^\ell}{n^\ell} \geq e^{-\mathbb{E}[R]} \cdot \mathbb{P}[R = \ell] \cdot \tfrac{u^\ell}{n^\ell}.$$

From above, $\mathbb{P}[X_1 = 0] \leq \mathbb{P}[Z_1 = 0] \leq \frac{u^\ell}{n^\ell}$. Hence the protocol satisfies the double exponential shrinking conditions with parameter $1 + \ell$. Again, by Theorems III.7, III.14, III.35, and III.42 and Lemmas III.4 and III.5, the expected rumor spreading time is $\log_{1+2\mathbb{E}[R]} n + \log_{1+\ell} \ln n \pm O(1)$. □

2.3 Dynamic Graphs

We now show that our method can also be applied to certain dynamic graph settings, that is, when the network structure may be different in each round. While it is generally agreed upon that dynamic problem settings are highly relevant for practical applications, it is still not so clear what is a good theoretical model

for dynamicity. For rumor spreading problems, the only work regarding dynamic graphs [CCD+16] considers the two models (i) that in each round independently the network is a $G(n,p)$ random graph and (ii) that each possible edge has its own independent two-state Markov chain describing how it changes between being present and not (edge-Markovian dynamic graphs). For both models, it is proven that the push protocol informs all nodes in logarithmic time with high probability (when the parameters are chosen reasonably).

2-regular Dynamic Graphs

To show that our methods also allow the analysis of dynamic graph models with more dependencies, we regard in this section the model where in each round independently the network G_t is chosen as a random 2-regular (simple) graph. Regular random graphs are notorious for the inherent dependencies which already make sampling them highly non-trivial. For this reason, it is quite clear that the classic rumor spreading analysis approach of trying to understand the distribution of the number of newly informed nodes will be tedious. For our method, however, we only need to take the local view of understanding how likely it is that a new node becomes informed. For the covariance condition, while we believe it to be true, we do not need to show that the events of two uninformed nodes becoming informed are negatively correlated (or independent). Since the covariance conditions allow a mild positive covariance, we may conveniently ignore rare events like the two nodes sharing a neighbor and may thus assume that the calls affecting the two nodes are disjoint and thus independent.

Nevertheless, it turns out that the rumor spreading process in this type of dynamic graphs is different from the one in complete graphs. This is visible from the rumor spreading times proven below, but also from the observation that even in the pull protocol the events that two uniformed nodes become informed are not independent.

In this subsection, due to the small node degrees, we skip the assumption that nodes may also call themselves, but assume that contacts are chosen uniformly at random from all neighbors.

Theorem IV.9. *Consider a dynamic graph setting where the network in each round t independently is a 2-regular random (simple) graph G_t. Then the rumor spreading times T of the three classic protocols are as follow.*

- *Push protocol:* $\mathbb{E}[T] = \log_2 n + \log_4 n \pm O(1)$.
- *Pull protocol:* $\mathbb{E}[T] = \log_2 n + \log_2 \ln n \pm O(1)$.
- *Push-pull protocol:* $\mathbb{E}[T] = \log_{5/2} n + \log_2 \ln n \pm O(1)$.

There also exist constants A', α' such that $\mathbb{P}[|T - \mathbb{E}[T]| > r] \leq A'e^{-\alpha' r}$.

Proof. We defer the proof of the covariance conditions to the very end. Consider a round t starting with k informed nodes. Consider a fixed uninformed node x. Let A_i, $i = 0, 1, 2$, be the event that i of its 2 neighbors are informed. Since in a 2-regular random graph the two neighbors of x form a random 2-set of the nodes different from x, we easily compute

$$\begin{aligned}\mathbb{P}[A_0] &= \tfrac{n-1-k}{n-1}\tfrac{n-2-k}{n-2},\\ \mathbb{P}[A_1] &= 2\tfrac{n-1-k}{n-1}\tfrac{k}{n-2},\\ \mathbb{P}[A_2] &= \tfrac{k}{n-1}\tfrac{k-1}{n-2}\,.\end{aligned}$$

For the push protocol, we compute $p_k = \frac{1}{2}\mathbb{P}[A_1] + \frac{3}{4}\mathbb{P}[A_2] = \frac{k}{n-1}(1 - \frac{1}{4}\frac{k-1}{n-2})$. Assuming n to be sufficiently large, the exponential growth condition (apart from the covariance condition) is satisfied in the whole range $k \in [1, n[$ with $\gamma_n = 1$. Rewriting the expression for p_k, we see that the probability to remain uninformed in a round starting with u uninformed nodes, is $1 - p_{n-u} = \frac{1}{4} + \frac{1}{2}\frac{u-1}{n-2} + \frac{1}{4}\frac{(u-1)(u-4)}{(n-1)(n-2)}$. Hence for, say $u \leq n/2$, the exponential shrinking conditions are satisfied with $\rho_n = \ln 4$. Apart from the covariance conditions, this shows our claim for the push protocol.

For the pull protocol, we have $p_k = \frac{1}{2}\mathbb{P}[A_1] + \mathbb{P}[A_2] = \frac{k}{n-1}$ and consequently $1 - p_{n-u} = \frac{u-1}{n-1}$, showing the exponential growth and double exponential shrinking conditions to be satisfied in overlapping ranges with $\gamma_n = 1$ and $\ell = 1$.

For the push-pull protocol, we have $p_k = \frac{3}{4}\mathbb{P}[A_1] + \mathbb{P}[A_2] = \frac{3}{2}\frac{k}{n-1}(1 - \frac{2}{3}\frac{k-1}{n-2})$ and consequently $1 - p_{n-u} = \frac{u-1}{n-1} + \frac{1}{2}\frac{u+1}{n-1} - \frac{1}{2}\frac{(u+1)(u-1)}{(n-1)(n-2)}$, showing the exponential growth and double exponential shrinking conditions to be satisfied in overlapping ranges with $\gamma_n = \frac{3}{2}$ and $\ell = 1$.

It remains to show the covariance conditions. Note first that the covariance condition of the exponential growth conditions implies the other covariance conditions, so it suffices to show the former.

Let x_1, x_2 be two uninformed nodes and let X_1, X_2 be the indicator random variables for the events of becoming informed in the current round. We have $\mathrm{Cov}[X_1, X_2] = \mathbb{P}[X_2](\mathbb{P}[X_1 \mid X_2] - \mathbb{P}[X_1]) = O(\frac{k}{n})(\mathbb{P}[X_1 \mid X_2] - \mathbb{P}[X_1])$, so it suffices to show $\mathbb{P}[X_1 \mid X_2] - \mathbb{P}[X_1] \leq c/n$ for some constant c.

Let B be the event that x_1 and x_2 have distance at least 3 in G_t. Since x_2 has at most 4 other nodes in distance 2 or closer, $\mathbb{P}[B] \geq 1 - \frac{4}{n-1}$. By $\mathbb{P}[X_1 \mid X_2] - \mathbb{P}[X_1] \leq \mathbb{P}[\neg B] + \mathbb{P}[X_1 \mid X_2 \wedge B] - \mathbb{P}[X_1 \mid B]$, we only need to consider the case that B holds. In this case, the targets of the calls of x_1 and its neighbors are independent of the event X_2. Consequently, the only correlation among $X_2 = 1$ and $X_1 = 1$ stems from the fact that $X_2 = 1$ has an influence on where the

informed nodes are in G_t. More formally, denoting by A_i^j the event that x_j has exactly i of its two neighbors informed, we have that $X_2 = 1$ has an influence on the distribution of $(A_i^2)_i$ which in turn has an influence on the distribution $(A_i^1)_i$. However, regardless which event $A_{i_2}^2$ we condition on, the probability distribution of $(A_{i_1}^1)_{i_1}$ is only mildly affected. The precise expression for $\mathbb{P}[A_{i_1}^1 \mid A_{i_2}^2]$ can be obtained from the one for $\mathbb{P}[A_{i_1}]$ above by replacing n by $n-2$ and k by $k-i_2$. Consequently, $\mathbb{P}[A_{i_1}^1 \mid A_{i_2}^2] = \mathbb{P}[A_{i_1}] \pm O(1/n)$ for all $i_1, i_2 \in \{0,1,2\}$. Therefore $\mathbb{P}[X_1 \mid X_2 \wedge B] = \mathbb{P}[X_1 \mid B] + O(1/n)$ and hence $\mathbb{P}[X_1 \mid X_2] - \mathbb{P}[X_1] = O(1/n)$ as aimed at. This shows the covariance condition of the exponential growth conditions in the whole range $k \in [1, n[$, and thus also the other two covariance conditions. □

Erdős-Rényi Dynamic Graphs

It is clear that the edge-Markovian model due to the time-dependence cannot be analyzed with our methods. For the other result, we now show that our method quite easily gives a very precise analysis. We only treat the case of $\Theta(1/n)$ edge probabilities, as this seems to be the most interesting one (the graph is not connected, but has nodes with degrees varying between 0 and $\Theta(\log(n)/\log\log(n))$; when $p \geq (1+\varepsilon)/n$, a giant component encompassing a linear number of nodes exists).

To make the model precise, we assume that in each round independently, before the communication starts, the communication graph is sampled as $G(n,p)$ random graph, where $p = a/n$ for some positive constant a. That is, between any two nodes there is an edge, independently, with probability a/n. In the communication part of the round, each informed node chooses a communication partner uniformly at random from its neighbors in the communication graph and sends a copy of the rumor to it. Isolated informed nodes, naturally, do not communicate in this round.

We introduce the following notation. We consider one round and aim at showing the exponential growth and shrinking conditions. Let E be the set of edges of the communication graph $G(n, \frac{a}{n})$ of this round. We write $xy \in E$ as shorthand for $\{x,y\} \in E$. We write $x \to y$ to denote the event that x calls y. By $\deg_{\text{inf}} x$ we denote the number of informed neighbors of x.

Lemma IV.10. *Consider an uninformed node x and an informed node y. Let $\ell \leq n/2$ and let A_ℓ be the event that $\{y_1 y, \dots, y_\ell y\} \cap E = \emptyset$. Then*

$$\mathbb{P}[y \to x \mid xy \in E \wedge A_\ell] = \tfrac{1-e^{-a}}{a} + (\ell+1) \cdot O\left(\tfrac{1}{n}\right).$$

Proof. Assume that $xy \in E$. Then the number of other neighbors of y, that is, the random variable $\deg y - 1$, has a binomial distribution with parameters $n-2-\ell$ and $\frac{a}{n}$. The probability that y calls x is equal to $\frac{1}{\deg y}$. Using the fact

that $\binom{m+1}{k+1} = \frac{k+1}{m+1}\binom{m}{k}$, we compute

$$\begin{aligned}
\mathbb{P}[y \to x \mid xy \in E \wedge A_\ell] &= \sum_{i=0}^{n-2-\ell} \tfrac{1}{i+1}\binom{n-2-\ell}{i}\left(\tfrac{a}{n}\right)^i\left(1-\tfrac{a}{n}\right)^{n-2-\ell-i} \\
&= \tfrac{n}{a}\cdot\tfrac{1}{n-2-\ell+1}\cdot\sum_{i=0}^{n-2-\ell}\binom{n-2-\ell+1}{i+1}\left(\tfrac{a}{n}\right)^{i+1}\left(1-\tfrac{a}{n}\right)^{n-2-\ell+1-(i+1)} \\
&= \tfrac{1}{a}\cdot\left(1-\tfrac{\ell+1}{n-\ell-1}\right)\cdot\left(1-\mathbb{P}[\mathrm{Bin}(n-2-\ell+1,\tfrac{a}{n})=0]\right) \\
&= \tfrac{1}{a}\cdot\left(1-\tfrac{\ell+1}{n-\ell-1}\right)\cdot\left(1-\left(1-\tfrac{a}{n}\right)^{n-\ell-1}\right) \\
&= \tfrac{1-e^{-a}}{a} + (\ell+1)\cdot O\left(\tfrac{1}{n}\right),
\end{aligned}$$

where above we denoted by $\mathrm{Bin}(m,p)$ a random variable having a binomial distribution with parameters m and p. □

Lemma IV.11. *Consider one round starting with $k < n$ informed nodes. The probability $1-p_k$ that an uninformed node x stays uninformed in this round is at most $(1-\frac{1-e^{-a}}{n})^k + k\cdot O(\frac{1}{n^2})$.*

Proof. Let A be the event that $G\left(n,\frac{a}{n}\right)$ contains no triangle formed by x and two other informed nodes. By the first moment method, $\mathbb{P}[A] \geq 1-k^2\cdot\frac{a^3}{n^3}$. Let X be the indicator random variable for the event that x is called by an informed node. Then

$$\mathbb{P}[X=0] \leq \mathbb{P}[\neg A] + \mathbb{P}[X=0\wedge A] \leq k^2\tfrac{a^3}{n^3} + \mathbb{P}[X=0\wedge A].$$

We compute $\mathbb{P}[X=0\wedge A]$ by conditioning on $\deg_{\inf} x$, which has a binomial distribution with parameters k and $\frac{a}{n}$. In addition, we observe that the conditioning on A makes the actions of the informed neighbors of x independent (in the probability space composed of the random actions of the nodes and the not yet determined random edges). Hence

$$\begin{aligned}
\mathbb{P}[X=0 \mid \deg_{\inf} x = \ell \wedge A] &= (1-\mathbb{P}[y\to x \mid xy\in E\wedge A_{\ell-1}])^\ell \\
&\leq \left(1-\tfrac{1-e^{-a}}{a}+O\left(\tfrac{1}{n}\right)\right)^\ell
\end{aligned}$$

by Lemma IV.10. We compute.

$$\begin{aligned}\mathbb{P}[X=0\wedge A] &= \sum_{\ell=0}^{k}\mathbb{P}[\deg_{inf} x=\ell]\cdot\mathbb{P}[A\mid \deg_{inf} x=\ell]\cdot\mathbb{P}[X=0\mid \deg_{\inf} x=\ell\wedge A]\\ &\le \sum_{\ell=0}^{k}\binom{k}{l}\left(\tfrac{a}{n}\right)^{\ell}\left(1-\tfrac{a}{n}\right)^{k-\ell}\cdot 1\cdot\left(1-\tfrac{1-e^{-a}}{a}+O\left(\tfrac{1}{n}\right)\right)^{\ell}\\ &\le \left[\tfrac{a}{n}\left(1-\tfrac{1-e^{-a}}{a}+O\left(\tfrac{1}{n}\right)\right)+1-\tfrac{a}{n}\right]^{k}\\ &= \left(1-\tfrac{1-e^{-a}}{n}\right)^{k}+k\cdot O\left(\tfrac{1}{n^2}\right).\end{aligned}$$

□

Lemma IV.12. *Consider one round starting with $k<n$ informed nodes. The probability p_k that an uninformed node x becomes informed in the current round is at most $\frac{k}{n}\cdot\left(1-e^{-a}+O\left(\frac{1}{n}\right)\right)$.*

