ABSTRACT

HUANG, YUFAN. Gossip-based Information Spreading in Mobile and Social Networks. (Under the direction of Dr. Huaiyu Dai.)

We study the performance of gossip-based information spreading in different network models. Our study is conducted in two major different circumstances: mobile networks and social networks. In mobile networks, we mainly study two problems: single-piece information spreading with general transmission range and multi-piece information spreading in connected networks. Towards understanding the gossip algorithm, first, we study single-piece message spreading with a general transmission range. While the results in connected networks are relatively easy to obtain, some critical approximation fails to hold in the disconnected scenario, and we propose general mobile conductance with a new analytical method.With general mobile conductance, we observe a penalty factor of nr2 induced by network connectivity deficiency. We also combine the concepts of mobile conductance and static k-conductance to define a new metric, mobile k-conductance, which enables us to analyze multi-piece message spreading in connected mobile networks.

In social networks, we consider the celebrated Preferential Attachment model. Inspired by the Preferential Attachment process, we propose a new non-uniform gossip algorithm called PA-gossip for effective information spreading in social networks. Large-scale network simulation is conducted to demonstrate its great potential in facilitating information spreading in social networks.

Gossip-based Information Spreading in Mobile and Social Networks

by
Yufan Huang

A thesis submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Master of Science

Electrical Engineering

Raleigh, North Carolina

2014

APPROVED BY:

Dr. Brian Hughes

Dr. Dror Baron

Dr. Huaiyu Dai
Chair of Advisory Committee

DEDICATION

To my parents and sister, your love keeps me moving forward.

BIOGRAPHY

Yufan Huang received his Bachelor of Engineering degree in Electrical and Information Engineering from Beihang University in 2012. He joined the Master of Science in Electrical Engineering program at North Carolina State University in 2012.

ACKNOWLEDGEMENTS

I would like to express my gratitude to my advisor Dr. Huaiyu Dai for his valuable guidance and suggestions for my thesis. I am thankful to his persistent help in my work and study.

I am also thankful to the rest of my thesis committee: Dr. Dror Baron and Dr. Brian Hughes, for their questions and insightful comments.

I am grateful to my colleagues Xiaofan, Huazi, and Juan who made my life in Raleigh enjoyable.

TABLE OF CONTENTS

LIST OF FIGURES

Chapter 1

Introduction

We are living in exciting times. Communication in human society has increased intensively during the last few decades. The appearance of mobile devices and Internet-based services continually allow more convenient and comfortable communication. Meanwhile, we are witnessing incredible growth of such communication networks both in size and complexity along with technological advancements. In these large-scale networks, a variety of scenarios require agents to share their individual information or resources with each other for mutual benefits, such as file sharing and rumor spreading, distributed computing and estimation, and cooperative scheduling and control. Recently, by constructing global Internet platforms and developing social applications over the traditional technological networks, large-scale social networks have been formed so that information can be exchanged more effectively. Considering traditional technological mobile networks and emerging social networks, one must understand how information propagates in these networks. In addition, another important application for these large-scale networks is the ability to spread information quickly and efficiently to the whole network. For example, in emerging vehicular ad hoc networks (VANETs), traffic control and safety information must be promptly disseminated. Similarly, in wireless ad hoc and sensor networks set up for disaster monitoring or rescue, it is of paramount importance to ensure that the emergency or life-critical message can be spread out in the shortest time.

Since these networks are complex and exhibit unpredictable dynamics, information should be disseminated and messages exchanged in a decentralized and asynchronous manner to combat unpredictable topology changes and the lack of a global state of information. This motivates the exploration of dissemination strategies that are inherently simple, distributed, and asynchronous, while achieving efficient spreading results. Among distributed asynchronous algorithms, the gossip algorithm [8] is particularly appealing, wherein information spreads by mimicking a rumor-spreading style; that is, in each round, every node randomly selects one of its neighbors for message exchange during the information propagation process. In contrast to other types of distributed algorithms, such as flooding, gossip algorithms can achieve near-optimal performance for a large class of static network graphs including random geometric graphs. Therefore, our work will mainly focus on gossip-based information spreading in large-scale dynamic networks, including mobile and social networks.

Information spreading in static networks has been well studied in existing literature [18][19]. Spreading time in static networks is closely related to a graph expansion property of the network called "conductance", which essentially represents the bottleneck for information propagation within a network. Information spreading time in networks with symmetric representing matrices is reciprocal to the conductance of the graph [19].

In the traditional study of wireless networks, mobility is viewed as a negative feature because it adds additional uncertainty to wireless networks and incurs more challenges in channel estimation. Recently, mobility has been revisited for its potential to improve network performance. In [13], mobility is shown to significantly increase the sum-throughput of the network under the fully random mobility model; later, the study is extended to a one-dimensional mobility model [9]. Subsequently, throughput-delay tradeoff is further investigated in the context of mobile ad-hoc networks [11]-[17]. Additionally, some interesting analytic results for information spreading in mobile networks have emerged. The scaling properties of information propagation between a pair of nodes in large mobile wireless networks are explored in [16]. An upper bound of the information propagation speed for the flooding mechanism is derived in [14] for the random walk mobility model, with emphasis on sparse networks. Most interestingly, by modeling the mobile network as a Markovian evolving graph, it extends the concept of "conductance" in static networks to corresponding "mobile conductance" in mobile networks [21]. In this work, spreading time in mobile networks is also seen as reciprocal to mobile conductance. Then, by evaluating the corresponding mobile conductance, information spreading time is quantified in some classic mobility models, such as Fully Random Mobility and Velocity Constrained Mobility.

Information spreading in social networks is another interesting research area. Built upon traditional technological networks, social networks possess many interesting features, such as preferential attachment, power-law degree distribution, small-world phenomena, and clustering/community structure [4]. Currently, a prevalent approach for studying information spreading in social networks relates to a classic epidemic model: the SI model. In the SI model, nodes can be categorized into two types: susceptible (S) and infected (I). The SI model uses the following differential equations to model information spreading:

$$\begin{aligned} \frac{dS(t)}{dt} &= -\beta S(t)I(t), \\ \frac{dI(t)}{dt} &= \beta S(t)I(t). \end{aligned} \tag{1.1}$$

Many studies have been done using the SI model. In [7], the authors study a modified SI model with time variant infection rates, which are determined by social ties in social networks. By keeping track of individual interests, [12] proposes an agent-based SI model. Apart from the SI model, gossip-based information spreading is also studied in social networks. Conductance and other similar quantities (i.e., min-cut) still play an important role in gossip-based information spreading in social networks [5][6]; while, in social networks, due to the human participation, the information propagation will be quite different from technological networks in which messages are digitized and little attention is paid to information content. However, in social networks, people are assumed to have different interests for different messages and may show preference from one to another. Even for the same type of messages, one users interest may change, depending on whether it is the first message he/she receives, the influence from his friends (such as the "0-1-2" effect [15]), and many other factors. Therefore, both information content and its timing will influence the spreading behavior of a message in social networks, which is particularly true for the multi-piece information spreading problem in social networks, where multiple messages may originate from different, potentially competing, parties.

Inspired by the above pioneering works on information spreading, this thesis addresses some interesting problems concerning information spreading in both technological and social networks. We realize that the majority of studies on information spreading are based on the assumption that networks are always connected, which may not always be true in practice, which motivates us to study single-piece message spreading in disconnected mobile networks and, then, present the results with general transmission radius.

Multi-piece message spreading, all nodes have information and hope to spread it to the other nodes in the networks, in static networks is studied [19] from the concept of k-conductance. We extend this concept to "mobile k-conductance" in order to study multi-piece message spreading in mobile networks.

Finally, the gossip algorithms performance in social networks is explored. Knowing that information can spread extremely fast in social networks due to inherent special properties, we propose a new gossip algorithm to further improve information spreading performance in social networks.

Our contributions are summarized as follows:

- We develop "general mobile conductance" of different mobility models, which is applied for analyzing single-piece information spreading in sparse mobile networks (when $nr^2 = o(1)$ for a network of size n and node transmission range of r). An interesting $\Theta(nr^2)$ penalty factor is revealed when compared to its counterpart in connected networks. The performance gap is justified with simulations.

- We give a quantitative tradeoff analysis between mobility and connectivity in terms of information spreading effectiveness, which provides insights into how mobility exploitation occurs in order to compensate for network connectivity deficiency.

- We combine the notions of "mobile conductance" and "k-conductance" to propose a new metric, mobile k-conductance, which facilitates the analysis of multi-piece message spreading in mobile networks. We evaluate the mobile k-conductance for various mobility models such as, One-Dimensional Mobility and Velocity Constrained Mobility, upon which the resulting multi-piece information spreading time is derived for these mobility models.

- By exploring the Preferential Attachment (PA) networks (typical models of social networks), we propose a new gossip algorithm, Preferential Attachment (PA) Gossip, to facilitate information spreading in PA networks. Large-scale network simulation has been conducted to illustrate performance improvement results yielded by our new algorithm.

This thesis is organized as follows: Chapter 2 gives the analysis of single-piece message spreading with general transmission range and corresponding simulation results. Chapter 3 presents the analysis of multi-piece message spreading in mobile networks with mobile k-conductance. Chapter 4 presents an efficient gossip algorithm for information spreading in social networks and corresponding large-scale network simulation results. Finally, Chapter 5 presents conclusions of study results.

Chapter 2

Single-piece message spreading with general transmission range

In [21], mobile conductance is proposed to facilitate the analysis of single-piece information spreading in connected mobile networks. In [20], we gradually transform network connectivity from connected to disconnected, and results indicate that information spreading performance will exhibit non-trivial changes. In this chapter, we present the analysis of information spreading in mobile networks with general transmission range.

2.1 System Model

We consider an n-node mobile network on a unit square, modeled as a time-varying graph $G_t = (V, E_t)$ evolving over discrete time steps. The set of nodes are represented as vertex set V and identified by the first n positive integers. One key difference between a mobile network and its static counterpart is that the locations of nodes change over time, according to certain mobility models, as do the connections between the nodes represented by the edge set E_t. Denote the position of node i at time t as $X_i(t)$ and the node transmission range as r. If $|X_i(t) - X_j(t)| < r$, (i, j) belongs to edge set E_t, and j belongs to $N_i(t)$, the neighboring set of i at the beginning of time slot t.

It is assumed that the moving process of all the nodes $\{X_i(t), t \in N\}, i \in [n]$ are independently and identically distributed (i.i.d.) stationary Markov chains with transition distribution q_i and collectively denoted by $\{X(t), t \in N\}$ with transition distribution $Q = \Pi_{i=1}^{n} q_i$. The transition matrix Q is defined by entries $q_{xy}^i = q_i(X_i(t+1) = y|X_i(t) = x), x, y \in \Omega, \forall i$. Here, we assume the celebrated random geometric graph (RGG) model for the initial node distribution for the concreteness (i.e., $G_0 = G(n, r)$, where r is the common transmission range). Under most existing random mobility models, nodes will maintain uniform distribution on state space over time.

2.2 Mobility Model

The unit square is discretized into the grid with sufficient high resolution $\delta : \Omega = \{(i\delta, j\delta)|0 \le i, j \le \lfloor 1/\delta \rfloor\}$. The following mobility models are considered in this study:

Fully Random Mobility: $X_i(t)$ is uniformly distributed on and i.i.d. over time. In this case, $q_{xy}^i = 1/|\Omega|, \forall i, \forall x, y \in \Omega$. This idealistic model is often adopted to explore the largest possible improvement produced by mobility.

Partially Random Mobility: k randomly chosen nodes are mobile, following the fully random mobility model, while the rest $n - k$ nodes stay static. This is one generalization of a fully random mobility model.

Velocity Constrained Mobility: This is another generalization of a fully random mobility model, with the node speed bounded as $|X_i(t+1) - X_i(t)| \le v_{\max}$. In this case, $q_{xy}^i = 1/|C(x)|, \forall i$ and $y \in C(x)$, where $C(x) = \{y \in \Omega||y - x| \le v_{\max}\}$; and $q_{xy}^i = 0$, otherwise.

One-Dimensional Area Constrained Mobility: In this model, the mobile nodes move either vertically (named V-nodes) or horizontally (named H-nodes), reminiscent of trains or automobiles moving on railways or city streets. It is assumed that both V-nodes and H-nodes are uniformly and randomly distributed on Ω, and the mobility pattern of each node is "fully random" on the corresponding one-dimensional path. Let $x \triangleq (x_\alpha, x_\beta) \in \Omega$ and $y \triangleq (y_\alpha, y_\beta) \in \Omega$. For a V-node, $q^i_{xy} = 1/(\lfloor 1/\delta \rfloor + 1)$, $\forall i$ and $y \in V(x)$, where $V(x) = \{y \in \Omega | y_\alpha = x_\alpha\}$; and $q^i_{xy} = 0$, otherwise. The transition probability for an H-node is similarly defined.

2.3 Spreading Model

Denote $S(t)$ as the set of informed nodes at the beginning of time t with $S(0) = \{s\}$ where s is the random source node. Here, we take the "Move-and-Gossip" paradigm proposed in [21], i.e., each time slot is divided into two phases: each node first moves and then gossips with one of its new neighboring nodes. Note that the node's position $X_i(t)$ changes in the middle of each time slot (after the move step), while $S(t)$ is updated at the end (after the gossip step), according to the gossip algorithm. Specifically, node i contacts one of its neighbors, say node $j \in N_i(t+1)$ with probability $P_{ij}(t+1)$. Here, $P_{ij}(t+1)$ is set as $1/|N_i(t+1)|$ for $j \in N_i(t+1)$, and 0, otherwise. Therefore, the metric of interest is ϵ-spreading time and is defined as:

$$T_{spr}(\epsilon) \triangleq \sup_{s \in V} \inf\{t : \Pr(|S(t)| \neq n | S(0) = s) \leq \epsilon\} \tag{2.1}$$

Recall the mobile conductance [21] is defined for a stationary Markovian evolving graph as

$$\Phi_m(Q) \triangleq \min_{S'(t)\subset V, |S'(t)|\le n/2} \left\{ E_Q \left(\frac{\sum_{i\in S'(t), j\in \overline{S'(t)}} P_{ij}(t+1)}{|S'(t)|} \right) \right\}. \tag{2.2}$$

After revisiting the derivation of T_{spr} in [21], we know that no connectivity requirement is imposed in it. Thus,

$$T_{spr}(\epsilon, Q) = O\left(\frac{\log n + \log \epsilon^{-1}}{\Phi_m(Q)}\right) \tag{2.3}$$

still holds for a disconnected network setting.