Proof. Consider an uninformed node x and an informed node y. Then, applying Lemma IV.10 with $\ell=0$, we compute

$$\mathbb{P}[y\to x]=\mathbb{P}[xy\in E]\cdot\mathbb{P}[y\to x\mid xy\in E]=\tfrac{a}{n}\cdot\left(\tfrac{1-e^{-a}}{a}+O\left(\tfrac{1}{n}\right)\right).$$

A union bound over the k informed nodes proves the claim. □

Lemma IV.13. *Consider one round starting with $k=\Omega(n)$ informed nodes. The probability $1-p_k$ that an uninformed node x stays uninformed in current round is at least $\left(1-\frac{1-e^{-a}}{n}\right)^{k}-O\left(\frac{\log^2 n}{n}\right)$.*

Proof. Let again A denote the event that $G\left(n,\frac{a}{n}\right)$ contains no cycle of length 3 formed by x and two other informed nodes, and let X be the indicator random variable for the event that x becomes informed. Then $\mathbb{P}[X=0]\ge\mathbb{P}[X=0\wedge A]$. Similar to the proof of Lemma IV.11, we compute $\mathbb{P}[X=0]$ by conditioning on the number $\deg_{\inf} x$ of its informed neighbors.

$$\begin{aligned}\mathbb{P}[X=0\wedge A] &= \sum_{\ell=0}^{k}\mathbb{P}[\deg_{\inf} x=\ell]\cdot\mathbb{P}[A\mid \deg_{\inf} x=\ell]\cdot\mathbb{P}[X=0\mid \deg_{\inf} x=\ell\wedge A]\\ &= \sum_{\ell=0}^{k}\binom{k}{l}\left(\tfrac{a}{n}\right)^{\ell}\left(1-\tfrac{a}{n}\right)^{k-\ell}\cdot\left(1-\tfrac{a}{n}\right)^{\ell^2}\cdot\left(1-\tfrac{1-e^{-a}}{a}-(\ell+1)\cdot O\left(\tfrac{1}{n}\right)\right)^{\ell}\end{aligned}$$

To simplify the notation, we denote $x_\ell := \binom{k}{l}\left(\frac{a}{n}\right)^\ell\left(1-\frac{a}{n}\right)^{k-\ell}$ and $q := 1-\frac{1-e^{-a}}{a}$. Then

$$\begin{aligned}\mathbb{P}[X=0\wedge A] &\geq \sum_{\ell=0}^{c\log n} x_\ell\cdot\left(1-\tfrac{a}{n}\right)^{\ell^2}\cdot\left(q-\ell\cdot O\left(\tfrac{1}{n}\right)\right)^\ell\\ &\geq \sum_{\ell=0}^{c\log n} x_\ell\cdot\left(1-\tfrac{a}{n}\right)^{c^2\log^2 n}\left(q-O\left(\tfrac{\log n}{n}\right)\right)^\ell\\ &\geq \left(1-O\left(\tfrac{\log^2 n}{n}\right)\right)\sum_{\ell=0}^{c\log n} x_\ell q^\ell.\end{aligned}$$

We recall the maximum vertex degree of $G(n,\frac{a}{n})$ is at most $O(\log n)$ with high probability. Thus, from a simple Chernoff bound it follows that there exists $c>0$ such that $\sum_{\ell=c\log n}^{k} x_\ell q^\ell \leq \frac{1}{n}$. Since $\sum_{\ell=0}^{k} x_\ell q^\ell = \left(1-\frac{1-e^{-a}}{n}\right)^k$, we have

$$\mathbb{P}[X=0\wedge A] \geq \left(1-O\left(\tfrac{\log^2 n}{n}\right)\right)\left(1-\tfrac{1-e^{-a}}{n}\right)^k.$$

□

Lemma IV.14. *Consider a round starting with k informed nodes. Let x_1 and x_2 be two uninformed nodes. Then the corresponding random indicator variables X_1 and X_2 for the events of these becoming informed are negatively correlated.*

Proof. By symmetry, we can assume that in this round we first generate the random communication graph, then we let each node choose a potential communication partner (uniformly among its neighbors), and then we decide randomly which k nodes are informed, and finally those nodes which are informed actually call the potential partner chosen before. In this joint probability space, let x_1 and x_2 be two nodes. We condition in the following on (i) the outcome of the random graph, (ii) the outcome of the potential communication partners, and (iii) x_1 and x_2 being uninformed. In other words, all randomness is already decided except which set I of k nodes different from x_1 and x_2 is informed.

Let S_1 and S_2 be the sets of nodes having chosen x_1 and x_2 as potential partner. Now we have $X_1 = 1$ if and only if $S_1 \cap I \neq \emptyset$. Similarly, $X_2 = 1$ is equivalent to $S_2 \cap I \neq \emptyset$. Since $S_1 \cap S_2 = \emptyset$ by construction, X_1 and X_2 are negatively correlated. □

Theorem IV.15. *The expected rumor spreading time for the push protocol in the dynamic $G(n,\frac{a}{n})$ graph is*

$$\log_{2-e^{-a}} n + \tfrac{1}{1-e^{-a}}\ln n \pm O(1).$$

In addition, there are constant $A'\alpha' > 0$ such that for any $r \in \mathbb{N}$ we have $\mathbb{P}[|T - \mathbb{E}[T]| \geq r] \leq A'e^{-\alpha' r}$.

Proof. By Lemma IV.14, the covariance conditions are satisfied for both exponential growth and exponential shrinking.

From Lemma IV.11 together with Corollary B.3 it follows that for any $k < n$ we have

$$p_k \geq \tfrac{k}{n}\left(1 - e^{-a}\right) - \tfrac{k^2}{2n^2}\left(1 - e^{-a}\right)^2 - k \cdot O\left(\tfrac{1}{n^2}\right).$$

Combining this with Lemma IV.12, we see that the process satisfies the exponential growth conditions with $\gamma_n = 1 - e^{-a}$ in interval $[1, fn]$ for any constant $0 < f < 1$.

For $k = \Theta(n)$, Lemma IV.11 and Lemma IV.13 yield that

$$\left(1 - \tfrac{1-e^{-a}}{n}\right)^k - O\left(\tfrac{\log^2 n}{n}\right) \leq 1 - p_k \leq \left(1 - \tfrac{1-e^{-a}}{n}\right)^k + k \cdot O\left(\tfrac{1}{n^2}\right).$$

Substituting k by $n - u$ and applying Corollary B.6, we obtain for any $u < n$ that

$$\exp\left(e^{-a} - 1\right) - O\left(\tfrac{\log^2 n}{n}\right) \leq 1 - p_{n-u} \leq \exp\left(e^{-a} - 1\right)\left(1 + 2\left(1 - e^{-a}\right)\tfrac{u}{n}\right) + O\left(\tfrac{1}{n}\right).$$

Therefore, the protocol satisfies the upper exponential shrinking conditions with $\rho_n = 1 - e^{-a}$ and the lower exponential shrinking conditions with $\rho_n = 1 - e^{-a} + O\left(\tfrac{\log^2 n}{n}\right)$ in the interval $[n - gn, n]$ for any $0 < g < 1$.

Since the intervals of exponential growth and exponential shrinking overlap, it follows from Theorems III.7, III.14, III.19, and III.28 that the expected spreading time $\mathbb{E}[T]$ is equal to $\log_{1-e^{-a}} n + \tfrac{1}{1-e^{-a}} \ln n \pm O(1)$ and $\mathbb{P}[|T - \mathbb{E}[T]| \geq r] \leq A'e^{-\alpha' r}$ for suitable constants $A', \alpha' > 0$. □

3 Limited Incoming Calls Capacity

For all the protocols discussed above the nodes are allowed to be called several times in one round. For some processes such as protocols considered in Section 2.3, the number of calls received by each node is at most constant. However in most of rumor spreading processes such number can be unbounded. For example, consider the basic push-pull protocol from Section 1.3 on the complete graph with n vertices. Since each round all nodes make calls, the maximum number of incoming calls received by the same node in one round is the same as the maximum load of a bin in the well-known problem of throwing uniformly and independently at random n balls into n bills, i.e., $\frac{\log n}{\log\log n} \cdot (1 + o(1))$ (see Appendix C). Such phenomenon can impact the scalability of the rumor spreading process: typically the time gap between rounds is bounded, but each round with high probability there is at least one node which have to finish $\omega(1)$ transactions.

The simplest solution is to limit the incoming "capacity" of nodes, i.e., the number of calls they can reply in one round. In this section we propose a *single incoming call* setting – any node can reply to only one incoming call per round chosen uniformly at random among all received calls in current round. All other calls are considered "dropped", i.e., they cannot transfer the rumor. Therefore, each node participates in at most two rumor transactions per round, whatever is the size of the network.

On the other hand, we expect the noticeable slowdown for the protocols based on the single incoming call setting compared to the usual unlimited "capacity" setting. Thus we will show in Section 3.1 that the single incoming call push-pull protocol satisfies the single exponential shrinking conditions instead of double exponential shrinking and the corresponding expected rumor spreading time is equal to $\log_{3-2/e} n + \frac{1}{2}\ln n \pm O(1)$. In Section 3.2 we argue that since $\Theta(n)$ nodes are informed, the push calls of informed nodes becomes inefficient and they are responsible for such considerable slowdown. Finally, in Section 3.3 we combine a single incoming call push-pull protocol with pull protocol and provide a not memoryless process with spreading time $\log_{3-2/e} n + \log_2 \ln n + O(1)$.

Before proceeding to the computations, we observe that the following setting is equivalent to the single incoming call model. In each round we choose uniformly at random a permutation $\sigma \in S_n$. The element σ_n is the *order* of the outgoing call of node x_i, we write $ord_i = \sigma_i$. Each node accepts the call with the lowest order among its received incoming calls. We call such construction the *ordered calls* setting.

3.1 Single Incoming Call Push-Pull Protocol

The following theorem contains the main result of this section and will be proved at the end.

Theorem IV.16. *The expected spreading time for the single incoming call push-pull protocol is*

$$\log_{3-2/e} n + \tfrac{1}{2}\ln n + O(1).$$

There also exist constants A', α' such that $\mathbb{P}[|T - \mathbb{E}[T]| > r] \leq A'e^{-\alpha' r}$.

In this section we keep the notation from the previous ones, i.e. X_i is the random indicator variable corresponding to the event "uninformed node x_i gets informed in the considered round". As usual, we denote by p_k the probability $\mathbb{P}[X_i = 1]$ for the round started with k informed nodes and any i. In addition we denote by Y_i, Z_i the indicator random variables for the following events.

Y_i "Node i is called and the first incoming call comes from an informed node."

Z_i "The outgoing call of node i is accepted by an informed node."

Lemma IV.17. *Suppose that the fraction f of nodes is informed. Suppose node i is uninformed. Then*

$$p_{fn} = 2f\left(1 - \tfrac{1}{e}\right) - f^2\left(1 - \tfrac{1}{e}\right)^2 + f \cdot O\left(\tfrac{1}{n}\right). \tag{IV.8}$$

Proof. First, we compute the probabilities of the events corresponding to Y_i and Z_i. Since each node makes a call in the round, the probability that node x_i is not called is equal to $(1 - \frac{1}{n})^n$. Therefore,

$$\mathbb{P}[Y_i = 1] = f\left(1 - \left(1 - \tfrac{1}{n}\right)^n\right) = f\left(1 - \tfrac{1}{e}\right) + f \cdot O\left(\tfrac{1}{n}\right).$$

To compute $\mathbb{P}[Z_i = 1]$ we will use the ordered call model. Suppose that $ord_i = \ell$. Then, the outgoing call of node x_i is accepted if all calls with orders less than ℓ do not call the same node. Since the probability that the outgoing call of node x_i has order ℓ is equal to $\frac{1}{n}$, we compute

$$\mathbb{P}[Z_i = 1] = f\sum_{\ell=1}^{n} \tfrac{1}{n}\left(1 - \tfrac{1}{n}\right)^{\ell-1} = f\left(1 - \left(1 - \tfrac{1}{n}\right)^n\right) = f\left(1 - \tfrac{1}{e}\right) + f \cdot O\left(\tfrac{1}{n}\right).$$

Since $X_i = \max\{Y_i, Z_i\}$, it remains to compute the probability of the event $Y_i = Z_i = 1$. Suppose that $ord_i = \ell$. Since the outgoing call of node x_i is accepted, all calls with order less than ℓ should go away from the x_i's target, i.e., they can have only $n-1$ possible targets. We also remark that node x_i calls informed node,

so it cannot call itself. Thus the probability that nobody calls node x_i is equal to $\left(1-\frac{1}{n-1}\right)^{i-1}\left(1-\frac{1}{n}\right)^{n-i}$. Therefore,

$$\begin{aligned}\mathbb{P}[Z_i = 1 | Y_i = 1,\ ord_i = \ell] &= f\left(1-\left(1-\tfrac{1}{n-1}\right)^{i-1}\left(1-\tfrac{1}{n}\right)^{n-i}\right)\\ &= f\left(1-\left(1-\tfrac{1}{n}\right)^{n}+O\left(\tfrac{1}{n}\right)\right).\end{aligned}$$

Since the probability above is independent of ℓ, we obtain immediately that node

$$\begin{aligned}\mathbb{P}[Y_i = Z_i = 1] &= f^2\left(1-\left(1-\tfrac{1}{n}\right)^{n}\right)^2 + f^2\cdot O\left(\tfrac{1}{n}\right)\\ &= f^2\left(1-\tfrac{1}{e}\right)^2 + f^2\cdot O\left(\tfrac{1}{n}\right).\end{aligned}$$

The claim of lemma follows by including-excluding formula. □

Lemma IV.18. *There exists $c \geq 0$ such that for any uninformed nodes $x_i \neq x_j$ we have*

$$\mathbb{P}[X_i = 1 | X_j = 1] \leq \mathbb{P}[X_i = 1] + \tfrac{c}{n}. \tag{IV.9}$$

Proof. We say that nodes x_i and x_j *interact* if one calls another or if they both call the same node. Clearly, $\mathbb{P}[x_i, x_j \text{ interact} | X_j = 1] = O\left(\frac{1}{n}\right)$. Since we need to bound $\mathbb{P}[X_i = 1 | X_j = 1]$ up to $O(\frac{1}{n})$, without loss of generality we assume for the rest of the proof that nodes x_i and x_j do not interact. We say that a call interacts with a node x_j if its target coincides with x_j or with x_j's target (by convention a call does not interact with it source). Denote by I_j the number of calls interacting with node x_j and observe that since x_i and x_j don't interact, no node can interact with both x_i and x_j. We split the probability $\mathbb{P}[X_i = 1 | X_j = 1]$ conditioning on the values of I_j as follows.

$$\mathbb{P}[X_i = 1 | X_j = 1] = \sum_{k=1}^{n} \mathbb{P}[X_i = 1 | X_j = 1, I_j = k] \cdot \mathbb{P}[I_j = k | X_j = 1].$$