In this work, my main contributions include the derivation of general mobile conductance for different mobility models, proposing a ring-graph benchmark for analyzing the effects of dynamism on disconnectedness, and simulation justification. Results are presented in the following part of this chapter.

2.4 Evaluation of Several Mobility Models

In [21], it is assumed that the network remains connected under mobility, i.e., at each instant, there exist some contact pairs between the message set and non-message set after the move. Under such assumption, the contact probabilities $P_{ij}(t+1)$s are on the order of $P(n,r) = \Theta(\frac{1}{n\pi r^2})$ for most random walk-based mobility models including our discussion models in RGG, and mobile conductance admits a simple expression. In sparse networks, this fundamental assumption no longer holds, and the previous method fails. Hence, we develop a new method to directly evaluate mobile conductance, which works for a general transmission range r; the results we obtain with this method agree with those in [21] and reveal a penalty factor of $\Theta(nr^2)$ when $nr^2 = o(1)$.

2.4.1 Fully Random Mobility

For this mobility model, the probability that an arbitrary node j belongs to $N_i(t+1), \forall i \in S'(t)$ is given by

$$p_{i \leftrightarrow j} = \pi r^2. \tag{2.4}$$

The probability that i has m neighbors is

$$p_{i,m} \triangleq \begin{pmatrix} n-1 \\ m \end{pmatrix} (p_{i \leftrightarrow j})^m (1 - p_{i \leftrightarrow j})^{n-1-m}. \tag{2.5}$$

Among the m neighbors, the probability that b of them come from $\overline{S'(t)}$ is

$$p_{m,b} = \begin{pmatrix} m \\ b \end{pmatrix} (p_{i,\overline{S'(t)}})^b (1 - p_{i,\overline{S'(t)}})^{m-b} \tag{2.6}$$

where $p_{i,\overline{S'(t)}}$ denotes the probability of meaningful-contact, in which a uniformly and randomly chosen edge connects $i \in S'(t)$ with another node from $\overline{S'(t)}$ and, thus, offers an effective information transfer. Thanks to the uniform node distribution of $S'(t)$ and $\overline{S'(t)}$ after the move, $p_{i,\overline{S'(t)}}$ is identical for all i's and given by $p_{i,\overline{S'(t)}} = |\overline{S'(t)}|/(n-1)$.

The expected sum of contact probabilities related to node $i \in S'(t)$ is given by

$$E_Q\left(\sum_{j \in \overline{S'(t)}} P_{ij}(t+1)\right) = \sum_{m=1}^{n-1} p_{i,m} \sum_{b=1}^{m} \left(\frac{b}{m} p_{m,b}\right). \tag{2.7}$$

By the uniformity of all nodes in $S'(t)$, the quantity of our interest may be evaluated as

$$
\begin{aligned}
&E_Q\left(\sum_{i\in S'(t),j\in\overline{S'(t)}} P_{ij}(t+1)\right)\\
&=|S'(t)|E_Q\left(\sum_{j\in\overline{S'(t)}} P_{ij}(t+1)\right)\\
&=|S'(t)|\sum_{m=1}^{n-1}\tfrac{1}{m}p_{i,m}\sum_{b=1}^{m} bp_{m,b}\\
&=|S'(t)|\sum_{m=1}^{n-1}\tfrac{1}{m}p_{i,m}\tfrac{|\overline{S'(t)}|}{n-1}\\
&=\tfrac{|S'(t)||\overline{S'(t)}|}{n-1}\left(1-(1-p_{i\leftrightarrow j})^{n-1}\right)
\end{aligned}
\tag{2.8}
$$

With (2.2), (2.4), and (2.8), we have

$$
\begin{aligned}
\Phi_m^{FR}(Q) &= \min_{\substack{S'(t)\subset V\\ |S'(t)|\le n/2}}\left\{\frac{|\overline{S'(t)}|}{n-1}\left(1-\left(1-\pi r^2\right)^{n-1}\right)\right\}\\
&\simeq \frac{1}{2}\left(1-\left(1-\pi r^2\right)^{n-1}\right)\\
&=\begin{cases}\Theta(nr^2), nr^2=o(1),\\ \Theta(1), nr^2=\Omega(1).\end{cases}
\end{aligned}
\tag{2.9}
$$

Remarks: As can be seen from (2.8), for sparse networks with $r=o(\sqrt{1/n})$, information spreading can still be achieved (thanks to the high node mobility), but with a penalty factor of $\Theta(nr^2)$ when compared to the connected scenario.

2.5 Partially Random Mobility

We follow the lines of [21], and adopt a divide-and-conquer strategy. The sum probability is separated into subsets according to mobility. The mobile node and static node set are denoted by $D(t)$ and $\overline{D(t)}$, respectively. Therefore, $D(t)=k$ and $\overline{D(t)}=n-k$.

$$E_Q\left(\frac{1}{|S'(t)|}\sum_{i\in S'(t),j\in\overline{S'(t)}}P_{ij}(t+1)\right)$$
$$=\frac{1}{|S'(t)|}E_Q\left[\begin{array}{l}\sum_{i\in S'(t)\cap\overline{D(t)},j\in\overline{S'(t)}\cap\overline{D(t)}}P_{ij}(t+1)\\+\sum_{i\in S'(t)\cap D(t),j\in\overline{S'(t)}\cap D(t)}P_{ij}(t+1)\\+\sum_{i\in S'(t)\cap D(t),j\in\overline{S'(t)}\cap\overline{D(t)}}P_{ij}(t+1)\\+\sum_{i\in S'(t)\cap\overline{D(t)},j\in\overline{S'(t)}\cap D(t)}P_{ij}(t+1)\end{array}\right].$$

Therefore

$$\Phi_m(Q)=\left(\frac{n-k}{n}\right)^2\Phi_s+\frac{1}{|S'(t)|}\min_{\substack{S'(t)\subset V\\|S'(t)|\le n/2}}\left\{E_Q\left[\begin{array}{l}\sum_{i\in S'(t),j\in\overline{S'(t)}\cap D(t)}P_{ij}(t+1)\\+\sum_{i\in S'(t)\cap D(t),j\in\overline{S'(t)}\cap\overline{D(t)}}P_{ij}(t+1)\end{array}\right]\right\}$$

where Φ_s is the static conductance of RGG, and the expected sum of contact probabilities can be derived following the lines from (2.4) to (2.8), however through a more complicated process.

In order to get $\sum_{i\in S'(t),j\in\overline{S'(t)}\cap D(t)}P_{ij}(t+1)$ and $\sum_{i\in S'(t)\cap D(t),j\in\overline{S'(t)}\cap\overline{D(t)}}P_{ij}(t+1)$, we need to know, among the m neighbors, the probability that b of them come from $\overline{S'(t)}\cap D(t)$ and $\overline{S'(t)}\cap\overline{D(t)}$. The results for these two probabilities are as follows:

$$p^{m,b}_{\overline{S'(t)}\cap D(t)}=\begin{pmatrix}m\\b\end{pmatrix}(p_{i,\overline{S'(t)}\cap D(t)})^b(1-p_{i,\overline{S'(t)}\cap D(t)})^{m-b}, \quad (2.10)$$

$$p^{m,b}_{\overline{S'(t)}\cap\overline{D(t)}} = \begin{pmatrix} m \\ b \end{pmatrix} (p_{i,\overline{S'(t)}\cap\overline{D(t)}})^b (1 - p_{i,\overline{S'(t)}\cap\overline{D(t)}})^{m-b}, \tag{2.11}$$

where $p_{i,\overline{S'(t)}\cap D(t)} = \frac{k|\overline{S'(t)}|}{n^2}$ and $p_{i,\overline{S'(t)}\cap\overline{D(t)}} = \frac{(n-k)|\overline{S'(t)}|}{n^2}$.

With (2.10) and (2.11), and the uniformity of all nodes in $S(t)$ after the move, we can have

$$\begin{aligned}
\Phi_m(Q) =& \left(\frac{n-k}{n}\right)^2 \Phi_s \\
&+ \frac{1}{|S'(t)|} \min_{\substack{S'(t)\subset V \\ |S'(t)|\le n/2}} \left\{ \left[\begin{array}{c} \sum_{m=1}^{n-1}\sum_{b=1}^{m}\left(\frac{1}{m}|S'(t)|\, p_{i,m} b p^{m,b}_{\overline{S'(t)}\cap D(t)}\right) \\ +\sum_{m=1}^{n-1}\sum_{b=1}^{m}\left(\frac{1}{m}|S'(t)\cap D(t)|\, p_{i,m} b p^{m,b}_{\overline{S'(t)}\cap\overline{D(t)}}\right) \end{array} \right] \right\} \\
=& \left(\frac{n-k}{n}\right)^2 \Phi_s + \min_{\substack{S'(t)\subset V \\ |S'(t)|\le n/2}} \left\{ \frac{k|\overline{S'(t)}|}{n^2} \sum_{m=1}^{n-1} p_{i,m} + \frac{k}{n}\left(\frac{(n-k)|\overline{S'(t)}|}{n^2}\right) \sum_{m=1}^{n-1} p_{i,m} \right\} \\
=& \left(\frac{n-k}{n}\right)^2 \Phi_s + \left(\frac{k}{n}\left(1-(1-\pi r^2)^n\right) + \frac{k(n-k)}{n^2}\left(1-(1-\pi r^2)^n\right)\right).
\end{aligned}$$

Following the order analysis in (2.9), we can get the final results

$$\Phi^{PR}_m(Q) = \begin{cases} \left(\frac{n-k}{n}\right)^2 \Phi_s + \Theta\left(\frac{k(2n-k)}{n} r^2\right), nr^2 = o(1), \\ \left(\frac{n-k}{n}\right)^2 \Phi_s + \Theta\left(\frac{k(2n-k)}{n^2}\right), nr^2 = \Omega(1). \end{cases} \tag{2.12}$$

The overall additional factor between the mobility-contributed terms in "connected mobile conductance" and "disconnected mobile conductance" is still $n\pi r^2$.

2.6 One-Dimensional Mobility

The same divide-and-conquer strategy is adopted here. The sum-probability is separated into subsets according to moving directions: the subset of V-nodes S_V and the subset of H-nodes S_H

$$\begin{aligned}
&E_Q\left(\frac{1}{|S'(t)|}\sum_{i\in S'(t),j\in\overline{S'(t)}} P_{ij}(t+1)\right)\\
=&\frac{1}{|S'(t)|}E_Q\left[\begin{array}{c}
\sum\limits_{i\in S'(t)\cap S_V,j\in\overline{S'(t)}\cap S_V} P_{ij}(t+1)\\
+\sum\limits_{i\in S'(t)\cap S_H,j\in\overline{S'(t)}\cap S_H} P_{ij}(t+1)\\
+\sum\limits_{i\in S'(t)\cap S_V,j\in\overline{S'(t)}\cap S_H} P_{ij}(t+1)\\
+\sum\limits_{i\in S'(t)\cap S_H,j\in\overline{S'(t)}\cap S_V} P_{ij}(t+1)
\end{array}\right].
\end{aligned}$$

The general mobile conductance for a one-dimensional mobility model may be written as

$$\Phi_m(Q)=\frac{n_V^2+n_H^2}{n^2}\Phi_s+\frac{1}{|S'(t)|}\min_{\substack{S'(t)\subset V\\|S'(t)|\le n/2}}\left\{E_Q\left[\begin{array}{c}\sum\limits_{i\in S'(t)\cap S_V,j\in\overline{S'(t)}\cap S_H} P_{ij}(t+1)\\+\sum\limits_{i\in S'(t)\cap S_H,j\in\overline{S'(t)}\cap S_V} P_{ij}(t+1)\end{array}\right]\right\}.$$

Again, we need to get $\sum\limits_{i\in S'(t)\cap S_V,j\in\overline{S'(t)}\cap S_H} P_{ij}(t+1)$ and $\sum\limits_{i\in S'(t)\cap S_H,j\in\overline{S'(t)}\cap S_V} P_{ij}(t+1)$; thus, we need to know, among the m neighbors, the probability that b of them come from $\overline{S'(t)}\cap S_H$ and $\overline{S'(t)}\cap S_V$. The results for these two probabilities are as follows:

$$p^{m,b}_{\overline{S'(t)}\cap S_H}=\begin{pmatrix}m\\b\end{pmatrix}(p_{i,\overline{S'(t)}\cap S_H})^b(1-p_{i,\overline{S'(t)}\cap S_H})^{m-b},\qquad(2.13)$$

$$p^{m,b}_{\overline{S'(t)}\cap S_V} = \begin{pmatrix} m \\ b \end{pmatrix} (p_{i,\overline{S'(t)}\cap S_V})^b (1 - p_{i,\overline{S'(t)}\cap S_V})^{m-b}, \tag{2.14}$$

where $p_{i,\overline{S'(t)}\cap S_H} = \frac{n_H|\overline{S'(t)}|}{n^2}$ and $p_{i,\overline{S'(t)}\cap S_V} = \frac{n_V|\overline{S'(t)}|}{n^2}$.

With (2.13) and (2.14), and the uniformity of all nodes in $S(t)$ after the move, we can have

$$\begin{aligned}
\Phi_m(Q) =& \frac{n_V^2 + n_H^2}{n^2}\Phi_s \\
&+ \frac{1}{|S'(t)|} \min_{\substack{S'(t)\subset V \\ |S'(t)|\le n/2}} \left\{ \left[\begin{array}{c} \sum\limits_{m=1}^{n-1}\sum\limits_{b=1}^{m} \left(\frac{1}{m}\left|\overline{S'(t)}\cap S_H\right| p_{i,m} b p^{m,b}_{\overline{S'(t)}\cap S_H}\right) \\ + \sum\limits_{m=1}^{n-1}\sum\limits_{b=1}^{m} \left(\frac{1}{m}\left|\overline{S'(t)}\cap S_V\right| p_{i,m} b p^{m,b}_{\overline{S'(t)}\cap S_V}\right) \end{array} \right] \right\} \\
=& \frac{n_V^2 + n_H^2}{n^2}\Phi_s + \min_{\substack{S'(t)\subset V \\ |S'(t)|\le n/2}} \left\{ \frac{n_V}{n}\frac{n_H|\overline{S'(t)}|}{n^2}\sum_{m=1}^{n-1} p_{i,m} + \frac{n_H}{n}\left(\frac{n_V|\overline{S'(t)}|}{n^2}\right)\sum_{m=1}^{n-1} p_{i,m} \right\} \\
=& \frac{n_V^2 + n_H^2}{n^2}\Phi_s + \left(\frac{n_V n_H}{n^2}\left(1 - (1-\pi r^2)^n\right) + \frac{n_H n_V}{n^2}\left(1 - (1-\pi r^2)^n\right)\right).
\end{aligned}$$

Then, we can get the final results

$$\Phi_m^{1D}(Q) = \begin{cases} \frac{n_V^2+n_H^2}{n^2}\Phi_s + \Theta\left(\frac{n_V n_H}{n} r^2\right), nr^2 = o(1), \\ \frac{n_V^2+n_H^2}{n^2}\Phi_s + \Theta\left(\frac{n_V n_H}{n^2}\right), nr^2 = \Omega(1). \end{cases} \tag{2.15}$$

Still, the overall additional factor between the connected and disconnected cases is $n\pi r^2$.

2.7 Velocity Constrained Mobility

Results indicate that the bisection of the unit square with $S(t)$ on the left half-plane serves as a bottleneck segmentation in this case. Given this setting, each node may move uniformly to any point within the circle of radius $v_{\max}$ centered at its original position. The node distribution of $S(t)$ and $\overline{S(t)}$ after the move has been well studied in [21], and the overall diffusion process may be illustrated in Figure.2.1 and Figure.2.2, in which the darkness level of the area represents the density of nodes belonging to $S(t)$.

From Figure.2.1 and Figure.2.2, we can see that the node distribution of $S(t)$ and $\overline{S(t)}$ is no longer uniform; instead, it will be decided by the position of the node. Thus, in order to find out the sum probabilities, we need to define the meaningful-contact probability as follows.

Definition: The $\{l, r, v_{\max}\}$ meaningful-contact $p_{i,\overline{S'}}(l, r, v_{\max})$ is the probability that, under transmission radius r and velocity constraint $v_{\max}$, a uniformly and randomly chosen edge of node $i \in S'(t)$ with X-coordinate l connects i with another node in $\overline{S'(t)}$.

Also, $p_{i,\overline{S'}}(l, r, v_{\max})$ corresponds to the proportion of nodes without a message among all is neighbors (the r-radius circle centered at is position after the move), which can be calculated by integrating the density of nodes belonging to $\overline{S(t)}$ within the circle, and takes positive value when $-v_{\max} - r < l < v_{\max}$. As seen in Figures 2.1 and 2.2, a piecewise integral will involve two or three segments, depending on whether $r < v_{\max}$ and $r > v_{\max}$, denoted by *Case* 1 and *Case* 2, respectively.

In *Case* 1, i.e., $r < v_{\max}$, the calculation of $p_{i,\overline{S'}}(l, r, v_{\max})$ may be further divided into three subcases: (a) $-v_{\max} - r < l < -v_{\max} + r$, (b) $-v_{\max} + r < l < v_{\max} - r$, and (c) $v_{\max} - r < l < v_{\max}$.

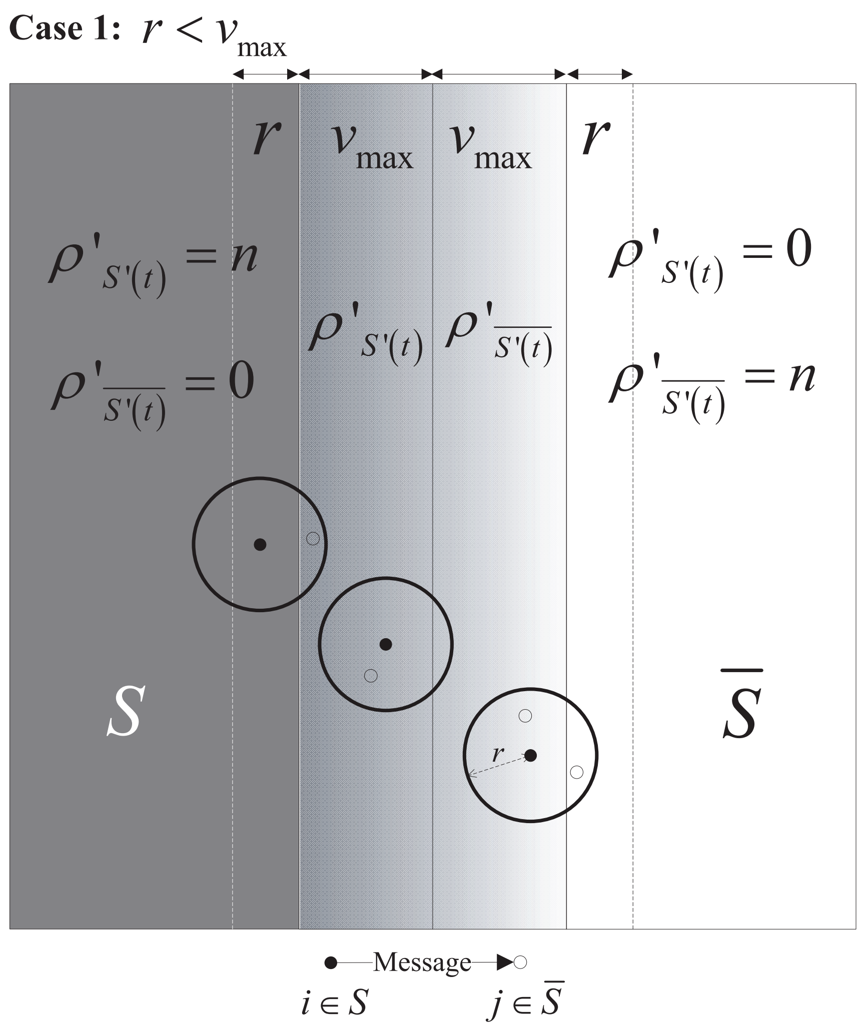

Figure 2.1: Case 1 for velocity constrained model

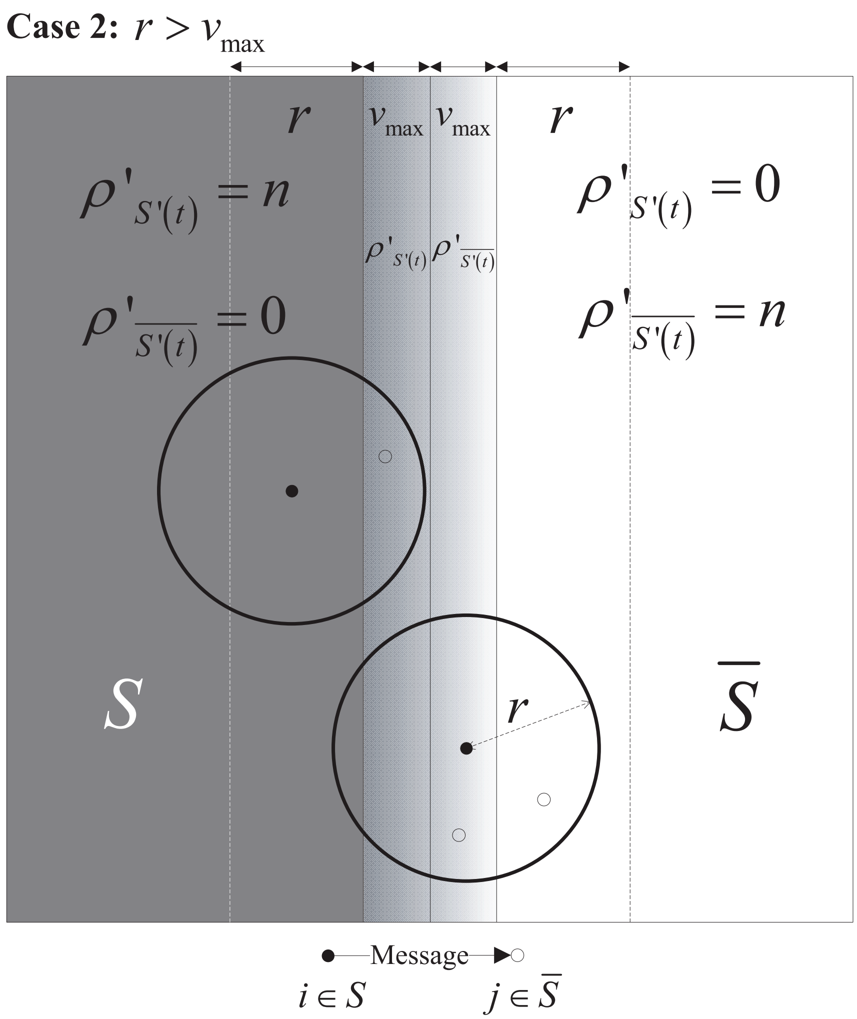

Figure 2.2: Case 2 for velocity constrained model

$$p_{i,\overline{S'}}(l,r,v_{\max}) = \begin{cases} \int\limits_{-v_{\max}}^{l+r} \frac{\rho'_{\overline{S}}(x)\left(2\sqrt{r^2-(x-l)^2}\right)}{n\pi r^2} dx, & \text{for } Case\ 1.a, \\ \int\limits_{l-r}^{l+r} \frac{\rho'_{\overline{S}}(x)\left(2\sqrt{r^2-(x-l)^2}\right)}{n\pi r^2} dx, & \text{for } Case\ 1.b, \\ \int\limits_{l-r}^{v_{\max}} \frac{\rho'_{\overline{S}}(x)\left(2\sqrt{r^2-(x-l)^2}\right)}{n\pi r^2} dx, & \text{for } Case\ 1.c. \\ +\frac{2(\alpha_l r^2-(v_{\max}-l)r\sin\alpha_l)}{n\pi r^2} & \end{cases} \quad (2.16)$$

where $\rho'_{\overline{S}}(\cdot)$ is the density of node without message as follows,

$$\rho'_{\overline{S}}(x) \triangleq \begin{cases} 0, & x \leq -v_{\max}, \\ (1+\frac{l\sin\beta_x}{\pi v_{\max}} - \frac{\beta_x}{\pi})n, & -v_{\max} < x < v_{\max}, \\ n, & x \geq v_{\max}, \end{cases} \quad (2.17)$$

and $\alpha_l \triangleq \arccos\frac{v_{\max}-l}{r}$, $\beta_x \triangleq \arccos\frac{x}{v_{\max}}$.

In *Case* 2, i.e., $r > v_{\max}$, the calculation of $p_{i,\overline{S'}}(l,r,v_{\max})$ may be further divided into two sub-cases: (a)$-v_{\max} - r < l < v_{\max} - r$, and (b) $v_{\max} - r < l < v_{\max}$.

$$p_{i,\overline{S'}}(l,r,v_{\max}) = \begin{cases} \int\limits_{-v_{\max}}^{l+r} \frac{\rho'_{\overline{S}}(x)\left(2\sqrt{r^2-(x-l)^2}\right)}{n\pi r^2} dx, & \text{for } Case\ 2.a, \\ \int\limits_{-v_{\max}}^{v_{\max}} \frac{\rho'_{\overline{S}}(x)\left(2\sqrt{r^2-(x-l)^2}\right)}{n\pi r^2} dx, & \text{for } Case\ 2.b. \\ +\frac{2(\alpha_l r^2-(v_{\max}-l)r\sin\alpha_l)}{n\pi r^2} & \end{cases} \quad (2.18)$$

For notation simplicity, we denote the three probabilities in (2.16) as $p^{1a}_{i,\overline{S'}}$, $p^{1b}_{i,\overline{S'}}$, and $p^{1c}_{i,\overline{S'}}$ for *Case* 1.a, *Case* 1.b, and *Case* 1.c, respectively, and denote two probabilities in (2.18) as $p^{2a}_{i,\overline{S'}}$ and $p^{2b}_{i,\overline{S'}}$ for *Case*2.a and *Case*2.b, respectively. Among the m neighbors, the probability that b of them come from $\overline{S(t)}$ is that

$$p^{m,b}_{\overline{S'(t)}\cap S_V} = \begin{pmatrix} m \\ b \end{pmatrix} (p_{i,\overline{S'}}(l,r,v_{\max}))^b (1 - p_{i,\overline{S'}}(l,r,v_{\max}))^{m-b} \tag{2.19}$$

According to (2.19) and based on the assumption that nodes keep uniform distribution all the time, we can have

$$\begin{aligned}
&E_Q(\sum_{i\in S'(t), j\in\overline{S'(t)}} P_{ij}(t+1)) \\
&= \int_{-v_{\max}-r}^{v_{\max}} \sum_{m=1}^{n-1} \sum_{b=1}^{m} \left(\tfrac{1}{m} p_{i,m} b p_{m,b}\right) ndl \\
&= n \int_{-v_{\max}-r}^{v_{\max}} \left(1-(1-\pi r^2)^{n-1}\right) p_{i,\overline{S'}}(l,r,v_{\max}) dl \\
&\overset{(a)}{\approx} \begin{cases} n^2\pi r^2 \int_{-v_{\max}-r}^{v_{\max}} p_{i,\overline{S'}}(l,r,v_{\max}) dl, nr^2 = o(1), \\ n \int_{-v_{\max}-r}^{v_{\max}} p_{i,\overline{S'}}(l,r,v_{\max}) dl, nr^2 = \Omega(1), \end{cases}
\end{aligned} \tag{2.20}$$

where (a) is because $1-(1-\pi r^2)^{n-1} \approx \begin{cases} n\pi r^2, nr^2 = o(1), \\ 1, nr^2 = \Omega(1). \end{cases}$ for large n.