Our goal is to study $\mathbb{P}[X_i = 1 | X_j = 1, I_j = k]$. Since k nodes interact with x_j, there are $n-k-1$ independent calls going uniformly to $n-2$ remaining targets (except x_j and x_j's target). In addition at least $n(f - \frac{k+1}{n})$ of calls are made by informed nodes. By these two observations we deduce

$$\begin{aligned}\mathbb{P}[Y_i = 1 | X_j = 1, I_j = k] &= \left(f-\tfrac{k+1}{n}\right)\left(1-(1-\tfrac{1}{n-2})^{n-k-1}\right)\\ &= f\left(1-\left(1-\tfrac{1}{n}\right)^{n}\right) + kO\left(\tfrac{1}{n}\right) = f\left(1-\tfrac{1}{e}\right) + k\cdot O\left(\tfrac{1}{n}\right).\end{aligned}$$

By the similar analysis we obtain that

$$\begin{aligned}\mathbb{P}[Z_i = 1 | X_j = 1, I_j = k] &= f\left(1-\tfrac{1}{e}\right) + k\cdot O\left(\tfrac{1}{n}\right);\\ \mathbb{P}[Y_i = Z_i = 1 | X_j = 1, I_j = k] &= f^2\left(1-\tfrac{1}{e}\right)^2 + k\cdot O\left(\tfrac{1}{n}\right).\end{aligned}$$

Therefore, $\mathbb{P}[X_i = 1 | X_j = 1, I_j = k] = \mathbb{P}[X_i = 1] + k \cdot O\left(\frac{1}{n}\right)$. Since $\mathbb{E}[I_j | X_j = 1] = O(1)$, we sum up by k and obtain

$$\begin{aligned}\mathbb{P}[X_i = 1 | X_j = 1] &= \mathbb{P}[X_i = 1] + \sum_{k=1}^{n} kO\left(\tfrac{1}{n}\right) \cdot \mathbb{P}[I_j = k | X_j = 1] \\ &= \mathbb{P}[X_i = 1] + O\left(\tfrac{1}{n}\right) \mathbb{E}[I_j | X_j = 1] = \mathbb{P}[X_i = 1] + O\left(\tfrac{1}{n}\right).\end{aligned}$$

□

Proof of Theorem IV.16. Consider a round started with k informed nodes. Substituting f by k/n in (IV.8), we obtain the probability part of the exponential growth conditions.

$$p_k = 2\left(1 - \tfrac{1}{e}\right) \cdot \tfrac{k}{n} + k^2 \cdot O\left(\tfrac{1}{n^2}\right).$$

Multiplying (IV.9) by p_k we get the covariance condition. Therefore the protocol satisfies the exponential growth conditions with $\gamma_n = 2(1 - \frac{1}{e})$.

Denote by $u := n - k$ the number of uninformed nodes. Substituting f by $1 - \frac{u}{n}$ in (IV.8), we compute

$$\mathbb{P}[X_i = 0] = 1 - \mathbb{P}[X_i = 1] = \tfrac{1}{e^2} + O\left(\tfrac{1}{n}\right).$$

Since the covariance condition follows from Lemma IV.18, the protocol satisfies the exponential shrinking conditions with $\rho_n = 2$. Therefore the expected spreading time is equal to $\log_{3-2/e} n + \frac{1}{2} \ln n + O(1)$. □

3.2 Single Incoming Call Pull-Only Protocol

We showed that the the single call push-pull protocol is significantly slower than the classic push-pull protocol. Although protocol based on the single incoming call setting cannot be faster than the classic independent call model, we can make it noticeably faster using the following trick. Let us consider one round of the exponential shrinking phase with u uninformed nodes. In such round there are $n - u$ push calls, each one hits uninformed node with small probability $\frac{u}{n}$. On the other hand, each of u pull calls touches some informed node with probability $1 - \frac{u}{n}$. One can conclude that push calls "spam" the network: they "occupy" other informed nodes making them inaccessible for pull calls of uninformed nodes. This observation is verified in the following theorem.

Theorem IV.19. *The spreading time for the single incoming call pull protocol is*

$$\log_{2-1/e} n + \log_2 \ln n + O(1).$$

There also exist constants A', α' such that $\mathbb{P}[|T - \mathbb{E}[T]| > r] \leq A' e^{-\alpha' r}$.

Proof. Consider one round of the protocol. Clearly, if x_1 becomes informed it "occupies" one informed node which cannot inform any other node in current round. Thus, if we condition on that $X_1 = 1$, then it is slightly less likely that x_2 becomes informed. Consequently, $\mathrm{Cov}[X_1, X_2] < 0$ and the covariance part of the exponential growth and double exponential shrinking conditions is satisfied.

Again, the call with order ℓ is accepted with probability $\left(1 - \frac{1}{n}\right)^{\ell-1}$. Since in the round started with k informed nodes only $n - k$ nodes perform calls, ord_i is uniformly distributed in $\{1, \dots, n-k\}$. Since the probability to call an informed node is $\frac{k}{n}$, we compute

$$p_k = \frac{k}{n} \sum_{\ell=1}^{n-k} \frac{1}{n-k} \left(1 - \frac{1}{n}\right)^{\ell-1} = \frac{k}{n-k} \left(1 - \left(1 - \frac{1}{n}\right)^{n-k}\right). \tag{IV.10}$$

By Corollary B.5, we have

$$\left(1 - \tfrac{1}{e}\right) \tfrac{k}{n} - 4\tfrac{k^2}{n^2} \le p_k \le \left(1 - \tfrac{1}{e}\right) \tfrac{k}{n} + 2\left(1 - \tfrac{1}{e}\right) \tfrac{k^2}{n^2}.$$

So the protocol satisfies the exponential growth conditions with parameter $\gamma_n = 1 - \frac{1}{e}$.

If we denote by u the number of uninformed nodes, from (IV.10) follows the following expression.

$$1 - p_{n-u} = \tfrac{n-u}{u} \left(1 - \left(1 - \tfrac{1}{n}\right)^u\right).$$

With Lemma B.2, we estimate $\frac{u}{n} \le 1 - p_{n-u} \le \frac{3u}{2n}$. The protocol hence satisfies the double exponential shrinking conditions with $\ell = 2$.

Therefore, the expected spreading time is equal to $\log_{2-1/e} n + \log_2 \ln n + O(1)$. □

3.3 Push-Pull Protocol with Transition Time

Comparing Theorems IV.16 and IV.19 we see that push-pull protocol still is more efficient until $\Theta(n)$ nodes are informed. Suppose now that we join to the rumor a counter which increases by one each round, so that each informed node knows the "age" of the rumor. Then the single incoming call push-pull protocol *with transition time* $R > 0$ acts as follows. While the age of the rumor is at most R, it acts as a single incoming call push-pull protocol. After R rounds of rumor spreading, all informed nodes stop calling simultaneously, so the protocol acts as the single incoming call pull protocol until nodes are informed.

Theorem IV.20. *The expected rumor spreading time of the single incoming call push-pull protocol with the transition time* $R = \lceil \log_{3-2/e} n \rceil$ *on the complete graph with* n *vertices is*

$$\log_{3-2/e} n + \log_2 \ln n + O(1).$$

There also exist constants A', α' such that $\mathbb{P}[|T - \mathbb{E}[T]| > r] \leq A'e^{-\alpha' r}$.

Proof. In the proof of Theorem IV.19 we showed that the single incoming call pull protocol satisfies the double exponential shrinking conditions for all $k \in [gn, n]$ for some $0 < g < 1$. Denote by I_t the number of informed nodes after t rounds. Let $t := \max\{R, t'\}$, where t' is the smallest time such that $I_{t'} \geq gn$. By construction, after round t the transition protocol acts as the pull protocol. Therefore,

$$\mathbb{E}[T(1,n)] \leq \mathbb{E}[t] + \mathbb{E}[T(fn,n)] \leq \mathbb{E}[t] + \log_2 \ln n + O(1).$$

It is easy to see that the transition protocol satisfies the conditions of Lemma III.4 with $\ell = fn$, $m = gn$ for any $0 < f < g < 1$. Thus, $\mathbb{E}[t] \leq \mathbb{E}[T(1, fn)] + O(1)$ for any constant $0 < f < 1$, i.e., it suffices to analyse the spreading time until fn informed nodes.

Let us consider a single incoming call push-pull protocol. In the proof of Theorem IV.16 we showed that the single incoming call push-pull protocol satisfies the exponential growth conditions with $\gamma_n = 2 - \frac{2}{e}$. In Section 2.1 we introduced a sequence k_j splitting the interval $[1, fn]$ into phases such that most of the rounds the rumor spreading process moves to exactly the next phase. Lemma III.10 claims that the biggest number of phase $J = \log_{1+\gamma_n} n + O(1)$. Since $\gamma_n = 2 - \frac{2}{e}$, we have $J = R + O(1)$. To simplify the proof we suppose that $R \leq J$ and $fn \leq k_R$.In the proof of Theorem III.7 we showed that $T(1, k_R) \leq R + \Delta r$, where Δr is stochastically dominated by a random variable with distribution $\mathrm{Geom}(1-q)$ for some constant $q < 1$. By construction, Δr is the number of rounds during which the process stayed it the same phase. Therefore, after at the end of round R when the protocol switches from push-pull to pull-only, we have $I_R \geq k_{R-\Delta r}$. By Lemma III.10, we have $k_{R-\Delta r} \geq \frac{fn}{(3-2/e)^{\Delta r}}$.

Consider now the single incoming call pull protocol. Let a sequence k'_j defines the phases for the single incoming call pull protocol. Suppose that R' is such that $k_{R'} \geq fn$ and that I_R belongs to the phase i of the single incoming call pull protocol. Since the single incoming call pull protocol satisfies the exponential growth conditions with $\gamma_n = 1 - 1/e$, we have $R' - i = \frac{3-2/e}{2-1/e}\Delta r + O(1)$. Therefore,

$$\mathbb{E}[T(I_R, fn)] \leq \mathbb{E}[T(k'_i, k'_{R'})] \leq \tfrac{3-2/e}{2-1/e}\Delta r + O(1).$$

Summing over all possible values of Δr we compute

$$\mathbb{E}[T(1, fn)] \leq r + \sum_{s=0}^{R} \mathbb{P}[\Delta r = s] \cdot \left(\tfrac{3-2/e}{2-1/e} s + O(1)\right) = R + O(1)$$

Since Δr is dominated by a random variable with distribution $\mathrm{Geom}(1-q)$, we have $\mathbb{E}[T(1, fn)] \leq R + O(1)$. Therefore, $\mathbb{E}[T(1,n)] \leq \log_{3-2/e} n + \log_2 \ln n + O(1)$.

To prove the lower bound we consider the following protocol. Suppose that any node knows the total number of informed nodes. The protocol acts as the single incoming call push-pull protocol until there are at least fn informed nodes for some $0 < f < 1$. Then the protocol acts as the single incoming call pull protocol. Since we proved Theorems IV.16 and IV.19, the expected spreading time of such protocol is at least $\log_{3-2/e} n + \log_2 \ln n + O(1)$. It is also easy to see that such protocol spreads the rumor slightly quicker that the protocol with the fixed transition time, so the expected spreading time is bounded from below by the same expression. □

Chapter V

Non-Homogeneous Rumor Spreading

Contents

1	Homogeneous Rumor Spreading Processes		**39**
2	Exponential Growth Regime		**42**
	2.1	Upper Bound	42
	2.2	Lower Bound	49
3	Exponential Shrinking Regime		**55**
	3.1	Upper Bound	55
	3.2	Lower Bound	60
4	Double Exponential Shrinking Regime		**63**
	4.1	Upper Bound.	63
	4.2	Lower Bound.	67

In Chapter III we restricted ourselves to the homogeneous rumor spreading processes. We recall that these are two-state memoryless processes requiring that all uninformed nodes learn the rumor with the same probability which depends only on the number k of currently informed nodes. This uniformity leads us to the studying the rumor spreading in the complete graphs. However, many processes cannot be analyzed by our method even in the complete graph. We cannot apply our method to the multi-state processes such as median counter algorithm proposed by Karp et al. [KSSV00], neither to the non-symmetric protocols such as the Panagiotou's et al. [PPS15] multiple call protocol, when nodes do not change their outgoing call capacity during the process. On the over hand, we could see in Section 2.3, Chapter IV that the complete networks can be relaxed to the dynamic ones, but the memorylessness property requires that the network is a random graph sampled independently each round.

In the rest of the chapter, we discuss how can we overcome the restriction to the homogeneous rumor spreading processes. We will show that the phase method still be promising for the *multi-parametric* protocols that require a constant amount of memory in addition to the number of informed nodes for the description of their progress.

1 Two-State Multi-Parametric Process

The first multi-parametric protocol that we consider is similar to the multiple call protocol considered in Chapter IV, Section 2.2. Assume that R is a random integer variable taking values in $[0, n[$ such that $\mathbb{E}[R] = \Theta(1)$ and $\mathrm{Var}[R] = O(1)$. When an uninformed node learns the rumor, it samples a value r from R independently from other nodes, keeps it and makes r push calls per round. In addition, each uninformed node makes one pull call per round. For the push calls, this process follows the setting of [PPS15]: each node has different capacity of outgoing calls that is unchanged during all rounds of the process. Unlike in [PPS15], for the pull calls we assume that any uninformed node makes exactly one call per round. The last assumption implies that the protocol above is faster than the basic pull protocol. Consequently, for any $f \in]0, 1[$, we have $\mathbb{E}[T([fn], n)] \leq \log_2 \ln n + O(1)$. Thus, it suffices to study only $\mathbb{E}[T(1, fn)]$ for some $f \in]0, 1[$.

Consider one round of the protocol starting from k informed nodes. We observe that the number ℓ of push calls in this round is different from k and unknown. Therefore, the success probability $p_{k\ell}$ that an uninformed node becomes informed depends on two parameters, k and ℓ, as follows.

$$p_{k\ell} = 1 - \left(1 - \tfrac{1}{n}\right)^{\ell} \left(1 - \tfrac{k}{n}\right)$$

In this section, we will show that the number of informed nodes in this protocol performs the exponential growth. Instead of directly applying Theorem III.7, we will repeat the main steps of its proof to show that the statements similar to ones from Chapter III, Section 2.1 hold for this multi-parametric protocol.

Theorem V.1. *The expected rumor spreading time for the protocol described above is at most* $\log_{2+\mathbb{E}[R]} n + \log_2 \ln n + O(1)$.

As usual, we will first state all preliminary lemmas, and state the proof of this theorem at the end of the section.