The simplification process is quite involved. In the interest of space, we move the detailed calculation to Appendix A. When $nr^2 = o(1)$, i.e., the disconnected case, we can obtain the final results as follows.

$$\Phi_m(Q) = \begin{cases} \Theta\left(nv_{\max}r^2 + nr^3 + \frac{nr^4}{v_{\max}}\right), \text{ for } Case\ 1, \\ \Theta\left(nv_{\max}r^2 + nr^3 + nv_{\max}^3\right), \text{ for } Case\ 2. \end{cases} \tag{2.21}$$

Since we assume $v_{\max} > r$ in *Case* 1, the general mobile conductance can be further simplified as $\Phi_m(Q) = \Theta(nv_{\max}r^2)$. When $v_{\max} < r$, i.e., *Case* 2, the conductance approximates $\Phi_m(Q) = \Theta(nr^3)$. Including the connected case, we have

$$\Phi_m^{VC}(Q) = \begin{cases} \Theta(nr^2 \max(v_{\max}, r)), nr^2 = o(1), \\ \Theta(\max(v_{\max}, r)), nr^2 = \Omega(1). \end{cases} \tag{2.22}$$

The $\Theta(nr^2)$ penalty factor is again observed.

2.8 Simulation Results

We conduct information spreading simulation on three networks with network size 600, $1,400$, and $2,000$. The simulation results for fully random, velocity constrained, one-dimensional and partially random mobility models are given in Figures 2.3, 2.4, 2.5, and 2.6, respectively. The x-axis of all figures represent the connectivity of the networks, while y-axis of all figures represent the time gap between fully connected networks and sparse networks. As we can see, the gap grows linearly with nr^2 in the disconnected zone and becomes saturated in the connected zone, confirming all of the theoretical results in the previous sections.

2.9 The Effects of Dynamism on Disconnectedness

Since we have obtained the mobile conductance within a more general framework, we may extract some useful observations and insights. One striking observation from the above study is that, even in a network with transmission range r far below the connectivity threshold, information can still be spread granted sufficient mobility. A question arises as to how much mobility is needed to facilitate information spreading given a certain degree of connectivity. Aided by our mobile conductance evaluation, some quantitative tradeoff between network dynamism and connectivity is revealed below, where we focus

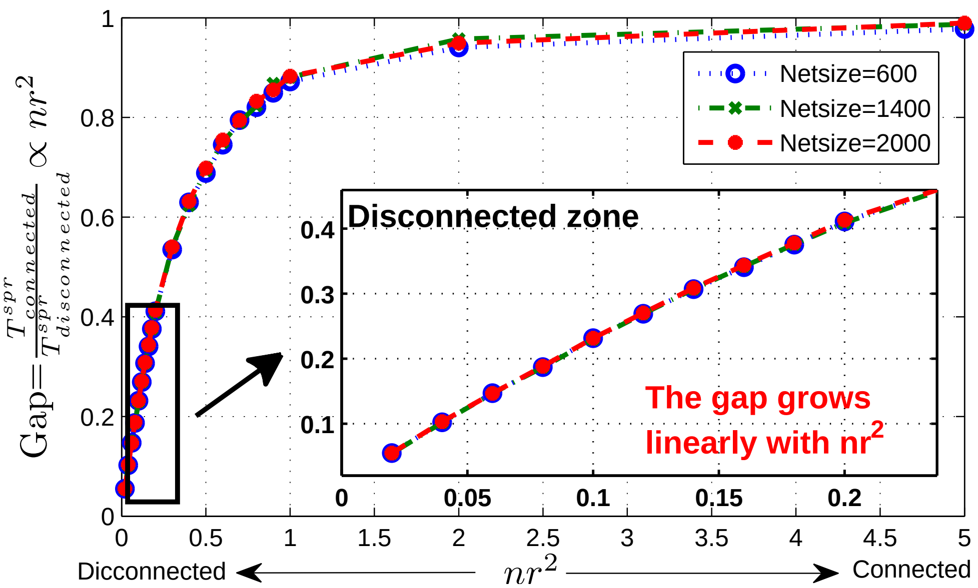

Figure 2.3: The gap vs. nr^2 in fully random mobility model

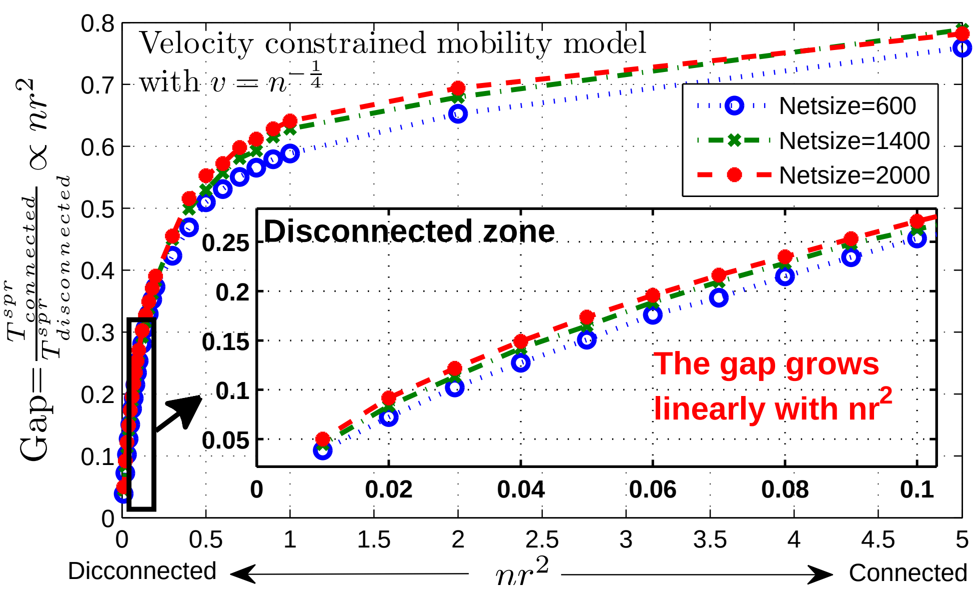

Figure 2.4: The gap vs. nr^2 in velocity constrained mobility model

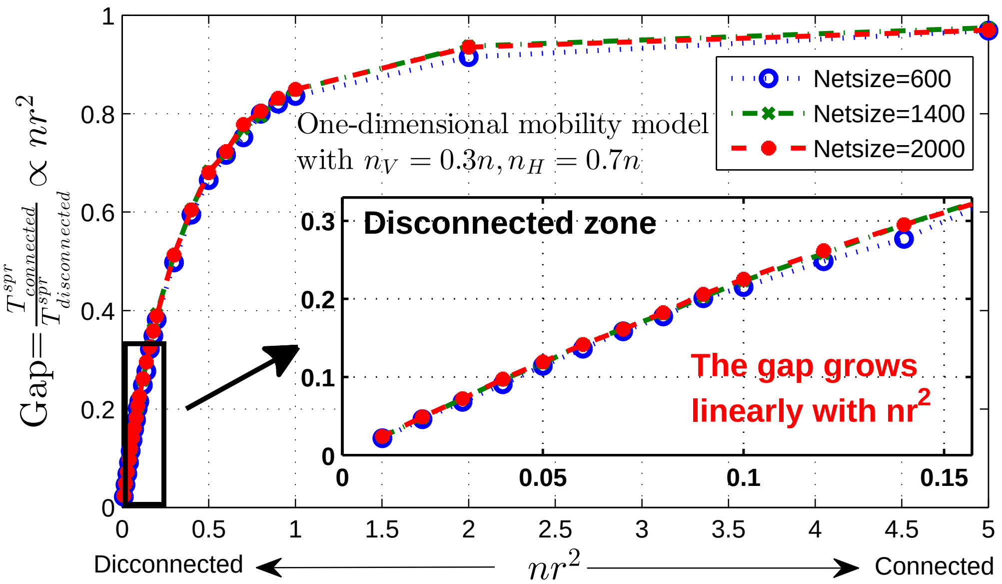

Figure 2.5: The gap vs. nr^2 in one-dimensional mobility model

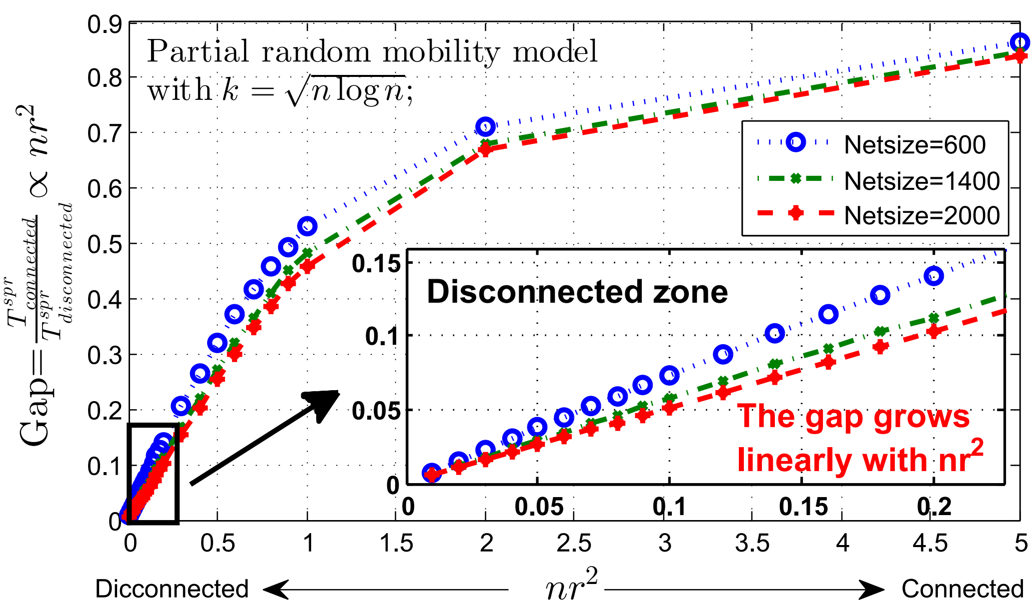

Figure 2.6: The gap vs. nr^2 in partially random mobility model

on a disconnected network with $nr^2 = o(1)$.

Here, the performance benchmark we consider is the information spreading time on a static ring graph. As indicated in, the ring graph is essentially the worst constrained graph in communications, and its gossip time, $\Theta(n \log n)$, can serve as an upper bound for distributed information spreading time. Note that other meaningful performance benchmarks may also be used in our following discussion.

2.9.1 Velocity Constrained Mobility

In this case, we are interested in determining how much velocity can make up for the deficiency in network connectivity due to a small transmission range. We mainly consider the $v_{\max} > r$ case. By comparing the results of (2.22) with the static conductance of a ring graph $\Theta(1/n)$, we can obtain the following velocity threshold for effective information spreading:

$$v_{\max}^{th} = \Omega\left(\frac{1}{n^2 r^2}\right). \tag{2.23}$$

Remarks: Since $v_{\max} = O(1)$ according to our model, a further analysis of (2.23) reveals the following interesting points. (1) When $r = \Omega(1/n)$, the above velocity is needed to compensate for the connectivity deficiency so that the same information spreading performance is achieved, as in a worst-case connected graph. In this case, there exists a tradeoff between $v_{\max}$ and r, in which the *effective velocity* is inversely proportional to the square of transmission range r. (2) When $r = O(1/n)$, even fully random mobility cannot recover the spreading time of $\Theta(n \log n)$. Nonetheless, information can still be spread to the whole network given sufficiently high velocity, only at a slower speed (see discussion in Fully Random Mobility Model).

2.9.2 Partially Random Mobility

Similarly, by comparing the results of (2.12) with the static conductance of a ring graph $\Theta(1/n)$ a necessary condition for achieving the same information spreading performance as in a worst-case connected graph can be obtained as

$$k(2n-k)r^2 = \Omega(1). \tag{2.24}$$

Remarks: A tradeoff exists between the mobility ratio k/n and r for effective message spreading. With the reasonable assumption that $k/n = o(1)$, the *effective mobility ratio* is given by

$$\left(\frac{k}{n}\right)^{th} = \Omega\left(\frac{1}{n^2r^2}\right) = \omega(1/n). \tag{2.25}$$

2.9.3 One-Dimensional Mobility

Following the same approach as above, for effective message spreading with this mobility model, a necessary condition is

$$n_V n_H r^2 = \Omega(1). \tag{2.26}$$

Remarks: A tradeoff exists between the mobility balance $\frac{n_V n_H}{n^2}$ and r for effective message spreading. Given r, the *effective mobility balance* is given by

$$\left(\frac{n_V n_H}{n^2}\right)^{th} = \Omega\left(\frac{1}{n^2r^2}\right). \tag{2.27}$$

Given that $n_V + n_H = n$, we may further infer that both n_V and n_H should, at least, remain growing with n when $r = o(1/\sqrt{n})$.

Single-piece message spreading with general transmission range is one extension of mobile conductance. Next, we will study another extension of mobile conductance to facilitate the analysis of multi-piece message spreading in mobile networks.

Chapter 3

Multi-piece message spreading

We consider the problem of multi-piece message dissemination through a natural randomized gossip algorithm in mobile connected networks and adopt the same system model, mobility model, and "Move-and-Gossip" paradigm in the previous section to facilitate the analysis.

3.1 Spreading Model

Different from the single-piece case, the message set in the multi-piece message dissemination is multi-dimensional. Let m_i denote the message of node i and $M = \{m_1, ..., m_n\}$ denote the set of all of these n messages. Let $S_i(t) \subset M$ denote the messages that node i has at the beginning of time t. Initially, $t = 0$ and $S_i(0) = \{m_i\}, \forall i$. Note that the nodes position $X_i(t)$ changes in the middle of each time slot (after the move step), while $S_i(t)$ is updated at the end (after the gossip step), according to the gossip algorithm. Specifically, node i contacts one of its neighbors, say node $j \in N_i(t+1)$ with probability $P_{ij}(t+1)$. Upon contact, node i sends an arbitrary message from $S_i(t) \setminus S_j(t)$ to node

j, if $S_i(t) \setminus S_j(t) \neq$; node j sends an arbitrary message from $S_j(t) \setminus S_i(t)$ to node i, if $S_j(t) \setminus S_i(t) \neq$. The metric of interest here is the ϵ-spreading time, defined as

$$T_{spr}^{all}(\epsilon) \triangleq \inf \left\{ t : \Pr \left(\bigcup_{i=1}^{n} \{S_i(t) \neq M\} \right) \leq \epsilon \right\}. \tag{3.1}$$

3.2 Mobile k-conductance

3.2.1 Preliminaries on Static Networks

In [19], the static k-conductance is defined as

$$\Phi_s^k = \min_{S \subset V, |S| \leq k} \left(\frac{\sum\limits_{i \in S, j \in \overline{S}}}{|S|} \right) = \min_{S \subset V, |S| \leq k} \left(\frac{P(r) N_S}{|S|} \right). \tag{3.2}$$

In (3.2), since we assume connected networks, $P_{ij} = P(r) = \Theta(\frac{1}{\log n})$ when $j \in N_i$, and N_S is the number of connecting edges between set S and its complement S. Note that N_S is a constant for a given set S in the static case, but becomes a random variable in the mobile case when the nodes in $S(t)$ and $\overline{S(t)}$ move at each time step.

Then, $\hat{\Phi}_s$, the harmonic conductance [19], is defined as the summation of the static k-conductance

$$\hat{\Phi}_s = \sum_{k=1}^{n-1} \Phi_s^k. \tag{3.3}$$

According to [19], the static spreading time for a multi-piece message case is

$$T_{static}^{all}(\epsilon) = O\left(\frac{\hat{\Phi}_s \log \epsilon^{-1}}{n} \right). \tag{3.4}$$

3.2.2 Mobile k-conductance and Mobile Spreading Time

As we mentioned above, the mobile networks are defined as Markovian evolving graphs, the expected values of $N_S(t+1)$; the number of contact pairs between $S(t)$ and $\overline{S(t)}$ after the moves is well defined, with respect to the transition matrix Q. Therefore, we can define the counterpart of mobile k-conductance as follows.

Definition: The mobile k-conductance of a stationary Markovian evolving graph with transition distribution Q is defined as:

$$\begin{array}{l} \Phi_m^k \overset{\Delta}{=} \min\limits_{|S'(t)| \le k} \left\{ E_Q \left(\frac{\sum\limits_{i \in S'(t), j \in \overline{S'(t)}} P_{ij}(t+1)}{|S'(t)|} \right) \right\} \\ \overset{(uniform)}{=} \min\limits_{|S'(t)| \le k} \left\{ \frac{P(n,r)}{|S'(t)|} E_Q \left[N_{S'}(t+1) \right] \right\} \end{array} . \tag{3.5}$$

Based on this definition, the multi-piece message spreading time admits the following theorem.

Theorem 3.1: For a mobile network with mobile harmonic conductance $\hat{\Phi}_m(Q)$, the multi-piece message spreading time scales as

$$T_{mobile,spr}^{all}(\epsilon, Q) = O\left(\frac{\hat{\Phi}_m(Q) \log \epsilon^{-1}}{n} \right). \tag{3.6}$$

Proof: We follow the proof strategy in , with suitable modifications from static to mobile networks. The message set at node i in the beginning of time t, denoted by $S_i(t)$, is a subset of M. Since M can have 2^n distinct subsets, effectively, each node can be in one of the 2^n distinct states, each corresponding to a distinct subset of M, at any time t. Thus, under the gossip algorithm, the overall network state $(S_1(t), ..., S_n(t))$ evolves over a state space of size $2^{n \times n}$. This is quite different from a single-piece message case where we only need to consider one monotonically increasing set.

To study this evolution over a high-dimensional space, the notion of 'type' is introduced. Specifically, two nodes, e.g., i and j, are called of the same 'type' at time t if $S_i(t) = S_j(t)$. At time t, this notion of 'type' defines a partition of all n nodes into disjoint 'type classes', with nodes having the same set of messages, i.e., nodes of the same 'type' belong to the same 'type class'. We denote the size of a maximal type class at time t by $A(t)$. When all nodes have received all messages, there is only one type class of size n. Therefore, we wish to find the following stopping time:

$$\inf\{t : |S_i(t)| = n, \forall i \in V\}. \tag{3.7}$$

The information spreads to all of the nodes when the overall dimension increases among all the nodes is $n(n-1)$. Then, we only need to study the overall dimension increase as the 'performance metric' of the algorithm. Define the dimension increase at time t, denoted by $D(t)$, as

$$D(t) = \sum_{i=1}^{n}(|S_i(t)| - 1) = (\sum_{i=1}^{n}|S_i(t)|) - n. \tag{3.8}$$

Therefore, $D(0) = 0$ and when all nodes have all messages $D(t) = n(n-1)$, we will bound $T^{all}_{mobile,spr}(\epsilon, Q)$ by bounding the time for $D(t)$ to become $n(n-1)$. We also define $L_k = \inf\{t : A(t) \geq k\}$ as the first time when any type class has at least k nodes. Similar to a single-piece case, where we study the lower bound of the expected set increase $|S(t+1)| - |S(t)|$ in each time slot, we study the lower bound of total dimension increase of all nodes by the end of gossip time $t+1$, denoted by $F(t+1) = D(t+1) - D(t)$. Let $I_{ij}(t+1)$ denote the indicator that node i dimension increases 1 by node i contacting node j. Thus, under the given mobile model Q, in the time interval $t \in [L_k, L_{k+1})$, and assuming there are b type classes, $C_1, ..., C_b$ at time t, we have

$$
\begin{aligned}
E[F(t+1)] &\geq E_Q\left(\sum_{i\in V(t+1)}\sum_{j>i}\left(E[I_{ij}(t+1)]+E[I_{ji}(t+1)]\right)\right)\\
&\geq E_Q\left(\sum_{i\in V(t+1)}\sum_{j>i:S_i(t+1)\neq S_j(t+1)} P_{ij}(t+1)\right)\\
&= E_Q\left(\frac{1}{2}\sum_{i,j\in V(t+1):S_i(t+1)\neq S_j(t+1)} P_{ij}(t+1)\right)\\
&= \frac{1}{2}E_Q\left(\sum_{l=1}^{b}\left(|C_l(t+1)|\frac{\sum_{i\in C_l(t+1),j\notin C_l(t+1)} P_{ij}(t+1)}{|C_l(t+1)|}\right)\right)\\
&= \frac{1}{2}E_Q\left(\sum_{l=1}^{b}\left(|C_l(t)|\frac{\sum_{i\in C_l(t+1),j\notin C_l(t+1)} P_{ij}(t+1)}{|C_l(t)|}\right)\right)\\
&\geq \frac{n\Phi_m^k(Q)}{2}.
\end{aligned}
\tag{3.9}
$$

This form is consistent with its counterpart in static networks. Therefore, we can follow the same lines in [19] to proceed with the rest of the proof we discovered.

3.3 Evaluation of Several Mobility Models

3.3.1 Static Network

Theorem 3.2: For a static random geometric graph, the static k-conductance scales as follows:

$$\Phi_s^k = \begin{cases} \frac{n-k}{n}, k \in [1, nr^2), \\ \Theta(\sqrt{\frac{n}{k}}r), k \in [nr^2, n/2), \\ \Theta(\frac{\sqrt{n(n-k)}}{k}r), k \in [n/2, n-nr^2), \\ \frac{n-k}{n}, k \in [n-nr^2, n-1]. \end{cases} \tag{3.10}$$

Therefore, the harmonic conductance scales as $\Theta(\frac{n^2}{r})$.

Proof: Recall the definition of k-conductance, we have

$$\Phi_s^k = \min_{S \subset V, |S| \leq k} \left(\frac{P(r)N_S}{|S|} \right). \tag{3.11}$$

In [21], it has been shown that the bottleneck segmentation is equivalent with the shortest straight line dividing S and $\overline{S}$. Therefore, the bottleneck segmentation for $|S| = k$ is the line with length $2\sqrt{\frac{k}{n}}$ in Figure 3.1, and it contains the minimum connecting edges between S and $\overline{S}$ for k-conductance, which should be

$$\begin{aligned} Cut_{\Phi_k}(S, \overline{S}) &\approx \textstyle\int_0^r \frac{1}{2} r^2 n \left[\arccos\left(\frac{x}{r}\right) - \sin\left(\arccos\left(\frac{x}{r}\right)\right)\right] n \sqrt{\frac{k}{n}} dx \\ &= \Theta(\sqrt{\frac{k}{n}} n^2 r^3). \end{aligned} \tag{3.12}$$

Thus, the conductance is

$$\Phi_s^k = \begin{cases} \frac{n-k}{n}, \quad k \in [1, nr^2), \\ \Theta(\sqrt{\frac{n}{k}}r), k \in [nr^2, n/2), \\ \Theta(\frac{\sqrt{n(n-k)}}{k}r), k \in [n/2, n-nr^2), \\ \frac{n-k}{n}, \quad k \in [n-nr^2, n-1]. \end{cases} \tag{3.13}$$

When $n/2 < k < n - nr^2$, due to symmetry, the minimum edge is the same for k and $n-k$; therefore, the k-conductance is $\frac{\sqrt{n(n-k)}}{k}r$.

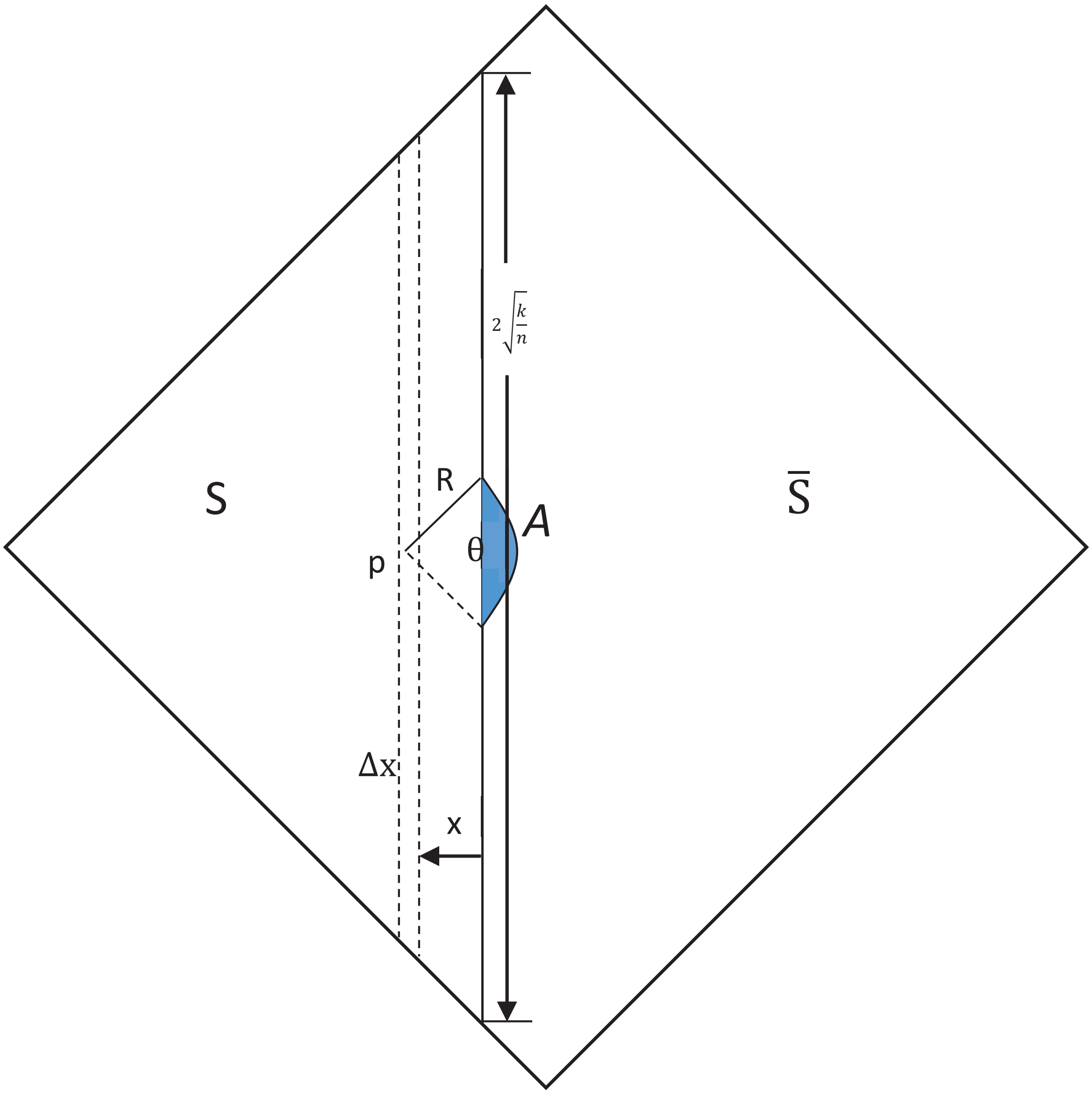

Figure 3.1: Approximation for k-conductance

When $1 \leq k < nr^2$ and $n - nr^2 < k \leq n-1$, all nodes in S or $\overline{S}$ can fully contact with S or $\overline{S}$, in this case, we can approximate the conductance by

$$\Phi_s^k = \frac{|S||\overline{S}|\pi r^2}{kn\pi r^2} = \frac{n-k}{n}. \tag{3.14}$$

After the summation, the harmonic conductance scales as, the harmonic conductance scales as $\hat{\Phi}_s = \sum_{k=1}^{n-1} \frac{k}{\Phi_s^k} \approx \Theta(\frac{n^2}{r})$.