Round Targets and Failure Probabilities

Consider an uniformed node in a round starting with $k < n/2$ informed nodes that make $\ell \leq k$ calls in current round. First, using the estimates from Lemma B.2, we

obtain

$$p_{k\ell} \geq \frac{\ell}{n} + \frac{k}{n} - \frac{(\ell+k)^2}{2n^2}.$$

Denote by $X(k,\ell)$ the number of newly informed nodes in current round. Since $\mathbb{E}[X] = (n-k)p_{k\ell}$, the expression above implies that

$$\mathbb{E}[X] \geq E(k,\ell) := k + \ell - \tfrac{3}{2} \cdot \frac{(k+\ell)^2}{n}.$$

Now we construct the target value $X_0(k,\ell)$ for the number of newly informed nodes such that $X(k,\ell) \geq X_0(k,\ell)$ with good probability. Let $A_1 > 0$ and $B \in]1/2, 1[$ are two constants to be chosen later. Similarly to (III.2), the target value X_0 is defined as follows.

$$X_0(k,\ell) := E(k,\ell) - A_1(k+\ell)^B. \tag{V.1}$$

Similarly to Lemma III.9, we bound the probability that X fails to achieve the target in one round.

Lemma V.2. *There exists $q(k,\ell) = O(1) \cdot (k+\ell)^{-2B+1}$ such that*

$$\mathbb{P}[X(k,\ell) \leq X_0(k,\ell)] \leq q(k,\ell).$$

Sketch of the proof. Using Chebyshev's inequality we compute

$$\begin{aligned}
\mathbb{P}[X(k,\ell) \leq X_0(k,\ell)] &= \mathbb{P}\left[X(k,\ell) \leq E(k,\ell) - A_1(k+\ell)^B\right] \\
&\leq \mathbb{P}\left[X(k,\ell) \leq \mathbb{E}[X(k,\ell)] - A_1(k+\ell)^B \cdot \tfrac{\mathbb{E}[X(k,\ell)]}{E(k,\ell)}\right] \\
&\leq \tfrac{\mathrm{Var}[X(k,\ell)]}{(A_1(k+\ell)^B)^2} \cdot \tfrac{E(k,\ell)^2}{\mathbb{E}[X(k,\ell)]^2}.
\end{aligned}$$

Since the covariance condition is satisfied for both basic pull and push-pull protocols, the same holds for the pull protocol with 1 push call. Thus, we have $\mathrm{Var}[X(k,\ell)] \leq \mathbb{E}[X] + O(1) \cdot (k+\ell)$. Therefore, the computation similar to one in the proof of Lemma III.9, implies that the failure probability is bounded by $O(1) \cdot (k+\ell)^{-2B+1}$. □

Each newly informed node i samples a random number r_i from R that contributes to the total number ℓ of push calls. Denote by $Y := \sum_{i=1}^{X} r_i$ the increase of this number at the end of current round. Since X and r_i are independent, we have $\mathbb{E}[Y] = \mathbb{E}[X] \cdot \mathbb{E}[R]$ and $\mathrm{Var}[Y] \leq \mathbb{E}[Y]$. Similarly to (V.1), we define the target value Y_0 for Y as follows.

$$Y_0(k,\ell) := \mathbb{E}[R] \cdot k + \mathbb{E}[R] \cdot \ell - \tfrac{3}{2}\mathbb{E}[R] \cdot \tfrac{(k+\ell)^2}{n} - A_2(k+\ell)^B. \tag{V.2}$$

Similarly to Lemma V.2, one can show using Chebyshev's inequality that

$$\mathbb{P}[Y \leq Y_0] = O(1) \cdot (k+\ell)^{-2B+1}. \tag{V.3}$$

The Phase Calculus

Since the progress of the protocol is described by two parameters, the number k of informed nodes and the number ℓ of push calls, we should construct the sequence of round targets for both parameters. This sequence is defined recursively as follows.

$$k_0 := 1, \quad k_{j+1} := k_j + X_0(k_j, \ell_j);$$
$$\ell_0 := 1, \quad \ell_{j+1} := \ell_j + Y_0(k_j, \ell_j).$$

This seems to be the typical structure of phase targets for the multi-parametric process: the recurrent sequences of round targets for k and ℓ are entangled with each other. To transform this into the more similar form to (III.3), we will employ the fact that the sequence above is almost linear. Denote by $\mathbf{x}_j := \binom{k_j}{\ell_j}$ the vector of the parameters and let

$$M := \begin{pmatrix} 2 & 1 \\ \mathbb{E}[R] & \mathbb{E}[R] \end{pmatrix}, \quad \mathbf{u} := \begin{pmatrix} 3/2 \\ 3\mathbb{E}[R]/2 \end{pmatrix}, \quad \mathbf{v} := \begin{pmatrix} A_1 \\ A_2 \end{pmatrix}.$$

Then, the recursive expression above can be written as

$$\mathbf{x}_{j+1} := M\mathbf{x}_j - \tfrac{1}{n}\mathbf{u}\|\mathbf{x}_j\|^2 - \mathbf{v}\|\mathbf{x}_j\|^B,$$

where $\|\mathbf{x}_j\| := k_j + \ell_j$ denotes the Manhattan norm of $\mathbf{x}_j$. By the elementary computation, we obtain that M has eigenvalues 1 and $2 + \mathbb{E}[R]$ and that the corresponding eigenvectors both has positive coordinates. Now we can reformulate Lemma III.10 for our protocol.

Lemma V.3. *After possibly lowering A_1 and A_2 from* (V.1)*, there exists $J = \log_{2+\mathbb{E}[R]} n + O(1)$ such that for any $j < J$ we have $k_j = \Theta(1)(2 + \mathbb{E}[R])^j$ and $\ell_j = \Theta(1)(2 + \mathbb{E}[R])^j$.*

The proof repeats ideas of the proof for Lemma III.10. By convention, we use symbols $\leqslant$ and $\geqslant$ for the element-wise comparison between vectors.

Proof. To ease the notation, we denote $\lambda := 2 + \mathbb{E}[R]$. By induction we obtain for all $j \geq 0$ that

$$\mathbf{x}_j = M^j\mathbf{x}_0 - \tfrac{1}{n}\sum_{i=1}^{j} M^i\mathbf{u}\|\mathbf{x}_{j-i}\|^2 - \sum_{i=1}^{j} M^i\mathbf{v}\|\mathbf{x}_{j-i}\|^B. \tag{V.4}$$

Let $J := \log_\lambda(fn) - \Delta r$ for some positive $\Delta r = O(1)$ determined later. Since A and elements of M are all nonnegative, we have $\mathbf{x}_j \leqslant M^j\mathbf{x}_0$ element-wise. Thus, $k_j = O(1)\lambda^j$ and $\ell_j = O(1)\lambda^j$.

We show by induction on j that there exists a constant vector $\mathbf{w}$ such that $\mathbf{x}_j \geqslant \lambda^j \mathbf{w}$ for all $j \leq J$. The base for $j = 0$ and $\mathbf{x}_0 = \binom{1}{1}$ is trivial. Let $1 \leq j \leq J$ and suppose $\mathbf{x}_i \geq \mathbf{w}\lambda^i$ for all $i < j$. Since the elements of M^i are $O(1) \cdot \lambda^i$, we compute both sums in (V.4).

$$\sum_{i=1}^{j} M^i \mathbf{v} \|\mathbf{x}_{j-i}\|^B = O(\mathbf{v}) \sum_{i=1}^{j} \lambda^i \cdot \lambda^{(j-i)B}$$
$$= O(\mathbf{v})\lambda^{jB} \sum_{i=1}^{j} \lambda^{(1-B)i} = O(\mathbf{v})\lambda^j.$$

By choosing A_1 and A_2 sufficiently small, we can make the coefficient at λ^j arbitrary small. To compute the second sum, we use the same idea together with the fact that $\frac{\lambda^j}{n} \leq \lambda^{-\Delta r}$ for all $j \leq J$.

$$\tfrac{1}{n} \sum_{i=1}^{j} M^i \mathbf{u} \|\mathbf{x}_{j-i}\|^2 = \tfrac{O(\mathbf{u})}{n} \sum_{i=1}^{j} \lambda^i \cdot \lambda^{2j-2i}$$
$$= \tfrac{O(\mathbf{u})}{n} \lambda^{2j} \sum_{i=1}^{j} \lambda^{-i} = O(\mathbf{u})\lambda^{-\Delta r} \cdot \lambda^j$$

By choosing Δr large enough, we can make the coefficient at λ^j arbitrary small. Therefore, $\mathbf{x}_j = M^j \mathbf{x}_0 - O(\mathbf{v})\lambda^j - O(\mathbf{u})\lambda^{-\Delta r} \cdot \lambda^j$. Thus we see that there exists $\mathbf{w}$ such that $\mathbf{x}_j \leq \mathbf{w}\lambda^j$. □

Now we are ready to estimate the rumor spreading time.

Proof of Theorem V.1. By Lemma V.3, there exists $J = \log_{2+\mathbb{E}[R]} n + O(1)$ such that both $k_J = \Theta(n)$ and $\ell_J = \Theta(n)$. The round targets k_j (resp. ℓ_j) cut the interval $[1, k_J]$ (resp. $[1, \ell_J]$) into J phases. We say that the process achieves phase j, if $\mathbf{x} \geqslant \mathbf{x}_j$, i.e., both $k \geq k_j$ and $\ell \geq \ell_j$. Note that this is another difference between the homogeneous rumor spreading and the multi-parametric process: the round is failed if any (not all) of the parameters does not reach its target value.

After some elementary computation, one can see that the probability that the process fails the round being in phase j is at most $q(k_j, \ell_j)$. Since by definition of the process, the number k of informed nodes, as well as the number ℓ of push calls never decreases, the number T_j of rounds to leave phase j is stochastically dominated by $1 + \mathrm{Geom}(1 - q(k_j, \ell_j))$. Consequently, the number of rounds until either k reaches k_J or ℓ reaches ℓ_J[1] is stochastically dominated by $\sum_{j=1}^{J} T_J$.

[1]It is easy to see that if the number ℓ of push calls has reached the final phase ℓ_J, then the next round at least $\Theta(n)$ nodes will be informed with high probability via only ℓ_J push calls.

By Lemma V.2 and V.3, the expected number of rounds before $\Theta(n)$ nodes are informed is at most $J + O(1)$.

Finally, since the protocol performs the double exponential shrinking, we obtain that $\mathbb{E}[T(1,n)] \leq \log_{2+\mathbb{E}[R]} n + \log_2 \ln n + O(1)$. □

2 Multi-State Rumor Spreading. Independent Stop Process

Another natural and more interesting example of the multi-parametric process is the *multi-state* rumor spreading, when the nodes can be in more than two possible states performing different behavior. For the motivation of the multi-state rumor spreading, we refer to the stoppage problem: how informed nodes can determine that most of the nodes in the network are informed, so that they can switch from the *active* state when they make push calls to the *passive* state when they only reply the pull requests. This requires at least three states: one uninformed and two informed states, active, and passive. Thus, the rumor spreading protocol that solves this problem has to be multi-parametric[2].

In this section we will consider the *independent stop protocol* in which an active informed node can become passive, but passive nodes never become active again. Unlike the protocol with the transition time, in the independent stop protocol each node makes the decision independently from others to become passive. We also require the protocol to be symmetric, so that the probability that one node transforms from its state A to some state B does not depend on the choice of the node, but only on the number of nodes in each state. Since the total number n of nodes in the network is a global parameter, the progress of the independent stop protocol can be described by at least two parameters, e.g., the number k of informed nodes and the number $\ell \leq k$ of active ones.

One of the examples of the independent stop protocol is the median counter algorithm [KSSV00]. This is a $O(\ln \ln n)$-parameter process, so that the precise analysis of such protocol is challenging. Instead of median counter algorithm, we will consider the following process which are simpler and for which we have the strong intuition about the rumor spreading time.

Pull protocol with C push calls: In this protocol, when an uninformed node learns the rumor, it becomes active for the next C rounds and makes one push

[2] Note that technically, the protocol with the transition time considered in Section 3.3 of Chapter IV is also multi-state. However, all informed nodes switch from the active state to the passive one simultaneously, so that these two states cannot be occurred in the same round, and the protocol can be analyzed as homogeneous.

call per round. After C rounds, it becomes passive. Note that in addition to the number k of informed nodes, we have to keep in memory how many nodes have been informed at each of the C last rounds. Hence, the pull protocol with C push calls is $C+1$-parameter protocol.

Random stop protocol: In this variation of the basic push-pull process, each round each informed node decides independently at random with probability q to stop making push calls. The success probability that one node gets the rumor depends not only on the number k of informed nodes, but also on the number ℓ of active ones. Thus, this is an example of 3-parameter process.

Note that from the outer point of view, the protocols described above are similar to one from Section 1: they are all the variations of the push-pull protocols when the number ℓ of push calls in the round evolves almost independently from the number k of informed nodes. By this reason, most of the technical statements are already proved in Section 1. The runtimes of independent stop protocols are also stochastically bounded between the runtime of the basic pull protocol and the push-pull one, that makes the analysis simpler. Since both pull and push-pull protocols performs the double exponential shrinking regime, the independent stop protocol also performs the double exponential shrinking with rate 2. In addition, we can omit the covariance numbers routine arguing that we already proved the corresponding conditions for the both basic protocols.

Observation V.4. *Consider the symmetric independent stop protocol in which each uninformed node makes one pull call per round. For any $f \in]0,1[$ we have*

$$\mathbb{E}[T([nf],n)] = \log_2 \ln n \pm O(1).$$

2.1 One Push Call per Node

We start the analysis of multi-state protocols by the *pull protocol with one push call* – the simplest protocol that solves the stoppage problem and does not require any sharing of the rumor's "age" among informed nodes. The protocol is a variation of the basic synchronous push-pull protocol: any node that learns a rumor makes one push call in the next round and then becomes inactive. Informed nodes reply all received pull requests even if they do not make push calls.

Note that since $n/2$ nodes are informed, the number of push calls in current round does not exceed the number of pull calls in previous round, so the push calls do not spam the network.

Since the pull protocol with one push call is not memoryless, the challenging part of the analysis is the exponential growth.

Round Targets and Failure Probabilities

Consider an uniformed node in a round of the pull protocol with 1 push call starting with $k < n/2$ informed nodes that make $\ell \le k$ calls in current round. Similarly to the protocol from Section 1, we have $p_{k\ell} \ge \frac{\ell}{n} + \frac{k}{n} - \frac{(\ell+k)^2}{2n^2}$.

Denote by $X(k,\ell)$ the number of newly informed nodes in a round started with k informed nodes that make ℓ push calls. The definition of the target value coincides with (V.1) and the arguments used for Lemma V.2 also hold.

Lemma V.5. *Let* $X_0 := k + \ell - \frac{3}{2} \cdot \frac{(\ell+k)^2}{n} - A(k+\ell)^B$ *for some* $A > 0$ *and* $B \in]1/2, 1[$. *There exists* $q(k,\ell) = O(1) \cdot (k+\ell)^{-2B+1}$ *such that*

$$\mathbb{P}[X(k,\ell) \le X_0(k,\ell)] \le q(k,\ell).$$

The Phase Calculus

Similarly to Section 1, the progress of the pull protocol with one push call is described by two parameters, the number k of informed nodes and the number ℓ of push calls. Observe that the number of push calls in the next round is equal to the number of newly informed nodes in the current round. Thus, we define recursively

$$\begin{aligned} k_0 &:= 1, \quad k_{j+1} := k_j + X_0(k_j, \ell_j); \\ \ell_0 &:= 1, \quad \ell_{j+1} := X_0(k_j, \ell_j). \end{aligned}$$

This seems to be the typical structure of phase targets for the multi-parametric process: the recurrent sequences of round targets for k and ℓ are entangled with each other. Again, we employ the fact that the sequence above is almost linear. Denote by $\mathbf{x}_j := \binom{k_j}{\ell_j}$ the vector of the parameters and let $M := \left(\begin{smallmatrix} 2 & 1 \\ 1 & 1 \end{smallmatrix}\right)$, $\mathbf{u} = \binom{3/2}{3/2}$, and $\mathbf{v} = \binom{A}{A}$. Then, the recursive expression above can be written as

$$\mathbf{x}_{j+1} := M\mathbf{x}_j - \tfrac{1}{n}\mathbf{u}\|\mathbf{x}_j\|^2 - \mathbf{v}\|\mathbf{x}_j\|^B,$$

where $\|\mathbf{x}_j\| := k_j + \ell_j$ denotes the Manhattan norm. By the elementary computation we obtain that M has two real eigenvalues $\frac{3\pm\sqrt{5}}{2}$ and that the corresponding eigenvectors both has positive coordinates. Therefore, Lemma V.3 transforms into the following one which can be proved using exactly the same arguments.

Lemma V.6. *Let* $\lambda := \frac{3+\sqrt{5}}{2}$. *After possibly lowering* A *from* (V.1), *there exists* $J = \log_\lambda n + O(1)$ *such that for any* $j < J$ *we have* $k_j = \Theta(1)\lambda^j$ *and* $\ell_j = \Theta(1)\lambda^j$.

Again, the round targets cut the interval $[1, fn]$ for some $f \in]0,1[$ into phases. We say that the process achieves the phase j if $\mathbf{x} \geqslant \mathbf{x}_j$, i.e., both $k \ge k_j$ and $\ell \ge \ell_j$.

The main difference between the process from Section 1 and the pull protocol with one push call is that in last one the values of parameters may decrease[3]. If, for example, nobody is informed at the end of current round, then there will be no push calls during the next one. Thus, we need to prove one more statement.

Lemma V.7. *Suppose that the process is in phase j. The expected number of rounds before it proceeds to phase $j+1$ or further is $1 + O(1) \cdot q(\mathbf{x}_j)$.*

Proof. With probability at least $1 - q(\mathbf{x}_j)$, the process proceeds to the phase $j+1$ or further in one round. If this did not happened, then we consider only pull calls until the process reaches at least phase $j+1$. Since pull protocol satisfies the exponential growth conditions, the expected number of rounds until at least $3k_j$ nodes are informed is $O(1)$. Since $k_j \geq \ell_j$, we have $3k_j \geq 2k_j + \ell_j$, so that $3k_j$ belongs to the phase $j+1$. Then we wait until at lest $2k_j \geq k_j + \ell_j$ nodes are informed in one round. It is easy to see that the expected number of rounds to wait is $O(1)$. □

Lemma V.5, V.6 and V.7 together with Observation V.4 imply the following upper bound for the rumor spreading time.

Theorem V.8. *The expected rumor spreading time for the pull protocol with one push call in the complete graph on n vertices is at most $\log_\lambda n + \log_2 \ln n + O(1)$, where $\lambda = \frac{3+\sqrt{5}}{2}$.*

Proof. Since by Lemma V.5 and V.6, we have $\sum_{j=0}^{J-1} q(\mathbf{x}) = O(1)$, Lemma V.7 implies that $\mathbb{E}[T(1, k_J)] = J + O(1)$. The rest of the claim follows immediately from Observation V.4. □

2.2 Pull Protocol with $C \geq 1$ Push Calls:

Let us now consider the general case of the pull protocol with $C \geq 1$ push calls. As it was mentioned before, this is a $C+1$ parameter process. Thus, we choose the following parameters to describe the progress of the protocol: the number k of informed nodes and $\ell^{(1)}, \ell^{(2)}, \ldots, \ell^{(C)}$, where $\ell^{(i)}$ is the number of nodes that were informed during round $t - i$ where t is current round. Thus, the number of calls in current round is $\ell := \ell^{(1)} + \ldots + \ell^{(C)}$. Therefore, the number $X(k, \ell)$ of newly informed nodes depends only on k and ℓ exactly in the same way as for the pull protocol with one pull call. Therefore, its target value X_0 coincides with (V.1):

$$X_0(k, \sum_{i=1}^{C} \ell^{(i)}) := \sum_{i=1}^{C} \ell^{(i)} + k - \tfrac{3}{2} \cdot \tfrac{(k+\sum_{i=1}^{C} \ell^{(i)})^2}{n} - A(k + \sum_{i=1}^{C} \ell^{(i)})^B.$$

[3]That is also true for all independent stop protocols considered in this section

The evolution of the parameters goes in the following way. The number k of informed nodes increases by X_0. The number $\ell^{(1)}$ of nodes that were informed during last round becomes equal to X_0. All other $\ell^{(i)}$ shrink on the right. Therefore, the sequence of round targets is the following.

$$\begin{aligned} k_0 &:= 1, \quad k_{j+1} := k_j + X_0(k_j, \sum_{i=1}^{C} \ell_j^{(i)}); \\ \ell_0^{(1)} &:= 1, \quad \ell_{j+1}^{(1)} := X_0(k_j, \sum_{i=1}^{C} \ell_j^{(i)}); \\ \ell_0^{(i)} &:= 0, \quad \ell_{j+1}^{(i)} := \ell_j^{(i-1)} \text{ for all } 2 \le i \le C. \end{aligned}$$

Denote by $\mathbf{x}_j$ the vector-column $(k_j, \ell_j^{(1)}, \dots, \ell^{(C)})^T$ of the parameters. Then, using the notation above, we can write $\mathbf{x}_{j+1} := M_C \mathbf{x_j} - \frac{1}{n}\mathbf{u}\|\mathbf{x_j}\|^2 - \mathbf{v}\|\mathbf{x_j}\|^B$, where

$$M_C := \begin{pmatrix} 2 & 1 & 1 & \dots & 1 & 1 \\ 1 & 1 & 1 & \dots & 1 & 1 \\ 0 & 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & 1 & \dots & 0 & 0 \\ 0 & 0 & 0 & \vdots & 0 & 0 \\ 0 & 0 & 0 & \dots & 1 & 0 \end{pmatrix}, \quad \mathbf{u} := \begin{pmatrix} 3/2 \\ 3/2 \\ \vdots \\ 3/2 \end{pmatrix}, \quad \mathbf{v} := \begin{pmatrix} A \\ A \\ \vdots \\ A \end{pmatrix}.$$

By induction on C we obtain that the characteristic polynomial of M_C is $P_C(\lambda) := -\lambda^{C+1} + 3\lambda^C - 1$. The proposition similar to Lemma V.3 and V.6 also holds for the pull protocol with C push calls. Therefore, there exists $J = \log_{\lambda_C} n + O(1)$ such that $k_J = \Theta(n)$.

Like in the pull protocol with one push call, the values of the parameters might decrease. Proving the statement similar to Lemma V.7 is the most challenging part. To simplify the technical details we slightly relax the claim of the following lemma.

Lemma V.9. *Suppose that the process is in phase j. The expected number of rounds before it proceeds to phase $j+1$ or further is $1 + O(C) \cdot q(\mathbf{x}_j)$.*

Sketch of the proof. With probability at least $1 - q(\mathbf{x}_j)$, the process proceeds to the phase $j+1$ or further in one round. If this did not happened, then we consider only pull calls until the process reaches at least phase $j+1$. Since the pull protocol satisfies the exponential growth conditions, the expected number of rounds until at least $(C+1)k_j$ nodes are informed is $O(C)$.

It is easy to see that $\ell_{j+1}^{(1)} \leq 2k_j$ and $\ell_{j+1}^{(i)} \leq k_j$ for all $2 \leq i \leq C$. Therefore, to proceed into phase $j+1$ or further, it suffices that during next $C-1$ rounds there are at least k_j newly informed nodes per round, and, $2k_j$ newly informed nodes during the Cth round. From Chebyshev's inequality it follows that with probability at least $1 - O(1/C)$, at least k_j uninformed nodes make successful pull calls in a round starting with at least $(C+1)k_j$ informed nodes. Consequently, with the probability $\Theta(1)$, the process will proceed to phase $j+1$ or further after C rounds.

Thus, the expected number of rounds until the process proceeds to the phase $j+1$ or further is at most $1 + O(C) \cdot q(\mathbf{x}_j)$. □

The bound for the rumor spreading time immediately follows, the proof of theorem coincides with the proof of Theorem V.8.

Theorem V.10. *The expected rumor spreading time for the pull protocol with C push calls in the complete graph on n vertices is at most $\log_{\lambda_C} n + \log_2 \ln n + O(C)$, where λ_C is the largest root of $\lambda^{C+1} - 3\lambda^C + 1$.*

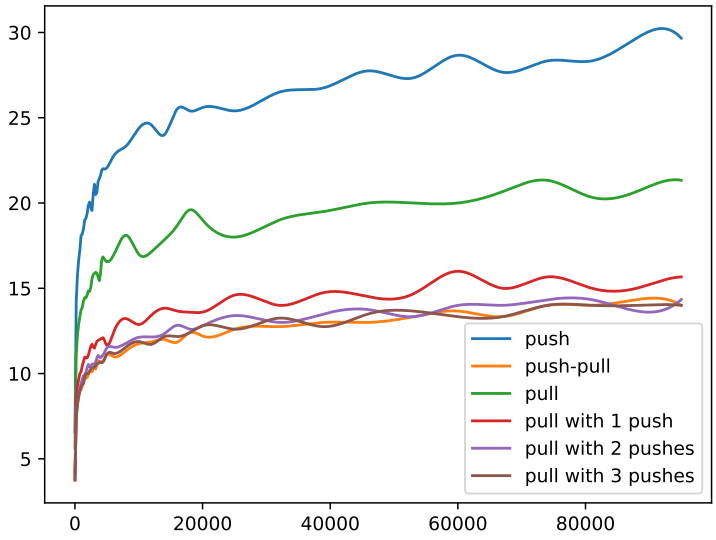

Figure 7: The rumor spreading time versus size of the network. We used the cubic spline interpolation to get the smooth curve.

It is easy to see that $\lambda_C \geq 3 - \frac{2}{3^C}$. Since the pull protocol with C push calls cannot be faster than the basic push-pull protocol, we have $\lambda_C \leq 3$. Therefore,

λ_C converges to 3 exponentially fast as C increases. We illustrate this effect on Figure 7 by plotting the rumor spreading time of the pull protocol with C push calls for $C = 1, 2, 3$ in the network with up to 100K nodes. For the comparison, we also plot the spreading time for the basic push, pull and push-pull protocols. We see that the allowance of even one push call per node in addition to the pull calls significantly decrease the rumor spreading time. If we allow 3 push calls per node, we cannot see any difference between the performances of the pull protocols with 3 push calls and the basic push-pull protocol when the network contains at most 100K nodes.

2.3 Random Stop Decision

In the pull protocol with C push calls, the active node should keep in memory how long they are informed, so that instead of three states, any node can be in $C + 2$ ones: uninformed, informed and passive, and C informed and active states. Conversely, the *random stop protocol* is a true three-state process in which nodes require only one bit of memory to keep their current state.

The random stop protocol acts similarly to the previous ones from the outer view. But instead of counting rounds, each active node each round decides with probability q independently from other nodes to become inactive. Thus, any round consists of two steps: first, nodes make calls according to their states, then, the active nodes decide to become inactive according to their random coins.

Note that the behavior of this protocol depends on the choice when nodes becomes informed: directly after learning the rumor or since the next round, i.e., whether newly informed nodes toss a coin immediately. If yes, then a newly informed node can become inactive bypassing the active state. We call this protocol *type I* random stop protocol. If not, then any newly informed node makes at least one push call. This is a *type II* protocol.

As above, we will study the random stop protocol based on the classic push-pull protocol[4]: active nodes make one random push call per round and uninformed ones make one pull call per round. Therefore, the rumor spreading time for the push-pull random stop protocol is stochastically bounded between the spreading time for the basic pull and push-pull protocols. In particular, the random stop push pull protocol also performs the double exponential shrinking phase with rate 2.

As before, we denote by n the number of nodes in the network. Consider one round starting with k informed nodes and suppose that ℓ of them are active. First,

[4] The random stop push protocol is much less reasonable than the push-pull. It is easy to see that if $q = \omega\left(\frac{1}{\ln n}\right)$, then the random stop push protocol fails to inform all nodes with high probability. Indeed, it follows from the coupon collector reduction that we need at least $\Theta(n \ln n)$ calls to inform all nodes. On the other hand, the expected number of the calls in the random stop push protocol is at most $n/q = o(n \ln n)$, that makes a contradiction.

we observe that the previous statements concerning the number of newly informed nodes (The expression for the success probability $p_{k\ell}$, the expression (V.1) for the target value $X_0(k,\ell)$, and Lemma V.2) hold for both types of the random stop protocol. Therefore, the number of newly informed nodes achieves its target value $X_0(k,\ell)$ with probability $q_1(k,\ell) = O(1)\cdot(k+\ell)^{1-2B}$ for some $B \in]1/2, 1[$. More precisely, there exists $A_1 > 0$ such that

$$\mathbb{P}[X \le k+\ell - \tfrac{3}{2n}(k+\ell)^2 - A_1(k+\ell)^B] \le q_1(k,\ell), \tag{V.5}$$

In addition, the proof of Lemma V.7 is also applicable for both types of the random stop protocol. Therefore, the expected number of rounds before the process proceeds to the next phase or further is at most $1 + O(1)\cdot q(\mathbf{x}_j)$ for the round target sequence $\mathbf{x}_j$ defined below.