3.3.2 Fully Random Mobility

Theorem 3.3: In fully random mobile networks, the k mobile conductance is

$$\Phi_m^k = 1 - \frac{k}{n}. \tag{3.15}$$

Thus, the mobile harmonic conductance scales as $\Theta(n^2 \log n)$, and corresponding mobile spreading time scales as $\Theta(n \log^2 n)$.

Proof: As mentioned before, the nodes are always uniformly distributed in full random mobility. Therefore, for each node in $S(t)$, the size of its neighborhood area is πr^2, and the expected number of contact pairs is

$$E_Q\left[N_{S'}(t+1)\right] = |S'(t)||\overline{S'(t)}|\pi r^2. \tag{3.16}$$

Noting that,

$$\Phi_m^k(Q) = \min_{|S'(t)| \leq k} \left\{ \frac{P(r)}{|S'(t)|} E_Q[N_{S'}(t+1)] \right\} = \min_{|S'(t)| \leq k} \left\{ \frac{\overline{S'(t)}}{n} \right\} = 1 - \frac{k}{n}. \tag{3.17}$$

The corresponding harmonic conductance is

$$\hat{\Phi}_m(Q) = \sum_{k=1}^{n-1} \frac{k}{\Phi_m^k(Q)} = \sum_{k=1}^{n-1} \frac{nk}{n-k} = \Theta(n^2 \log n). \tag{3.18}$$

Thus, the multi-message spreading time is $\Theta(n \log^2 n)$.

3.3.3 Partially Random Mobility

Theorem 3.4: For the partially random mobility model, where b out of n nodes are fully mobile and the rest $n-b$ nodes are static, the mobile k-conductance is $\Phi_m^k = \left(\frac{n-b}{n}\right)^2 \Phi_s^k + \frac{(n-k)(2n-b)b}{n^3}$, and the harmonic mobile conductance scales as

$$\hat{\Phi}_m = \begin{cases} O(\frac{n^2}{r}), b = o(\sqrt{n \log n}), \\ O(\frac{n^3 \log n}{b}), b = \Omega(\sqrt{n \log n}). \end{cases} \tag{3.19}$$

Proof: For each node that already has the message, e.g., i, among all its neighbors, on average, there are $(n-b)\pi r^2$ static nodes and $b\pi r^2$ mobile nodes. We denote the set of b mobile nodes at time t as $D(t)$, the set of $n-b$ static nodes at time t as $\overline{D(t)}$, and we also notice that the edges between static nodes remain unchanged after the move; therefore, $I_{ij}(t+1) = I_{ij}$ for $i \in S(t) \cap \overline{D(t)}$, $j \in \overline{S(t)} \cap \overline{D(t)}$. Since the k mobile nodes are fully random, the links involving the mobile nodes can be estimated in the same way as the fully random model; then, we have

$$E_Q\left[N_{S'}(t+1)\right] = E_Q\left[\sum_{\substack{i\in S'(t)\cap\overline{D(t)}\\ j\in\overline{S'(t)}\cap\overline{D(t)}}} I_{ij}\right] + \begin{pmatrix} \frac{n-k}{n}|\overline{S'(t)}|n\pi r^2\frac{|S'(t)|}{n}\frac{k}{n} \\ +\frac{n-k}{n}|S'(t)|n\pi r^2\frac{|\overline{S'(t)}|}{n}\frac{k}{n} \\ +\frac{k}{n}|S'(t)|n\pi r^2\frac{|\overline{S'(t)}|}{n}\frac{k}{n} \end{pmatrix}. \tag{3.20}$$

Then we can obtain the mobile k-conductance

$$\begin{aligned}\Phi_m^k = &\min_{S'(t)\subset V,|S'(t)|\le k}\left\{\frac{P(r)}{S'(t)}\sum_{\substack{i\in S'(t)\cap\overline{D(t)}\\ j\in\overline{S'(t)}\cap\overline{D(t)}}} I_{ij}\right\} \\ &+\min_{S'(t)\subset V,|S'(t)|\le k}\left\{2\frac{n-b}{n}\frac{\overline{S'(t)}}{n}\frac{b}{n}+\frac{b}{n}\frac{\overline{S'(t)}}{n}\frac{b}{n}\right\} \\ &=\left(\frac{n-b}{n}\right)^2\Phi_s^k+\frac{(n-k)(2n-b)b}{n^3}.\end{aligned} \tag{3.21}$$

Then, the mobile harmonic conductance becomes

$$\begin{aligned}\widehat{\Phi}_m(Q) &= \sum_{k=1}^{n-1}\frac{k}{\Phi_m^k} \\ &\approx \sum_{k=1}^{n/2}\frac{k}{\left(\frac{n-b}{n}\right)^2\sqrt{\frac{n}{k}}r+\frac{(n-k)(2n-b)b}{n^3}} \\ &+\sum_{n/2}^{n-1}\frac{k}{\left(\frac{n-b}{n}\right)^2\frac{\sqrt{n(n-k)}}{k}r+\frac{(n-k)(2n-b)b}{n^3}}.\end{aligned} \tag{3.22}$$

There is no closed form for the summation in (3.22); but, we can approximate it in order sense. The second term of (3.22) is always larger than the first term; therefore, we only need to calculate the second term. When $b = o(\sqrt{n\log n})$, $\left(\frac{n-b}{n}\right)^2\frac{\sqrt{n(n-k)}}{k}r >$

$\frac{(n-k)(2n-b)b}{n^3}$ in the denominator. In this case, results will be $O\left(\frac{n^2}{r}\right)$; similarly, when $b = \Omega(\sqrt{n \log n})$, the result will be $O\left(\frac{n^3 \log n}{b}\right)$.

Remarks: With the above results, we know that the number of mobility nodes b should achieve $\Omega\left(\sqrt{n \log n}\right)$ so as to obtain a significant performance gain over the static case. We also notice that when $k \to n$, $\hat{\Phi}_m(Q) \to \Theta\left(n^2 \log n\right)$, which is the fully random mobility harmonic conductance.

3.3.4 One-Dimensional Mobility

Theorem 3.5: For the one-dimensional area constrained mobility model, the mobile k-conductance $\Phi_m^k = \left(\frac{n_V^2+n_H^2}{n^2}\right)\Phi_s^k + \frac{2n_V n_H (n-k)}{n^3}$; the corresponding mobile harmonic conductance will be

$$\hat{\Phi}_m = \begin{cases} O(\frac{n^2}{r}), n_V = \Theta(n), n_H = o(\sqrt{n}), \\ O(\frac{n^2}{r}), n_H = \Theta(n), n_V = o(\sqrt{n}), \\ O(\frac{n^4 \log n}{n_V n_H}), \Theta(\sqrt{n}) < n_V, n_H < \Theta(n). \end{cases} \tag{3.23}$$

Proof: We follow the lines in [21] and adopt a divide-and-conquer strategy according to moving directions. Then, the mobile k-conductance for the one-dimensional mobility model is:

$$
\begin{aligned}
\Phi_m^k(Q) &= \left(\frac{n_V}{n}\right)^2 \Phi_s^k + \left(\frac{n_H}{n}\right)^2 \Phi_s^k \\
&\quad + \min_{S'(t)\subset V, |S'(t)|\le k} \left\{ \frac{P(r)}{|S'(t)|} \left(\begin{array}{c} p_{H-V}|S'(t)|\frac{n_H}{n}|\overline{S'(t)}|\frac{n_V}{n} \\ +p_{V-H}|S'(t)|\frac{n_V}{n}|\overline{S'(t)}|\frac{n_H}{n} \end{array} \right) \right\} \\
&= \frac{n_V^2+n_H^2}{n^2}\Phi_s^k + \min_{S'(t)\subset V, |S'(t)|\le k} \left\{ \frac{2n_V n_H |\overline{S'(t)}|}{n^3} \right\} \\
&= \frac{n_V^2+n_H^2}{n^2}\Phi_s^k + \frac{2n_V n_H (n-k)}{n^3}.
\end{aligned} \tag{3.24}
$$

where $p_{H-V} = p_{V-H} = \pi r^2$.

Then the mobile harmonic conductance should be

$$
\begin{aligned}
\widehat{\Phi}_m(Q) &= \sum_{k=1}^{n-1} \frac{k}{\Phi_m^k} \\
&\approx \sum_{k=1}^{n/2} \frac{k}{\frac{n_V^2+n_H^2}{n^2}\sqrt{\frac{n}{k}}r + \frac{2n_V n_H(n-k)}{n^3}} \\
&\quad + \sum_{n/2}^{n-1} \frac{k}{\frac{n_V^2+n_H^2}{n^2}\frac{\sqrt{n(n-k)}}{k}r + \frac{2n_V n_H(n-k)}{n^3}}.
\end{aligned} \tag{3.25}
$$

Thus, following the same analysis method in (3.22), we know that when $n_V = \Theta(n), n_H = o(\sqrt{n})$ or $n_H = \Theta(n), n_V = o(\sqrt{n})$, the mobile harmonic conductance will be $O\left(\frac{n^2}{r}\right)$; when $\Theta(\sqrt{n}) < n_V, n_H < \Theta(n)$, the result will be $O\left(\frac{n^4 \log n}{n_V n_H}\right)$.

Remarks: We can see that, when all of the nodes move in the same direction, the result is the same as the static case. When half of the nodes are moving vertically, and half of the nodes are moving horizontally, the mobile harmonic conductance achieves its minimum of $\Theta(n^2 \log n)$, which is the same order as a fully random mobility model.

3.3.5 Velocity Constrained Model

Theorem 3.6: In the velocity constrained model, the mobile k-conductance scales as

$$\Phi_m^k = \begin{cases} \frac{n-k}{n}, & k \in [1, n(r+v_{\max})^2), \\ \Theta(\sqrt{\frac{n}{k}} \max_(r, v_{\max})), & k \in [n(r+v_{\max})^2, n/2), \\ \Theta(\frac{\sqrt{n(n-k)}}{k} \max(r, v_{\max})), & k \in [n/2, n-n(r+v_{\max})^2), \\ \frac{n-k}{n}, & k \in [n-n(r+v_{\max})^2, n-1]. \end{cases} \tag{3.26}$$

The corresponding mobile harmonic conductance will be

$$\hat{\Phi}_m = \begin{cases} O(\frac{n^2}{\max(r, v_{\max})}), v_{\max} = o(1), \\ O(n^2 \log n), v_{\max} = \Theta(1). \end{cases} \tag{3.27}$$

Proof: For connected mobile networks, since nodes are always uniformly distributed in the unit square Ω, we can view the velocity of a node as a transmission amplifier. Therefore, for a velocity constrained model with $v_{\max}$, it is equivalent to a static network with transmission range of $v_{\max} + r$. Therefore, we can use the result in Theorem 3.2, and the corresponding mobile k-conductance for a velocity constrained model with $v_{\max}$ should be

$$\Phi_m^k = \begin{cases} \frac{n-k}{n}, & k \in [1, n(r+v_{\max})^2), \\ \Theta(\sqrt{\frac{n}{k}} \max_(r, v_{\max})), & k \in [n(r+v_{\max})^2, n/2), \\ \Theta(\frac{\sqrt{n(n-k)}}{k} \max(r, v_{\max})), & k \in [n/2, n-n(r+v_{\max})^2), \\ \frac{n-k}{n}, & k \in [n-n(r+v_{\max})^2, n-1]. \end{cases} \tag{3.28}$$

Thus, the corresponding mobile harmonic conductance will be

$$
\begin{aligned}
\widehat{\Phi}_m(Q) &= \sum_{k=1}^{n-1} \frac{k}{\Phi_m^k(Q)} \\
&= \sum_{k=1}^{n(r+v_{\max})^2} \frac{kn}{n-k} + \sum_{n-n(r+v_{\max})^2}^{n-1} \frac{kn}{n-k} \\
&+ \sum_{n(r+v_{\max})^2}^{n/2} \sqrt{\frac{n}{k}}(r+v_{\max}) + \sum_{n/2}^{n-n(r+v_{\max})^2} \frac{\sqrt{n(n-k)}}{k}(r+v_{\max}).
\end{aligned} \tag{3.29}
$$

When $v_{\max} = o(1)$, the summation will be dominated by the last two terms. Therefore, the result will be $\Theta\left(\frac{n^2}{r}\right)$; while, if $v_{\max} = \Theta(1)$, the summation will be dominated by the first two terms, and the result will be $\Theta(n^2 \log n)$.

Remarks: Therefore, when $v_{\max} > r$, the multi-piece message spreading can gain significant benefits from node velocity. When node velocity is comparable to 1 (tends to be the same as fully random mobility), the information spreading time will have the same order as a fully random mobility model.

Chapter 4

Information spreading with PA-gossip in social networks

4.1 Problem Formulation

In this work, we have some preliminary results of information spreading in a classic graph theoretical model for social networks with a non-uniform gossip algorithm. Specifically, we are interested in the *preferential attachment* (PA) model originally proposed by Barabasi and Albert [2], which builds on the paradigm that new vertices attach to existing vertices with probabilities proportional to their degrees. It has been shown that this model indeed enjoys many properties observed in social networks (e.g., a power law distribution of the vertex degrees and a small diameter).

4.2 System Model

For an undirected graph G with the degree of a vertex v, denoted by $deg_v(G)$, the PA graph is defined as follows [3][10].

Definition: (Preferential Attachment Graph). *Let $m \geq 2$ be a fixed parameter. The random graph G_m^n is an undirected graph on the vertex set $V := 1, ..., n$, inductively defined as follows:*

- *G_m^1 consists of a single vertex with m self loops.*
- *For all $n > 1$, G_m^n is built from G_m^{n-1} by adding new node n together with m edges $e_n^1 = n, v_1, ..., e_n^m, v_m$ inserted one after the other in this order. Let $G_{m,i-1}^n$ denote the graph right before the edge e_n^i is added. Let $M = \sum_{v \in V} deg_{G_{m,i-1}^n}(v)$ be the sum of the degrees of all of the nodes in $G_{m,i-1}^n$. The end point v_i is selected randomly, such that $v_i = u$ with probability $deg_{G_{m,i-1}^n}(u)/(M_i + 1)$ except for itself that is selected with probability $deg_{G_{m,i-1}^n}(n) + 1/(M_i + 1)$.*

This definition implies that when a new edge e_n^i is inserted, the vertex v_i will be chosen with a probability proportional to its degree (except for $v_i = n$). Since many real-world social networks are conjectured to evolve using similar principles, the PA-model can serve as a model for social networks. Another property observed in many real-world networks has been formally proven for preferential attachment graphs, namely, that the degree distribution follows a power-law.

It can be seen that, for $m = 1$, the graph is disconnected with high probability; so, we focus on the case $m \geq 2$. Under this assumption, [3] shows that the diameter of PA-model graphs is only $\Theta\left(\frac{\log n}{\log \log n}\right)$ with high probability.

4.3 Spreading Model

For a classic gossip algorithm, at each time slot, each node will randomly pick one of its neighbors to exchange information with push-pull strategy. Specifically, in each round, one node, e.g., node i, will select one of its neighbors, e.g., node j, to push its information; and, then, node i will also pull information from node j. Such a gossip algorithm is also called uniform gossip. In [10], the single-piece information spreading time with uniform gossip in PA-models is bounded by $O(\log n)$. Later, it adds uniform gossip with memory M, which means that, in each round, each node will randomly pick one of its neighbors except for neighbors which have already been picked in the previous M rounds. Thus, it proves that, for gossip-with-memory, the single-piece information spreading time is $\Theta\left(\frac{\log n}{\log \log n}\right)$.

In addition to the fact that gossip-with-memory can improve information spreading time in order sense, we discovered two other interesting facts about gossip-with-memory in PA-models, as follows: (1) The larger the parameters m, the smaller the effects brought by adding memory into a gossip algorithm. (2) In [10], increasing the memory has only minimal influence on the total information spreading time. Therefore, we studied whether there are other possible gossip models to help facilitate information spreading in PA-models. Results indicate that preferential attachment behavior not only exists in the graph building process but also influences the way in which we obtain information. Specifically, in real life, people tend to gather information from famous people. Inspired by this human behavior, we propose the following PA-gossip algorithm.

- At the beginning of information spreading, each node i gathers the degree information $deg_G(N_i(1)), ...deg_G(N_i(m))$ from all of its neighbors N_i.
- Let $D_i = \sum_{k=1}^{m} N_i(k)$. In the middle of each time slot, every node i picks its infor-

mation exchange target $N_i(u)$ from its neighbors $N_i(1), ..., N_i(m)$ with probability $deg_G(N_i(u))/D_i$.

- Every node i exchanges information with its transmission target according to Push-Pull strategy.

where $m = deg_G(i)$ is the number of node i's neighbor.

We will see in the simulation that PA-gossip has a significant advantage to both uniform gossip and gossip-with-memory. In the next section, we provide some preliminary theoretical explanations for fast information spreading with PA-gossip in PA graphs.

4.4 Theoretical Intuition for PA-gossip

First, we will present some proven properties of PA graphs in order to explain the advantage of PA-gossip in PA graphs.

4.4.1 Alternative Model

In the random process generating G_m^n, the random decisions made at each step depend heavily on the previous random decisions. To solve this problem, in [3], an alternative way of generating G_m^n that is more accessible is discussed. We, first, describe the model for $m = 1$ and, then, generalize it to arbitrary m. We refer to [3] for proof that both models are equivalent.

Let (x_i, y_i) for $i \in [n] : 1, 2, ..., n$ be n independently and uniformly chosen pairs from $[0, 1] \times [0, 1]$. With probability one, all of these numbers are distinct. By reordering with each pair, we assume that $x_i < y_i$ for every $i \in [n]$. Suppose that, after relabelling,

$y_1 < y_2 < ... < y_n$. We set $W_0 := 0$ and $W_i := y_i$ for $i \in [n]$. The graph G_1^n is now defined by having an edge (i, j) if, and only if, $W_{j-1} < x_i < W_j$. Define $\omega_j := W_j - W_{j-1}$.

Similarly, for G_m^n, we sample mn pairs $(x_{i,j}, y_{i,j})$ independently and uniformly from $[0, 1] \times [0, 1]$ with $x_{i,j} < y_{i,j}$ for $i \in [n]$ and $j \in [m]$. We relabel the variables such that $y_{i,j}$ is increasing in the following order:

$$y_{1,1} < y_{1,2} < ... < y_{1,m} < y_{2,1} < ... < y_{n,1} < ... < y_{n,m}. \tag{4.1}$$

We set $W_0 := 0$ and $W_i := y_{i,m}$ for $i \in [n]$. The graph is now defined by having an edge (i, j) for each $k \in [m]$ such that $W_{j-1} < x_{i,k} < W_j$. As in the previous paragraph, define $\omega_j = W_j - W_{j-1}$. We write $l_{i,k}$ for the node j such that $W_{j-1} < x_{i,k} < W_j$.

Note that given $y_{1,1}, ..., y_{n,m}$, the random variables $x_{1,1}, ..., x_{n,m}$ are independent with $x_{i,k}$ being chosen uniformly from $[0, y_{i,k}]$. For readability, we, instead, assume that if $y_{1,1}, ..., y_{n,m}$ are given, then each $x_{i,k}$ is chosen independently and uniformly from $[0, W_i]$. Let $s = 2^a$ be the smallest power of 2 larger than $\log^7 n$, and let 2^b be the largest power of 2 smaller than 2n/3. Let $I_t = [2^t + 1, 2^{t+1}]$.

With all of the definitions and alternative models described above, we can illustrate some properties of G_m^n.

4.4.2 Special Properties of PA graph G_m^n

Theorem 4.1: ([3]) *Let $m \geq 1$ be fixed, Let*

$$\alpha_m(d) := \frac{2m(m+1)}{(d+m)(d+m+1)(d+m+2)}, \tag{4.2}$$

and let $\epsilon > 0$ be fixed. Then, with probability $1 - o(1)$, we have

$$(1-\epsilon)\alpha_m(d) \le \frac{\#_m^n(d)}{n} \le (1+\epsilon)\alpha_m(d), \tag{4.3}$$

for every d in the range $0 \le d \le n^{1/5}$.

Remarks: It can now be seen that the node degree distribution of PA-models obeys the power-law. Therefore, most nodes in a PA graph will have comparatively small degrees while only small nodes in the graph will have comparatively large degrees.

Lemma 4.1: ([3]). *Let $m \ge 2$ be fixed. For graph G_m^n, each of the following holds with probability $1 - o(1)$.*

- $E_1 := \{|W_i - \sqrt{i/n}| \le \frac{1}{10}\sqrt{i/n}$ *for all* $i \in [s, n]\}$,
- $E_2 := \{|\{i \in I_t | \omega_i \ge 1/(10\sqrt{in})\} \ge 2^{t-1}|$ *for all* $t \in [a, b)\}$,
- $E_3 := \{\omega_1 \ge \frac{4}{\log n\sqrt{n}}\}$,
- $E_4 := \{\omega_i \ge \frac{\log^2 n}{n}$ *for all* $i < n^{1/5}\}$,
- $E_5 := \{\omega_i \le n^{-4/5}$ *for all* $i > n/(\log n)^5\}$.

Remarks: It is proven in [3] that all five properties above hold for G_m^n with high probability. Precisely speaking, let $0 < W_1 < ... < W_n < 1$ be distinct real numbers and let $\omega_i = W_i - W_{i-1}$. Assume that $W_1, ..., W_n$ satisfy the properties $E_1, ..., E_5$. A typical social network $G_m(W_1, ..., W_n)$ is obtained by connecting each node i with the nodes $l_{i,1}, ..., l_{i,m}$, where each $l_{i,k}$ is a node chosen randomly with $P[l_{i,k} = j] = \omega_j / W_i$ for all $j \le i$.

In [3], a node i is *useful* if $i \le n/(\log n)^5$ and $\omega_i \ge (\log n)^2/n$. Nodes without *useful* property are called *non-useful.* Then, we have the following properties.

Lemma 4.2: ([10]). *Assuming* $E_1, ..., E_5$, *the following events hold with prob.* $1 - o(1)$

- $E_6 := \{deg_G v \geq 2mn^{2/5} \log^2 n \textit{ for all useful } v\}$,
- $E_7 := \{deg_G v \leq 5m \log^2 n \textit{ for all non-useful } v\}$,
- $E_8 := \{G \textit{ contains } (\log n)^{O(\log^{3/4} n)} \textit{ non-useful cycles of length at most } \log^{3/4} n\}$.

Remarks: From the above lemma, we can see that a *useful* node is equivalent to a node with a large degree; while a *non-useful* node corresponds to one with a small degree. Therefore, most nodes in the graph are *non-useful*, and there are many cycles in a *non-useful* node.

Lemma 4.3: ([3]). *Assuming* $E_1, ..., E_5$, *with prob.* $1 - o(1)$ *every node* v *(non-useful) of* G *is joined by a (descending) path of length at most* $8 \log \log n$ *to a useful node.*

Lemma 4.4: ([3]). *Assuming* $E_1, ..., E_5$, *let* $\epsilon > 0$ *be fixed, and let* v, $1 \leq v \leq n$, *be a useful node. With prob.* $1 - o(1)$, *there is a path in* G *between* v *and* 1 *of length at most* $(1/2 + \epsilon) \log n / \log \log n$.

Remarks: From all of the above properties, we can see that most nodes in the graph are *non-useful* and that there are some long cycles (length $log^{3/4}n$) in *non-useful* nodes. With a unform gossip algorithm, it is with high probability that a message will be trapped in *non-useful* nodes. However, for PA-gossip, it will force a direction from smaller degree nodes to larger degree ones, thereby avoiding the possible cycles in *non-useful* nodes, greatly decreasing the time required for information to be spread to *useful* nodes and then disseminated to the whole graph. Besides, due to the push-pull strategy, when information reaches to a large degree node, spreading efficiency will be significantly higher than when it stays in small degree node. Actually, the small-to-large information direction brought by PA-gossip can effectively utilize scale-free property (specifically, the large-hub-like structure). Therefore, we may also consider that it will be effective in other scale-free network models, even in real-life Internet and social networks.

4.5 Simulation Results

We simulated information spreading with three different gossip algorithms to illustrate the advantage of PA-gossip in spreading information in PA-models, and results confirm our conjectures. We simulated networks with up to 10^6 nodes. The curves in Figure 4.1, from the bottom up, correspond to uniform gossip in a complete graph, uniform gossip in a PA graph, gossip-with-memory in a PA graph, and, finally, PA-Gossip in a PA graph. As the figure shows, improvement produced by gossip-with-memory is very limited; even in a 10^6 nodes network, it is only 0.5 step faster than uniform gossip. For PA-gossip, improvement is about 5 steps faster. Also, if we compare the curves horizontally, in 11 steps, uniform gossip can only spread information to 3×10^3 nodes, and gossip-with-memory can spread information to 6×10^3 nodes; however, with PA-gossip, information can be spread to 1×10^6 nodes in 11 steps. Therefore, information propagation has been strongly facilitated in PA-models with PA-gossip.

To further verify the advantage of PA-gossip, we investigated its performance on the Slashdot social network data set from [1] (with a network size around $70,000$ and an average node degree of 11.5 after removal of some erroneous data) in Figure 4.2. Realizing that generally fast information spreading may partially obscure the advantage of the PA network structure and our PA-gossip algorithm, in this figure, we demonstrate the average number of informed nodes during the propagation process. In particular, with seven rounds of information spreading, the original message can reach about $2,000$ users with uniform gossip on the complete graph, about $10,000$ users with uniform gossip on

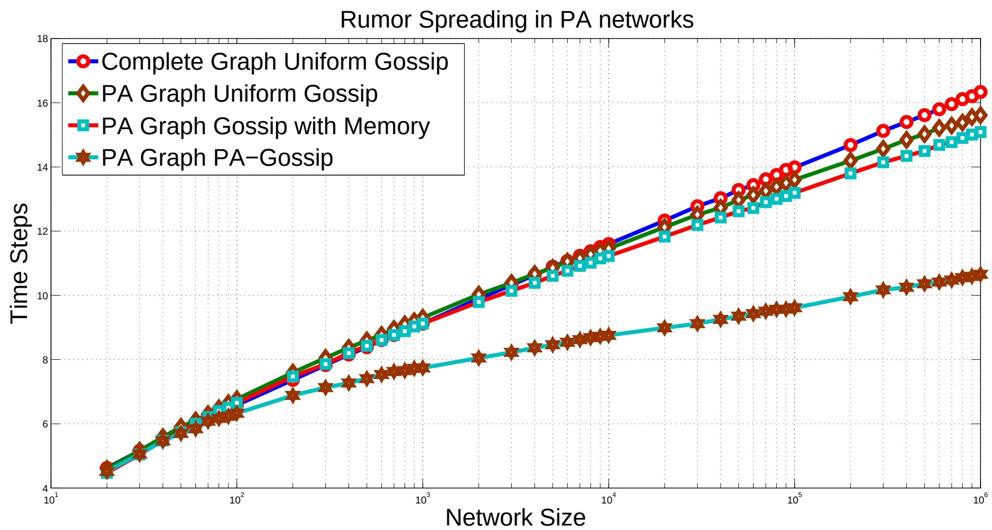

Figure 4.1: Average information spreading time for different gossip algorithms over PA networks ($m = 10$)

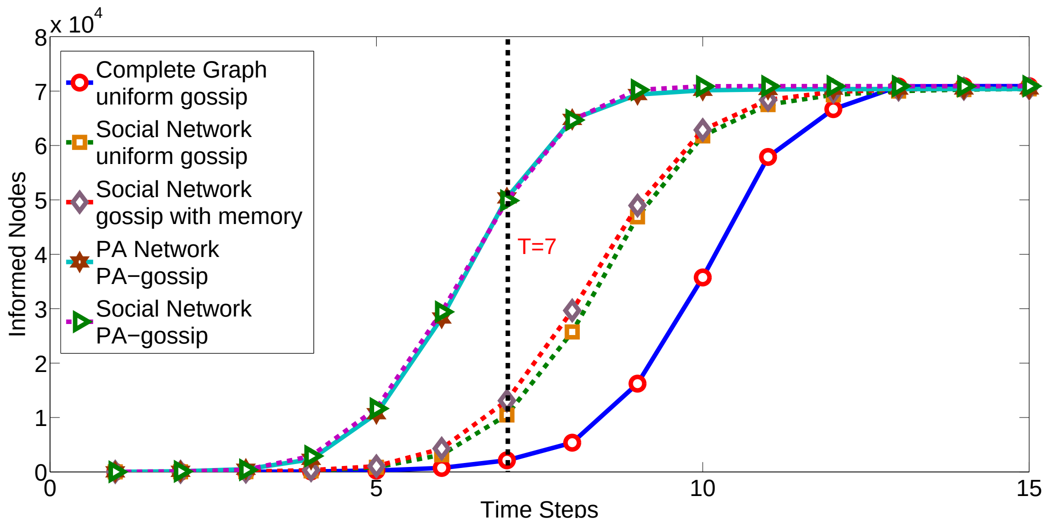

Figure 4.2: Average number of informed nodes over time for different gossip algorithms over a social network

the social graph, and about $50,000$ users with PA-gossip on the social graph. The result for PA-gossip on a PA graph with $m = 5$ (average degree of 10) is also known, which agrees well with that for the social graph, indicating that the PA model works well with this social network. As revealed by our study results, our PA-gossip algorithms have great potential.