Type I Protocol

The second step when nodes decide to become inactive depends on the type of the protocol. For the random stop protocol of type I, ℓ active nodes toss coins together with all newly informed ones. Let Y denote the number of active nodes at the end of the round and suppose that $X \ge X_0(k,\ell)$. Since the probability that active node becomes inactive is q, we have $\mathbb{E}[Y] \ge (1-q)(\ell + X_0(k,\ell))$. Since all nodes make decisions independently, $\mathrm{Var}[Y] \le \mathbb{E}[Y]$. Therefore, applying the Chebyshev's inequality, we obtain that there exists $A_2 > 0$ such that

$$\mathbb{P}[Y \le (1-q)(\ell+X_0(k,\ell)) - A_2\ell^B] \le q_2(k,\ell), \quad \text{where } q_2 = O(1)\cdot(k+\ell)^{1-2B}. \tag{V.6}$$

The expressions (V.5) and (V.6) imply the following construction of the round target sequence.

$$\begin{aligned} k_0 &:= 1, \quad k_{j+1} := k_j + X_0(k_j,\ell_j);\\ \ell_0 &:= 1, \quad \ell_{j+1} := (1-q)(\ell_j + X_0(k_j,\ell_j)) - A_2\ell^B. \end{aligned}$$

Let the parameter vector $\mathbf{x}_j$ be $\binom{k_j}{\ell_j}$. Then, using the previous notation, we can write $\mathbf{x_0} = \binom{1}{1}$ and $\mathbf{x}_{j+1} := M\mathbf{x_j} - \frac{1}{n}\mathbf{u}\|\mathbf{x_j}\|^2 - \mathbf{v}\|\mathbf{x_j}\|^B$, where

$$M := \begin{pmatrix} 2 & 1 \\ 1-q & 2-2q \end{pmatrix}, \quad \mathbf{u} := \begin{pmatrix} \frac{3}{2} \\ (1-q)\frac{3}{2} \end{pmatrix}, \quad \mathbf{v} := \begin{pmatrix} A_1 \\ (1-q)A_1 + A_2 \end{pmatrix}.$$

Matrix M has eigenvalues $\lambda = 2 - q \pm \sqrt{q^2-q+1}$. Then, similarly to the previously regarded processes, we conclude the rumor spreading time.

Theorem V.11. *Let $\lambda = 2-q+\sqrt{q^2-q+1}$. Then the expected rumor spreading time for the type I push-pull protocol with stop probability q in the complete graph on n vertices is at most*

$$\log_\lambda n + \log_2 \ln n + O(1).$$

Although, unlike the pull protocol with C push calls, the parameters achieve their round targets independently, we employ the fact that $k \geq \ell$, so that no matter which is parameters achieves first its Jth round target, we have $k \geq \ell_J = \Theta(n)$.

Type II Protocol:

The analysis for the protocol of type II is similar to the previous one. The only difference is that during the second step only ℓ nodes that were active at the beginning of the round toss coins. Therefore, the expected number of active nodes at the end of the round is at least $(1-q)\ell + X_0(k,\ell)$ with good probability. Thus, (V.6) transforms into

$$\mathbb{P}[Y \leq (1-q)\ell + X_0(k,\ell) - A_2\ell^B] \leq q_2(k,\ell), \quad \text{where } q_2 = O(1)\cdot(k+\ell)^{1-2B}.$$

The expression above implies the following definition of the round target sequence.

$$\begin{aligned} k_0 &:= 1, \quad k_{j+1} := k_j + X_0(k_j,\ell_j); \\ \ell_0 &:= 1, \quad \ell_{j+1} := (1-q)\ell_j + X_0(k_j,\ell_j) - A_2\ell^B. \end{aligned}$$

We denote by $\mathbf{x}_j$ the parameter vector $\binom{k_j}{\ell_j}$. With the usual notation, we rewrite the expression above as $\mathbf{x}_0 := \binom{1}{1}$ and $\mathbf{x}_{j+1} := M\mathbf{x_j} - \frac{1}{n}\mathbf{u}\|\mathbf{x_j}\|^2 - \mathbf{v}\|\mathbf{x_j}\|^B$, where

$$M := \begin{pmatrix} 2 & 1 \\ 1 & 2-q \end{pmatrix}, \quad \mathbf{u} := \begin{pmatrix} 3/2 \\ 3/2 \end{pmatrix}, \quad \mathbf{v} := \begin{pmatrix} A_1 \\ A_1 + A_2 \end{pmatrix}.$$

The eigenvalues of M are $\lambda = 2 - \frac{q}{2} - \frac{1}{2}\sqrt{q^2+4}$. Then, we conclude the rumor spreading time.

Theorem V.12. *Let $\lambda = 2 - \frac{q}{2} - \frac{1}{2}\sqrt{q^2+4}$. Then the expected rumor spreading time for the type II push-pull protocol with stop probability q in the complete graph on n vertices is at most*

$$\log_\lambda n + \log_2 \ln n + O(1).$$

Remark V.13. Clearly, if $q = 0$, then $\lambda = 3$ for both protocols. In this case nodes never become inactive, so that both protocols are equivalent to the basic push-pull protocol.

For the type I protocol, $q = 1$ implies $\lambda = 2$. In this case all informed nodes are inactive since they are informed, that is equivalent to the basic pull protocol.

For the type II protocol, $q = 1$ implies $\lambda = \frac{3+\sqrt{5}}{2}$. In this case any node that learns the rumor stays active for one round and makes exactly one push call. This is equivalent to the pull protocol with 1 push call.

3 Multiparametric exponential growth

The analysis of the processes above yields the intuition that the phase method is suitable for the multi-parametric exponential growth. However, it is challenging to provide a general, easy-to-use construction that, similarly to the exponential growth and shrinking conditions from previous chapters, relate the micro-parameters such as the probabilities p_k of a node to become informed with the macro-parameters of the process such as the number of informed nodes and, consequently, the rumor spreading time.

Comparing the processes above we see two main problems that prevent from obtaining such a general method. First, the parameters in different processes may have different nature. Although for all processes considered above, ℓ denotes the number of push calls in current round, it behaves differently in different processes. In the pull protocol with one push call it is equal to the number of nodes informed during the previous round. In the random stop protocol as well as in the protocol from Section 1, this number ℓ changes at the end of the round and has an additional source of the randomness independent of the number k of informed nodes. The second problem is that the parameters in the multi-parametric processes might decrease, as we have seen for the independent stop process. Thus, it might be challenging to show that the expected number of rounds until the process moves from phase j to phase $j+1$ or further is at most $1+q(\mathbf{x}_j)$.

By this reason, we have the only choice to remove the atomic success probabilities for single nodes from the phase analysis. We build the multi-parametric exponential growth conditions based on the macro-parameters described by $\mathbf{x}$. These conditions should be understood as a plan with the list of statements that, once proved, yield together an estimate for the spreading time.

To formulate the multi-parametric exponential growth conditions we will use the following notation. Let n be the number of nodes in the network. Suppose that the progress of the rumor spreading process is described by m parameters forming the vector $\mathbf{x}$ such that the number k of informed nodes equals $\mathbf{K}\mathbf{x}$, for some constant vector $\mathbf{K}$. By $\|\cdot\|$ we denote the Manhattan norm. Finally, let $\mathbf{y}(\mathbf{x})$ be the values of the parameters at the end of the round staring from $\mathbf{x}$. The multi-parametric exponential growth conditions contain in the following four constructions.

Round target operator: Let M be a $m \times m$ real matrix. Let $\mathbf{u}$, $\mathbf{v}$ be two vectors with non-negative coordinates. Let $1/2 < B < 1$. Then we define *round target operator* as

$$E_0(\mathbf{x}) := M\mathbf{x} - \tfrac{1}{n}\mathbf{u}\|\mathbf{x}\|^2 - \mathbf{v}\|\mathbf{x}\|^2.$$

Error condition: We say that the round target operator E_0 satisfies the *error conditions*, if there exists a function $q(\mathbf{x}) = O(1) \cdot \|\mathbf{x}\|^{1-2B}$ and a vector $\mathbf{f}$ with positive coordinates such that for all $\mathbf{x} \leqslant n\mathbf{f}$ element-wise, we have

$$\mathbb{P}[\mathbf{y}(\mathbf{x}) \geqslant E(\mathbf{x}) \text{ elementwise}] \geq 1 - q(\mathbf{x}).$$

Exponential-target conditions: Let matrix M has m eigenvalues and the largest one is real. Denote it by λ. We define the sequence $\mathbf{x}_j$ of the round targets recursively as $\mathbf{x}_{j+1} := E_0(\mathbf{x}_j)$, where $\mathbf{x}_0$ consists of the initial values of the parameters. We say that the round target operator E_0 satisfies the *exponential target conditions* if there exists $J = \log_\lambda n + O(1)$ such that (i) the sequence $\mathbf{x}_j$ is increasing for $0 \leq j \leq J$ and (ii) $\mathbf{K} \cdot \mathbf{x}_J = \Theta(n)$.

Transition condition: Suppose that the round target operator E_0 satisfies the exponential-target conditions. Suppose that the protocol is in phase j, i.e., $\mathbf{x} \geqslant \mathbf{x}_j$ element-wise. We say that the *transition condition* is satisfied if for all $j \leq J$, the expected number of rounds until the process proceeds to phase j or further is $1 + O(1) \cdot q(\mathbf{x}_j)$.

Remark.

(i) The conditions above are equivalent to the main statements in Section 2.1 of Chapter III. The error condition corresponds to Lemma III.9. The exponential-target conditions correspond to Lemma III.10. The transition condition corresponds to Lemma III.11. Lemma III.12 follows from the error condition together with the transition conditions.

(ii) In the exponential-target conditions we require that $\mathbf{K} \cdot \mathbf{x}_J = \Theta(n)$. To prove that this is satisfied, one need to compute $\mathbf{K} \cdot M^j \mathbf{x}_0$ and verify explicitly that for the given initial value $\mathbf{x}_0$, the expression for the number of informed nodes contains the λ^j term.

Theorem V.14. *Consider a multi-parametric rumor spreading protocol. Suppose that the round target operator satisfies the error condition, the exponential-target conditions, and the transition condition. Then there exists $f \in]0, 1[$ such that the expected spreading time $\mathbb{E}[T(1, fn)] = \log_\lambda n + O(1)$.*

Proof. From the exponential-target conditions, it follows that the sequence of round targets $\mathbf{x}_j$ cuts the progress of the process until at least $\Theta(n)$ nodes are informed into $J = \log_\lambda n + O(1)$ phases. From the transition condition it follows that $\mathbb{E}[T(1, \Theta(n))] = J + O(1) \cdot \sum_{j=0}^{J-1} q(\mathbf{x}_j)$.

Finally, the error condition together with exponential-target conditions imply that $q(\mathbf{x}_j)$ form a decreasing geometric series. Consequently, $\sum_{j=0}^{J-1} q(\mathbf{x}_j) = O(1)$, that finishes the proof. □

The multi-parametric exponential growth conditions for the lower bound can be easily obtained from the upper bound conditions similarly to the exponential growth conditions for the homogeneous rumor spreading in Chapter III. However, we do not know how to generalize the proof of Theorem III.14 to the multi-parametric rumor spreading, so we do not provide them.

Chapter VI

Summary

Contents

1 Classic Protocols **71**
1.1 Push Protocol 72
1.2 Pull Protocol 73
1.3 Push-Pull Protocol 75
2 Robustness, Multiple Calls, and Dynamic Graphs **76**
2.1 Transmission Failures 77
2.2 Multiple Calls 80
2.3 Dynamic Graphs 84
3 Limited Incoming Calls Capacity **92**
3.1 Single Incoming Call Push-Pull Protocol 93
3.2 Single Incoming Call Pull-Only Protocol 95
3.3 Push-Pull Protocol with Transition Time 96

1 Outlook

We gave a general, easy-to-use method to analyze homogeneous rumor spreading processes on complete networks. Such processes are important in many applications, among others, due to the use of random peer sampling services in many distributed systems. Such processes also correspond to the fully mixed population model in mathematical epidemiology.

The two main strengths of our method are (i) that it requires only understanding the probability that an uninformed node becomes informed and a mild covariance condition, whereas a deeper knowledge of the process is not required (as opposed to all previous works on this topic), and (ii) that it can determine the expected rumor spreading time precise apart from additive constants (the only such analysis for one particular protocol is the predecessor work [DK14]). The key to our results is distilling the right growth and shrinking conditions, which allow us to describe essentially all previously regarded homogeneous processes, and to show, based on these conditions, that the usually present mild deviations from a perfect exponential growth or shrinking in total cost only a constant number of rounds.

From the viewpoint of rumor spreading, this work leaves open two desires, namely overcoming the restrictions to complete networks and to processes without memory. For the former, random graphs might be a good first object of investigation as there similar rumor spreading times have been observed as in complete networks. Concerning the memory issue, it has been observed that already a mild use of memory (not calling the same neighbor twice in a row) can make a substantial difference, so potentially this is an interesting first object for further research.

From a broader perspective, this work shows that the traditional approach to randomized processes of splitting the analysis in several phases and then trying to understand each phase with uniform arguments might not be the ideal way to capture the nature of processes with a behavior changing continuously over time. While we demonstrated that the more careful round-target approach is better suited for homogeneous rumor spreading processes, one can speculate if similar ideas are profitable for other randomized algorithms or processes regarded in computer science.

2 Open Problems

We showed the proof of concept that the phase analysis is a reasonable approach for the multi-parametric rumor spreading and, especially, for the multi-state processes. However, we leave the following problems open for the further studies. Solving them is necessary to transform our intuition into a general method, similar as we proposed for the single parameter homogeneous rumor spreading.

Improving the exponential growth conditions: For the analysis of the arbitrary multi-parametric process, the transition condition seems to be the most hard to prove. It is also might be interesting to have some easier-to-use version of the exponential growth conditions. Our intuition is that it is possible to find

such conditions for the multi-state protocol when the parameters are the numbers of nodes belonging to the corresponding states. Also it is worth to provide the general multi-parametric lower bound conditions.

Multi-parametric shrinking: For all protocols we discussed in this chapter, the double exponential shrinking behavior trivially follows from the fact that the independent stop process is "sandwiched" between basic pull and push-pull protocols. However, for the general method we need the criteria of the behavior of the process since $\Theta(n)$ nodes are informed until the last node learns the rumor. We suppose that similarly to the single parameter homogeneous rumor spreading, there are two possible regimes: the exponential shrinking and the double exponential shrinking. Since the single parameter exponential growth conditions are almost linear, it is probably that the construction for the exponential shrinking will remind the exponential growth: some parameter vector $\mathbf{x}$ is multiplied each round by almost a constant matrix M with eigenvalues less than 1 by the absolute value. At the same time, the double exponential shrinking conditions are not linear. The possible construction can be based on the analysis of the behavior of some dimensionless quantity similar to $\frac{\|\mathbf{x}\|}{n}$ which is supposed to be powered each round.

Connection between regimes: Typically, it is hard to prove that the shrinking regime starts directly the next round after the growth regime ends. For the single parameter homogeneous rumor spreading, we used Lemma II.17 and II.18 claiming that the for any $0 < f < g < 1$, the transition from fn to gn informed nodes takes $O(1)$ rounds with high probability. Using them, it suffices to prove that the growth and shrinking conditions are satisfied only within intervals $[1, fn]$ and $[(1-g)n, n]$ of informed nodes correspondingly, where f and g might be arbitrarily small. Proving the similar statements for the multi-parametric rumor spreading seems to be challenging but important for the simplification of the method, because now we need to glue together not only the number of informed nodes, but also the value of all remaining parameters.

Appendix A

Probabilistic notions

Geometric Distribution and Stochastic Domination

Definition. We say that a random integer variable G has a *geometric distribution* with success probability p and write $G \sim \mathrm{Geom}(p)$ if $\mathbb{P}[G = k] = p(1-p)^k$ for any $k \geq 0$.

The geometric distribution corresponds the number of failed Bernoulli trials until the first success. Recall that if $G \sim \mathrm{Geom}(p)$, then we have $\mathbb{E}[G] = \frac{1-p}{p}$ and $\mathrm{Var}[G] = \frac{1-p}{p^2}$.

Another important concept is the stochastic domination. Informally, a random variable X dominates a random variable Y if X's distribution is "to the right" of the Y's distribution.

Definition. Let a pair of random variables X, Y be given. We say that X *stochastically dominates* Y, and write $Y \preceq X$, if $\mathbb{P}[X \geq x] \geq \mathbb{P}[Y \geq x]$ for all x.

The stochastic domination satisfies the following elementary properties.

- if $X \preceq Y$ and $Z \preceq T$, then $X + Y \preceq Z + T$.
- if $X \preceq Y$ then $\mathbb{E}X \leq \mathbb{E}Y$.

Lemma A.1. *Let $G_1, \ldots, G_n$ be independent random variables with $G_i \sim \mathrm{Geom}(1-q_i)$. Then $\sum_{i=1}^n G_i$ is stochastically dominated by a random variable G with $G \sim \mathrm{Geom}(1 - \sum_{i=1}^n q_i)$*

Proof. Let G_1, G_2 be independent geometrically distributed random variables with success probabilities $1 - q_1$ and $1 - q_2$, respectively. By law of total probability,

we compute for all $t \geq 0$,

$$\mathbb{P}[G_1 + G_2 \geq t] = \left(\sum_{k=0}^{t-1} \mathbb{P}[G_1 = k]\mathbb{P}[G_2 \geq t-k]\right) + \mathbb{P}[G_1 \geq t] =$$
$$= \left(\sum_{k=0}^{t-1}(1-q_1)q_1^k q_2^{t-k}\right) + q_1^k \leq \sum_{k=0}^{t} q_1^k q_2^{t-k} \leq \sum_{k=0}^{t}\binom{t}{k} q_1^k q_2^{t-k} = (q_1+q_2)^t.$$

Hence, $G_1 + G_2 \preceq \mathrm{Geom}(1-(q_1+q_2))$. By successive application of this fact, we obtain $\sum_{i=1}^n G_i \preceq \mathrm{Geom}(1-\sum_{i=1}^n q_i)$. □

The following lemma contains a high probability bound for the sum of geometrically distributed variables in the case when $\sum_i q_i = O(1)$, but not necessarily less than 1.

Lemma A.2. *Let $\varepsilon, \delta \in]0,1[$ and $s > 0$. Let $q_j := \min\{1-\varepsilon, s\delta^j\}$, for any j. Let G be stochastically dominated by $\sum_{j=0}^{J-1} G_j$, where $G_j \sim \mathrm{Geom}(1-q_j)$. Then there exist constant $A, \alpha > 0$ such that for any integer $r > 0$ we have $\mathbb{P}[G > r] \leq Ae^{-\alpha r}$.*

Proof. Let j_0 is the smallest such that $\sum_{j \geq j_0} q_j < 1 - \varepsilon$. By construction, $j_0 = O(1)$. By Lemma A.1, $\sum_{j=j_0}^{J-1} G_j$ is stochastically dominated by a random variable with distribution $\mathrm{Geom}(\varepsilon)$. Therefore, for any integer $r > 0$ we have

$$\mathbb{P}\left[\sum_{j=j_0}^{J-1} G_j > \tfrac{r}{j_0+1}\right] \leq (1-\varepsilon)^{r/(j_0+1)}.$$

Similarly, for any $j < j_0$ we have $\mathbb{P}[G_j > \frac{r}{j_0+1}] \leq (1-\varepsilon)^{r/(j_0+1)}$. We conclude,

$$\mathbb{P}[G > r] \leq (j_0+1) \cdot (1-\varepsilon)^{r/(j_0+1)}.$$

□

Variance. Chebyshev's and Cantelli's Inequalities

We recall that the *variance* of a discrete random variable X is $\mathrm{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$. By definition it is a measure of how well X is concentrated around its mean.

We recall that the *covariance* of two discrete random variables X and Y is $\mathrm{Cov}[X,Y] = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]$. It is a measure of how strong X and Y depends from each other. If X, Y are independent, then $\mathrm{Cov}[X,Y] = 0$.

In this work we often need to bound a variance of a sum of indicator random variables. We provide a simple method get the bound if we know the covariance between indicator random variables.

Lemma A.3. *Let a random variables $X = \sum_{i=1}^{n} X_i$, where X_i are indicator random variables. Suppose, for any $i \neq j$ we have $\mathrm{Cov}[X_i, X_j] \leq c$ for some constant c. Then $\mathrm{Var}[X] \leq \mathbb{E}[X] + cn^2$.*

Proof. Since X_i is a binary random variable, $\mathrm{Var}[X_i] \leq \mathbb{E}[X_i]$. Therefore,

$$\mathrm{Var}[X] \leq \sum_{i=1}^{n} \mathrm{Var}[X_i] + \sum_{i \neq j} \mathrm{Cov}[X_i, X_j] \leq \mathbb{E}[X] + cn^2.$$

□

The two following inequalities gives the bounds for the "tail" probabilities for any random variable X.

Lemma A.4 (Chebyshev's inequality)**.** *For all $\lambda > 0$,*

$$\mathbb{P}\left[|X - \mathbb{E}[X]|] \geq \lambda\sqrt{\mathrm{Var}[X]}\right] \leq \tfrac{1}{\lambda^2}.$$

There is a one-sided version of the Chebyshev inequality attributed to Cantelli, replacing $\frac{1}{\lambda^2}$ by $\frac{1}{\lambda^2+1}$.

Lemma A.5 (Cantelli's inequality)**.** *For all $\lambda > 0$,*

$$\mathbb{P}\left[X - \mathbb{E}[X] \geq \lambda\sqrt{\mathrm{Var}[X]}\right] \leq \tfrac{1}{1+\lambda^2}.$$

We remark that Cantelli's inequality gives the bound which is less than one for any positive λ.

Appendix B

First Order Bounds

In this appendix we collected all simple estimates that we used in Chapter IV and VI while proving that certain protocols satisfy the corresponding shrinking or growth conditions.

Lemma B.1. *For any $n > 0$ we have $\frac{1}{e} - \frac{1}{en} \leq \left(1 - \frac{1}{n}\right)^n \leq \frac{1}{e}$.*

Proof. Let $a_n := 1 - \frac{1}{n}$. It is easy to see that a_n increases for $n \geq 1$. Since $a_n \to \frac{1}{e}$ as $n \to \infty$, we have $a_n \leq \frac{1}{e}$.

Observe now that $a'_n n := \left(1 - \frac{1}{n}\right)^{n-1}$ decreases for $n > 1$. Since $a'_n \to \frac{1}{e}$ as $n \to \infty$, we have $a'_n \geq \frac{1}{e}$, that is equivalent to the first inequality of the claim. □

Lemma B.2. *For any $k < n$ we have $1 - \frac{k}{n} \leq \left(1 - \frac{1}{n}\right)^k \leq 1 - \frac{k}{n} + \frac{k^2}{2n^2}$.*

Proof. The first part of the claim is well-known as Bernoulli's inequality. The second part directly follows from the expansion of $\left(1 - \frac{1}{n}\right)^k$ and from the fact that $k/n < 1$. It suffices to observe that the coefficient at $\frac{1}{n^i}$ in obtained alternating sum does not exceed in absolute value $\frac{k^i}{i!}$.

$$\left(1 - \tfrac{1}{n}\right)^k = 1 - \tfrac{k}{n} + \tbinom{k}{2} \cdot \tfrac{1}{n^2} - \tbinom{k}{3} \cdot \tfrac{1}{n^3} + \dots$$

□

Corollary B.3. *Let $p > 0$. For any $k < n/p$ we have $1 - p\frac{k}{n} \leq \left(1 - \frac{p}{n}\right)^k \leq 1 - p\frac{k}{n} + \frac{p^2k^2}{2n^2}$.*

Proof. Follows directly from Lemma B.2, if we substitute n/p by n for some $p > 0$. □

Lemma B.4. *For any $0 \leq x < 1$ we have $\frac{1}{1-x} \geq 1 + x$.*
For any $0 \leq x \leq \frac{1}{2}$ we have $\frac{1}{1-x} \leq 1 + 2x$.

Proof. The first inequality is trivially equivalent to $1 \geq (1-x)(1+x) = 1-x^2$. The second one is equivalent to $1 \leq (1+2x)(1-x) = 1+x-2x^2$. Clearly, $x-2x^2$ is non-negative if and only if $x \in [0, 1/2]$. □

Combining the three lemmas above we obtain the following corollary.

Corollary B.5. *For any $1 < u < n$ we have $\frac{1}{e} \leq \left(1-\frac{1}{n}\right)^{n-u} \leq \frac{1}{e} + \frac{2u}{en}$.*

Proof. $\frac{1}{e} \leq \left(1-\frac{1}{n}\right)^{n-u}$ follows directly from Lemma B.1.

By Lemma B.2, we have $\left(1-\frac{1}{n}\right)^{u} \geq 1-\frac{u}{n}$. Then, from Lemma B.4 it follows that $\left(1-\frac{1}{n}\right)^{-u} \geq 1+\frac{2u}{n}$. Multiplying this expression by the statement of Lemma B.1, we obtain that $\left(1-\frac{1}{n}\right)^{n-u} \leq \frac{1}{e} + \frac{2u}{en}$. □

Similarly to Corollary B.6, we have the following.

Corollary B.6. *Let $p > 0$. For any $u < n/p$ we have $e^{-p} \leq \left(1-\frac{p}{n}\right)^{n-u} \leq e^{-p}\left(1+2p\frac{u}{n}\right)$.*

Proof. Follows directly from Corollary B.5, if we substitute n/p by n for some $p > 0$. □

Appendix C

Coupon Collector and Ball into Bins

In this appendix we recall two classical combinatoric problem that we used in the proofs.

Balls into Bins

The problem involves m balls and n bins. Each time, a single ball is placed into one of the bins. After all balls are placed, we count the load (i.e., the number of the balls inside) of each bin. The balls into bins problems asks for the maximum load on a single bin.

For the case $m = n$ that corresponds to a rumor spreading protocol in which each node makes exactly one call, the result is well known [Gon81].

Lemma C.1. *With probability $1-o(1)$, the maximum load of the bin in the instance of the problem with n balls and n bins is equal to $\frac{\ln n}{\ln \ln n} \cdot (1 + o(1))$.*

Coupon Collector

The coupon collector problem is formulated as follows. *A company issues coupons of n different types, each type having a certain probability of being issued. The coupon collector problem asks for the expected number of coupons that need to be gathered before a full collection is obtained.* The same problem often occurs in the analysis of the push rumor spreading protocol, e.g., in a star graph or in a complete graph since most of the nodes are informed.

Let C_n denotes the number of draws necessary to pick a full collection of coupon starting from the empty initial collection. By $C_n(m)$ we denote the number of draws needed to collect m remaining coupons, i.e., starting from the collection of $n - m$ different coupons. The average and high probability bounds for $C_n(m)$

are well known, the formulation provided here are from Doerr and Künnemann's article [DK14].

Lemma C.2. $\mathbb{E}[C_n(m)/n] = \ln n + \gamma + O(1/m)$, *where* $\gamma = 0.577\ldots$ *is the Euler-Mascheroni constant.*

Proof. Suppose that we collected all types of coupons except i ones. Since the probability that the next drawn coupon is not in the collection is i/n, the number of drawings until we pick a coupon of a new type has the geometric distribution with success probability i/n. Therefore, $C_n = X_1 + \ldots + X_m$, where X_i has distribution Geom(i/n). Since $\mathbb{E}[X_i] = n/i$, we have that $\mathbb{E}[C_m] = nH_m$, where $H_m = \left(1 + \frac{1}{2} + \ldots + \frac{1}{m}\right)$ is the mth harmonic number. The claim follows from the fact that $H_m = \ln m + \gamma + O(1/m)$. □

Lemma C.3. *Let* $1 \leq m < n$ *and* $r > 0$. *Then,*

$$\mathbb{P}[C_n(m) \geq n \ln(m) + rn] \leq e^{-r}.$$

Proof. The probability that one coupons is never drawn after t attempts is $\left(1 - \frac{1}{n}\right)^t$. Thus, the probability that one of m missing coupons is never seen after $t := n \ln m + rn$ is at most

$$m \left(1 - \tfrac{1}{n}\right)^{n \ln m + rn} \leq m \cdot e^{-\ln m - r} = e^{-r}.$$

□

The two-sided formulation of lemma above is possible by using Chebyshev inequality.

Lemma C.4. $\mathbb{P}[|C_n(n) - nH_n| \geq \lambda n] \leq \frac{\pi^2}{6\lambda^2}$.

We also note that even in a case of non-equal probabilities of drawing each coupons the result is possible. Flajolet, Gardy, Thimonier [FGT92] proved that if one draws coupons with respect to some probability distribution p_i, then the expectation of C_n can be estimated as follows.

$$\mathbb{E}[C_n] = \int_0^\infty \left(1 - \prod_{i=1}^{n} (1 - e^{-p_i t})\right) dt.$$

Bibliography

[ACMW14] Hüseyin Acan, Andrea Collevecchio, Abbas Mehrabian, and Nick Wormald. On the push&pull protocol for rumour spreading. *CoRR*, abs/1411.0948, 2014.

[BA99] Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *Science*, 286:509–512, 1999.

[Bai57] Norman T. J. Bailey. *The mathematical theory of epidemics*. Griffin London, 1957.

[BBCS05] Noam Berger, Christian Borgs, Jennifer T. Chayes, and Amin Saberi. On the spread of viruses on the internet. In *Proceedings of the Sixteenth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 301–310. SIAM, 2005.

[BEPS14] Pawel Brach, Alessandro Epasto, Alessandro Panconesi, and Piotr Sankowski. Spreading rumours without the network. In *Proceedings of the Second ACM Conference on Online Social Networks, (COSN)*, pages 107–118, 2014.

[BGPS06] Stephen P. Boyd, Arpita Ghosh, Balaji Prabhakar, and Devavrat Shah. Randomized gossip algorithms. *IEEE Trans. Information Theory*, 52:2508–2530, 2006.

[BR03] Béla Bollobás and Oliver Riordan. Robustness and vulnerability of scale-free random graphs. *Internet Mathematics*, 1:1–35, 2003.

[CCD+16] Andrea E. F. Clementi, Pierluigi Crescenzi, Carola Doerr, Pierre Fraigniaud, Francesco Pasquale, and Riccardo Silvestri. Rumor spreading in random evolving graphs. *Random Structures and Algorithms*, 48:290–312, 2016.