Currently, we have not completed theoretical proof for this significant improvement brought by PA-gossip in PA-models, which we will study in future work. Also, the scale-free property (large-hub-like structure) generally exists in most Internet and social topologies, and it can be effectively used by PA-gossip. Therefore, we also plan to explore PA-gossip performance in other social network models.

Chapter 5

Conclusion

We studied the information spreading in large-scale network models from two network layers: technological and social networks. In the technological networks layer, we focused on the mobile networks. We extended mobile conductance in two directions. First, we modified mobile conductance and its derivation process to get the new definition of general mobile conductance, which led to studying single-piece information spreading in mobile networks with general transmission range. We observed a $\Theta(nr^2)$ penalty factor caused by transmission range deficiency (i.e., $nr^2 = o(1)$). Finally, with general mobile conductance, we analyzed its resulting capability in compensating transmission deficiency for information spreading.

When analyzing the multi-piece information spreading in mobile networks, we combined the static k-conductance in [19] and mobile conductance together to create a new metric mobile k-conductance. Then, we rated the mobile k-conductance in different mobility models and analyzed the mobility patterns effect on mobile k-conductance. Finally, we evaluated the multi-piece information spreading time with respect to the new metric for these models.

In social networks, with the knowledge of a PA models special properties and reputable human behavior (i.e., preferential attachment), we created an effective information spreading algorithm, PA-gossip. We provided some preliminary explanations for the effectiveness of PA-gossip in PA-models and confirmed our hypothesis with a large-size network simulation.

REFERENCES

[1] "Stanford Large Network Dataset Collection". http://snap.stanford.edu/data.

[2] A. Barabási and R. Albert. "Emergence of scaling in random networks". *science*, 286(5439):509–512, 1999.

[3] B. Bollobás and O. Riordan. "The diameter of a scale-free random graph". *Combinatorica*, 24(1):5–34, 2004.

[4] K. Chen, M. Chiang, and V. Poor. "From technological networks to social networks". *accepted and to be published in IEEE J. Sel. Areas Commun*, 31(9):548–572, 2013.

[5] F. Chierichetti, S. Lattanzi, and A. Panconesi. "Almost tight bounds for rumour spreading with conductance". In *Proceedings of the 42nd ACM symposium on Theory of computing*, pages 399–408. ACM, 2010.

[6] F. Chierichetti, S. Lattanzi, and A. Panconesi. "Rumor spreading in social networks". *Theoretical Computer Science*, 412(24):2602–2610, 2011.

[7] Y. Chou, H. Huang, and R. Cheng. "Modeling information dissemination in generalized social networks". *Communications Letters, IEEE*, 17(7):1356–1359, 2013.

[8] A. Demers, D. Greene, C. Hauser, W. Irish, J. Larson, S. Shenker, H. Sturgis, D. Swinehart, and D. Terry. "Epidemic algorithms for replicated database maintenance". In *Proceedings of the sixth annual ACM Symposium on Principles of distributed computing*, pages 1–12. ACM, 1987.

[9] S. Diggavi, M. Grossglauser, and D. Tse. “Even one-dimensional mobility increases the capacity of wireless networks”. *Information Theory, IEEE Transactions on*, 51(11):3947–3954, 2005.

[10] B. Doerr, M. Fouz, and T. Friedrich. “Social networks spread rumors in sublogarithmic time”. In *Proceedings of the 43rd annual ACM symposium on Theory of computing*, pages 21–30. ACM, 2011.

[11] A. El Gamal, J. Mammen, B. Prabhakar, and D. Shah. “Optimal throughput-delay scaling in wireless networks-part I: The fluid model”. *Information Theory, IEEE Transactions on*, 52(6):2568–2592, 2006.

[12] E. Frias-Martinez, G. Williamson, and V. Frias-Martinez. “An agent-based model of epidemic spread using human mobility and social network information”. In *Privacy, security, risk and trust (PASSAT), 2011 IEEE third international conference on and 2011 IEEE third international conference on social computing (SocialCom)*, pages 57–64. IEEE, 2011.

[13] M. Grossglauser and D. Tse. “Mobility increases the capacity of ad-hoc wireless networks”. In *INFOCOM 2001. Twentieth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*, volume 3, pages 1360–1369. IEEE, 2001.

[14] P. Jacquet, B. Mans, and G. Rodolakis. “Information propagation speed in mobile and delay tolerant networks”. *Information Theory, IEEE Transactions on*, 56(10):5001–5015, 2010.

[15] J. Kleinberg. “The convergence of social and technological networks”. *Communications of the ACM*, 51(11):66–72, 2008.

[16] Z. Kong and E. Yeh. "On the latency for information dissemination in mobile wireless networks". In *Proceedings of the 9th ACM international symposium on Mobile ad hoc networking and computing*, pages 139–148. ACM, 2008.

[17] J. Mammen and D. Shah. "Throughput and delay in random wireless networks with restricted mobility". *Information Theory, IEEE Transactions on*, 53(3):1108–1116, 2007.

[18] D. Mosk-Aoyama and D. Shah. "Fast distributed algorithms for computing separable functions". *Information Theory, IEEE Transactions on*, 54(7):2997–3007, 2008.

[19] D. Shah. "Gossip Algorithms". *Foundations and Trends® in Networking*, 3(1):1–125, 2009.

[20] H. Zhang, H. Dai, Z. Zhang, and Y. Huang. "Mobile conductance under general transmission radius r". Technical report, Department of Electrical Engineering, NC State University, 2013, available at `http://www4.ncsu.edu/~hdai/InformationSpreading-HZ-TP2.pdf`.

[21] H. Zhang, Z. Zhang, and H. Dai. "Gossip-based Information Spreading in Mobile Networks". *Wireless Communications, IEEE Transactions on*, 11(12):5918–5928, 2013.

APPENDIX

Appendix A

Simplification of General Mobile Conductance

Although too complicated to accurately evaluate, we may get relatively tight results in the order sense. The key lies in calculating $\int_{-v_{\max}-r}^{v_{\max}} p_{i,\bar{S}'}(l,r,v_{\max})dl$, which may be further expanded by replacing $p_{i,\overline{S'}}(l,r,v_{\max})$ with the expression obtained in (2.16) and (2.18),

$$\int_{-v_{\max}-r}^{v_{\max}} p_{i,\bar{S}'}(l,r,v_{\max})dl = \begin{cases} \int_{-v_{\max}-r}^{-v_{\max}+r} p_{i,\bar{S}'}^{1a}dl + \int_{-v_{\max}+r}^{v_{\max}-r} p_{i,\bar{S}'}^{1b}dl + \int_{v_{\max}-r}^{v_{\max}} p_{i,\bar{S}'}^{1c}dl, & \text{for } Case\ 1, \\ \int_{-v_{\max}-r}^{v_{\max}-r} p_{i,\bar{S}'}^{2a}dl + \int_{v_{\max}-r}^{v_{\max}} p_{i,\bar{S}'}^{2b}dl, & \text{for } Case\ 2. \end{cases} \tag{A.1}$$

Applying a similar approximation method in , we replace $\rho'_{\overline{S}}(\cdot)$ by

$$\rho'_{\overline{S}}(x) \approx \frac{x+v_{\max}}{2v_{\max}}n. \tag{A.2}$$

Note that this approximation does not change the result in order form. Thus, by inputting the approximation into $p_{i,\overline{S'}}(l,r,v_{\max})$, we, first, consider case 1, i.e., $v_{\max} > r$, as follows

$$
\begin{aligned}
p^{1a}_{i,\overline{S'}} &= \frac{1}{n\pi r^2}\int_{-v_{\max}}^{l+r} n\left(\frac{x+v_{\max}}{2v_{\max}}\right)\left(2\sqrt{r^2-(x-l)^2}\right)dx \\
&= \frac{1}{v_{\max}\pi r^2}\int_{-v_{\max}}^{l+r} x\left(\sqrt{r^2-(x-l)^2}\right)dx + \frac{1}{\pi r^2}\int_{-v_{\max}}^{l+r}\left(\sqrt{r^2-(x-l)^2}\right)dx \\
&\overset{x'=x-l}{=} \frac{1}{v_{\max}\pi r^2}\left(\int_{-v_{\max}-l}^{r} x\left(\sqrt{r^2-x^2}\right)dx + \int_{-v_{\max}-l}^{r} l\left(\sqrt{r^2-x^2}\right)dx\right) \\
&\quad + \frac{1}{\pi r^2}\int_{-v_{\max}-l}^{r}\left(\sqrt{r^2-x^2}\right)dx \\
&= \frac{1}{v_{\max}\pi r^2}\left(\frac{2}{3}\left(r^2-(v_{\max}+l)^2\right)^{\frac{3}{2}}\right) + \frac{(v_{\max}+l)^2}{v_{\max}\pi r^2}\sqrt{r^2-(v_{\max}+l)^2} + \Theta(r^2)
\end{aligned}
\tag{A.3}
$$

We can see that $p^{1a}_{i,\overline{S'}} = O(\frac{r}{v_{\max}}+1) = O(1)$. Thus, when $nr^2 = o(1)$, the first term in (A.1) may be expressed as

$$
\begin{aligned}
\int\limits_{-v_{\max}-r}^{-v_{\max}+r} f(p^{1a}_{i,\overline{S'}})dl &= \Theta\left(\frac{1}{\pi r^2}\int\limits_{-v_{\max}-r}^{-v_{\max}+r}\left(\begin{array}{c}\frac{1}{v_{\max}}\left(\frac{2}{3}\left(r^2-(v_{\max}+l)^2\right)^{\frac{3}{2}}\right)+ \\ \frac{(l+v_{\max})^2}{v_{\max}}\sqrt{r^2-(v_{\max}+l)^2}+\Theta(r)\end{array}\right)dl\right) \\
&= \Theta\left(\frac{1}{\pi r^2}\int\limits_{-r}^{r}\left(\frac{2}{v_{\max}}\left(\frac{2}{3}\left(r^2-l^2\right)^{\frac{3}{2}}\right)+\frac{2l^2}{v_{\max}}\sqrt{r^2-l^2}\right)dl\right) \\
&= \Theta\left(\frac{r^2}{v_{\max}}+r\right)
\end{aligned}
\tag{A.4}
$$

Following the same method, when $nr^2 = o(1)$, we may get the simplification of $p^{1b}_{i,\overline{S'}}$ and $p^{1c}_{i,\overline{S'}}$ as

$$
\begin{aligned}
p^{1b}_{i,\bar{S}'} &= \frac{\pi}{v_{\max}} \left(l + v_{\max} \right), \\
p^{1c}_{i,\bar{S}'} &= \frac{2}{3 v_{\max} r^2} \left(r^2 - (v_{\max} - l)^2 \right)^{\frac{3}{2}} \\
&\quad + \frac{l^2 - v_{\max}^2}{v_{\max} r^2} \sqrt{r^2 - (v_{\max} - l)^2} + \Theta\left(1 + \frac{v_{\max} - l}{r} \right),
\end{aligned} \tag{A.5}
$$

respectively, where the last term in (A.5) is because $\alpha_l \in [0, 2\pi)$ and, thus, $\sin \alpha_l \in [-1, 1]$

Finally, we can get the general mobile conductance for *Case* 1 as

$$
\begin{aligned}
\Phi_m(Q) &= \frac{2}{n} E_Q \left(\sum_{i \in S'(t), j \in \overline{S'(t)}} P_{ij}(t+1) \right) \\
&= \Theta\left(n v_{\max} r^2 + n r^3 + n \frac{r^4}{v_{\max}} \right), \text{ for } n r^2 = o(1).
\end{aligned} \tag{A.6}
$$

As for *Case* 2, i.e., $v_{\max} < r$, we can derive $p^{2a}_{i,\bar{S}'}$ and $p^{2b}_{i,\bar{S}'}$ similarly as

$$
\begin{aligned}
p^{2a}_{i,\bar{S}'} &= \frac{2}{v_{\max} r^2} \left(r^2 - (v_{\max} + l)^2 \right)^{\frac{3}{2}} + \frac{2 (v_{\max} + l)^2}{v_{\max} r^2} \sqrt{r^2 - (v_{\max} + l)^2} + \Theta\left(1 \right), \\
p^{1c}_{i,\bar{S}'} &= \Theta\left(\frac{v_{\max} + l}{r} \right).
\end{aligned} \tag{A.7}
$$

The general mobile conductance for *Case* 2 can be similarly obtained as

$$
\Phi_m(Q) = \Theta\left(n r^3 + n v_{\max} r^2 + n v_{\max}^3 \right), \text{ for } n r^2 = o(1). \tag{A.8}
$$