[CLP09] Flavio Chierichetti, Silvio Lattanzi, and Alessandro Panconesi. Rumor spreading in social networks. In *Proceedings of the 36th International Colloquium on Automata, Languages and Programming (ICALP)*, pages 375–386. Springer, 2009.

[CLP10] Flavio Chierichetti, Silvio Lattanzi, and Alessandro Panconesi. Almost tight bounds for rumour spreading with conductance. In *Proceedings of the 42nd ACM Symposium on Theory of Computing (STOC)*, pages 399–408. ACM, 2010.

[DFF11] Benjamin Doerr, Mahmoud Fouz, and Tobias Friedrich. Social networks spread rumors in sublogarithmic time. In *Proceedings of the 43rd ACM Symposium on Theory of Computing (STOC)*, pages 21–30. ACM, 2011.

[DFF12a] Benjamin Doerr, Mahmoud Fouz, and Tobias Friedrich. Asynchronous rumor spreading in preferential attachment graphs. In *Proceedings of the 13th Scandinavian Symposium and Workshops on Algorithm Theory (SWAT)*, pages 307–315. Springer, 2012.

[DFF12b] Benjamin Doerr, Mahmoud Fouz, and Tobias Friedrich. Experimental analysis of rumor spreading in social networks. In *Proceedings of the Design and Analysis of Algorithms - First Mediterranean Conference on Algorithms (MedAlg)*, pages 159–173, 2012.

[DFF12c] Benjamin Doerr, Mahmoud Fouz, and Tobias Friedrich. Why rumors spread so quickly in social networks. *Commununications of the ACM*, 55:70–75, 2012.

[DFS08] Benjamin Doerr, Tobias Friedrich, and Thomas Sauerwald. Quasirandom rumor spreading. In *Proceedings of the Nineteenth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 773–781, 2008.

[DGH+87] Alan J. Demers, Daniel H. Greene, Carl Hauser, Wes Irish, John Larson, Scott Shenker, Howard E. Sturgis, Daniel C. Swinehart, and Douglas B. Terry. Epidemic algorithms for replicated database maintenance. In *Proceedings of the Sixth Annual ACM Symposium on Principles of Distributed Computing (PODC)*, pages 1–12. ACM, 1987.

[DHL13] Benjamin Doerr, Anna Huber, and Ariel Levavi. Strong robustness of randomized rumor spreading protocols. *Discrete Applied Mathematics*, 161:778–793, 2013.

[DK14] Benjamin Doerr and Marvin Künnemann. Tight analysis of randomized rumor spreading in complete graphs. In *Proceedings of the Eleventh Workshop on Analytic Algorithmics and Combinatorics (ANALCO)*, pages 82–91. SIAM, 2014.

[DK17] Benjamin Doerr and Anatolii Kostrygin. Randomized rumor spreading revisited. In *Automata, Languages, and Programming - 44th International Colloquium, (ICALP)*, page to appear, 2017.

[DKM+10] Alexandros G. Dimakis, Soummya Kar, José M. F. Moura, Michael G. Rabbat, and Anna Scaglione. Gossip algorithms for distributed signal processing. *Proceedings of the IEEE*, 98:1847–1864, 2010.

[DKM15] Sebastian Daum, Fabian Kuhn, and Yannic Maus. Rumor spreading with bounded in-degree. *CoRR*, abs/1506.00828, 2015.

[Ede61] Murray Eden. A two-dimensional growth process. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 4: Contributions to Biology and Problems of Medicine*, pages 223–239. University of California Press, 1961.

[ES09] Robert Elsässer and Thomas Sauerwald. On the runtime and robustness of randomized broadcasting. *Theoretical Computer Science*, 410:3414–3427, 2009.

[FG85] Alan M. Frieze and Geoffrey R. Grimmett. The shortest-path problem for graphs with random arc-lengths. *Discrete Applied Mathematics*, 10:57–77, 1985.

[FG09] Nazim Fatès and Lucas Gerin. Examples of fast and slow convergence of 2d asynchronous cellular systems. *J. Cellular Automata*, 4:323–337, 2009.

[FGT92] Philippe Flajolet, Danièle Gardy, and Loÿs Thimonier. Birthday paradox, coupon collectors, caching algorithms and self-organizing search. *Discrete Applied Mathematics*, 39:207–229, 1992.

[FHP10] Nikolaos Fountoulakis, Anna Huber, and Konstantinos Panagiotou. Reliable broadcasting in random networks and the effect of density. In *Proceedings of the 29th International Conference on Computer Communications (INFOCOM)*, pages 2552–2560. IEEE, 2010.

[FP10] Nikolaos Fountoulakis and Konstantinos Panagiotou. Rumor spreading on random regular graphs and expanders. In *13th International*

Workshop on Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX), pages 560–573. Springer, 2010.

[FPRU90] Uriel Feige, David Peleg, Prabhakar Raghavan, and Eli Upfal. Randomized broadcast in networks. *Random Structures and Algorithms*, 1:447–460, 1990.

[FPS12] Nikolaos Fountoulakis, Konstantinos Panagiotou, and Thomas Sauerwald. Ultra-fast rumor spreading in social networks. In *Proceedings of the Twenty-Third Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1642–1660. SIAM, 2012.

[Gia11] George Giakkoupis. Tight bounds for rumor spreading in graphs of a given conductance. In *28th International Symposium on Theoretical Aspects of Computer Science (STACS)*, pages 57–68. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, 2011.

[Gia14] George Giakkoupis. Tight bounds for rumor spreading with vertex expansion. In *Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 801–815. SIAM, 2014.

[GKM03] Ayalvadi J. Ganesh, Anne-Marie Kermarrec, and Laurent Massoulié. Peer-to-peer membership management for gossip-based protocols. *IEEE Trans. Computers*, 52:139–149, 2003.

[GN64] William Goffman and Vaun A. Newill. Generalization of epidemic theory: An application to the transmission of ideas. *Nature*, 204:225–228, 1964.

[Gon81] Gaston H. Gonnet. Expected length of the longest probe sequence in hash code searching. *ACM*, 28:289–304, 1981.

[GS12] George Giakkoupis and Thomas Sauerwald. Rumor spreading and vertex expansion. In *Proceedings of the Twenty-Third Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1623–1641. SIAM, 2012.

[IvS10] Konrad Iwanicki and Maarten van Steen. Gossip-based self-management of a recursive area hierarchy for large wireless sensornets. *IEEE Transactions on Parallel and Distributed Systems*, 21:562–576, 2010.

[Jan99] Svante Janson. One, two and three times $\log n/n$ for paths in a complete graph with random weights. *Combinatorics, Probability and Computing*, 8:347–361, 1999.

[Jel11] Márk Jelasity. Gossip. In *Self-organising Software - From Natural to Artificial Adaptation*, pages 139–162. 2011.

[KC17] Marcos A. Kiwi and Christopher Thraves Caro. FIFO queues are bad for rumor spreading. *IEEE Trans. Information Theory*, 63:1159–1166, 2017.

[KDG03] David Kempe, Alin Dobra, and Johannes Gehrke. Gossip-based computation of aggregate information. In *Proceedings of the 44th Symposium on Foundations of Computer Science (FOCS)*, pages 482–491, 2003.

[Kle08] Jon M. Kleinberg. The convergence of social and technological networks. *Communications of the ACM*, 51:66–72, 2008.

[KM27] W. O. Kermack and Ag McKendrick. A Contribution to the Mathematical Theory of Epidemics. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, 115:700–721, 1927.

[KSC78] Valentin F. Kolchin, Boris A. Sevastianov, and Vladimir P. Chistiakov. *Random allocations; translation ed., A. V. Balakrishnan.* V. H. Winston; distributed solely by Halsted Press Washington: New York, 1978.

[KSSV00] Richard M. Karp, Christian Schindelhauer, Scott Shenker, and Berthold Vöcking. Randomized rumor spreading. In *Proceedings of the Annual Symposium on Foundations of Computer Science (FOCS)*, pages 565–574. IEEE, 2000.

[MP14] Abbas Mehrabian and Ali Pourmiri. Randomized rumor spreading in poorly connected small-world networks. In *Proceedings of the 28th International Symposium on Distributed Computing (DISC)*, pages 346–360. Springer, 2014.

[MRRS01] Dahlia Malkhi, Ohad Rodeh, Michael K. Reiter, and Yaron Sella. Efficient update diffusion in byzantine environments. In *Proceedings of the 20th Symposium on Reliable Distributed Systems (SRDS)*, pages 90–98, 2001.

[MS06] Damon Mosk-Aoyama and Devavrat Shah. Computing separable functions via gossip. In *Proceedings of the Twenty-Fifth Annual ACM Symposium on Principles of Distributed Computing, (PODC)*, pages 113–122. ACM, 2006.

[MS08] Damon Mosk-Aoyama and Devavrat Shah. Fast distributed algorithms for computing separable functions. *IEEE Trans. Information Theory*, 54:2997–3007, 2008.

[MSF+12] Miguel Matos, Valerio Schiavoni, Pascal Felber, Rui Oliveira, and Etienne Riviere. BRISA: combining efficiency and reliability in epidemic data dissemination. In *26th IEEE International Parallel and Distributed Processing Symposium (IPDPS)*, pages 983–994, 2012.

[Pit87] Boris Pittel. On spreading a rumor. *SIAM Journal on Applied Mathematics*, 47:213–223, 1987.

[PPS15] Konstantinos Panagiotou, Ali Pourmiri, and Thomas Sauerwald. Faster rumor spreading with multiple calls. *The Electronic Journal of Combinatorics*, 22:P1.23, 2015.

[Ric73] Daniel Richardson. Random growth in a tessellation. *Mathematical Proceedings of the Cambridge Philosophical Society*, 74:515–528, 1973.

[SMK+01] Ion Stoica, Robert Morris, David R. Karger, M. Frans Kaashoek, and Hari Balakrishnan. Chord: A scalable peer-to-peer lookup service for internet applications. In *SIGCOMM*, pages 149–160, 2001.

[TXD03] Chunqiang Tang, Zhichen Xu, and Sandhya Dwarkadas. Peer-to-peer information retrieval using self-organizing semantic overlay networks. In *Proceedings of the ACM Conference on Applications, Technologies, Architectures, and Protocols for Computer Communication (SIGCOMM)*, pages 175–186, 2003.

[VKMvS04] Spyros Voulgaris, Anne-Marie Kermarrec, Laurent Massoulié, and Maarten van Steen. Exploiting semantic proximity in peer-to-peer content searching. In *Proceedings of the 10th IEEE International Workshop on Future Trends of Distributed Computing Systems (FTDCS)*, pages 238–243, 2004.

université PARIS-SACLAY
ÉCOLE DOCTORALE
Sciences et technologies de l'information et de la communication (STIC)

Titre : Analyse précise des algorithmes épidemiques

Mots clefs : Algorithmes épidémiques, algorithmes basés sur les ragots, graphes aléatoires

Résumé : Dans cette thèse nous étudions le *problème de propagation de rumeur*, c'est-à-dire la dissémination collaborative, robuste et anonyme d'une information entre tous les agents d'un système distribué. Ce problème est à la base de nombreux algorithmes de communication sur des réseaux de capteurs sans-fil [Dimakis et al. (2010)] ou des réseaux mobiles ad-hoc. Il est aussi une brique centrale pour les nombreux algorithmes distribués avancés [Mosk-Aoyama et Shah (2008)].
Nous proposons une méthode générale d'analyse des protocoles de la propagation de rumeur dans les graphes complets. Contrairement aux résultats précédents basés sur la structure précise des processus étudiés, notre analyse est basée sur la probabilité et la covariance des évènements correspondants au fait qu'un agent non-informé s'informe. Cela nous permet de reproduire les résultats basiques concernant les protocoles classiques de push, pull et push-pull ainsi qu'analyser les certaines variations telles que les échecs de communications ou les communications multiples réalisées par chaque agent. De plus, nous sommes capables d'analyser les certains modèles dynamiques quand le réseaux forme un graphe aléatoire échantillonné à nouveau à chaque étape [Clementi et al. (ESA 2013)]. Notre méthode nous permet de déterminer l'espérance du temps de la diffusion à une constante additive près, ce qu'il est plus précis que la plupart des résultats précédents. Nous montrons que la déviation du temps de la diffusion par rapport à son espérance est inférieure d'une constante r avec la probabilité au moins $1 - \exp(\Omega(r))$.
Nous discutons d'une hypothèse classique que les agents peuvent répondre à plusieurs appels entrants. La restriction à un seul appel entrant par agent provoque un relantissement important de la diffusion pour un protocole de push-pull. Nous proposons une variation simple du protocole de push-pull qui rétablit une phase double logarithmique à nouveau et donc le nombre de messages passés redescend sur sa valeur optimal.

Title : Precise Analysis of Epidemic Algorithms

Keywords : Distributed algorithm, epidemic algorithm, rumor spreading, dynamic graph

Abstract : In epidemic algorithms the agents in the network involve peers similarly to the spread of epidemics. In this work, we focus on the randomized rumor spreading – a class of epidemic algorithms. We suppose that nodes of the network communicate with neighbors chosen at random. This approach has found numerous applications from the consistency maintenance of replicated databases to news spreading in social networks. Numerous mathematical analyses of different rumor spreading protocols can be found in the literature. Although some of them provide extremely sharp estimates for the performance of such processes, most of them are based on the inherent properties of concrete protocols.
We develop new simple and generic method to analyze randomized rumor spreading processes in complete graphs. In contrast to all previous works, we only need to understand the probability and the covariance of the events that uninformed nodes learn the rumor. This universality allows us to easily analyze the classic push, pull, and push-pull protocols in their pure version as well as in several variations such as when messages fail with constant probability or when nodes call a random number of others each round. Some dynamic models can also be analyzed, e.g., when the network is a random graph sampled independently each round [Clementi et al. (ESA 2013)]. Despite this generality, our method determines the expected rumor spreading time precisely apart from additive constants, i.e., more precise than almost all previous works. We show that a deviation from the expectation by more than r rounds occurs with probability at most $\exp(-\Omega(r))$.
We further use our method to discuss the assumption that nodes can answer any number of incoming calls. Restricting that only one call can be answered, we observe a significant increase of the runtime of the push-pull protocol. In particular, the double logarithmic end phase of the process now takes logarithmic time. This also increases the message complexity from the asymptotically optimal $\Theta(n \log \log n)$ [Karp, Shenker, Schindelhauer, Vöcking (FOCS 2000)] to $\Theta(n \log n)$. We propose a simple variation of the push-pull protocol that reverts back to the double logarithmic end phase and thus to the $\Theta(n \log \log n)$ message complexity.

Université Paris-Saclay
Espace Technologique / Immeuble Discovery
Route de l'Orme aux Merisiers RD 128 / 91190 Saint-Aubin, France