國 立 交 通 大 學

應用數學系

碩　士　論　文

搜尋散佈謠言者的數學模型

A Mathematical Model for Finding the Culprit Who Spreads Rumors

研 究 生：李姿慧

指導老師：傅恆霖　教授

中 華 民 國 一 〇 一 年 六 月

搜尋散佈謠言者的數學模型

A Mathematical Model for Finding the Culprit Who Spreads Rumors

研究生：李姿慧 Student: Zi-Hui Lee
指導教授：傅恆霖 教授 Advisor: Hung-Lin Fu

國 立 交 通 大 學

應 用 數 學 系

碩 士 論 文

A Thesis
Submitted to Department of Applied Mathematics
College of Science
National Chiao Tung University
in Partial Fulfillment of the Requirements
for the Degree of
Master
in
Applied Mathematics
June 2012
Hsinchu, Taiwan, Republic of China

中 華 民 國 一 〇 一 年 六 月

搜尋散佈謠言者的數學模型

研究生：李姿慧　　　　　　　指導教授：傅恆霖 教授

國 立 交 通 大 學

應 用 數 學 系

摘要

在這篇論文中，我們介紹謠言傳播模型，它的設計是根據一個在流行病學領域著名的易感–感染模型。我們描述在一個圖上散佈謠言的源頭的最大概似估計值並計算預測到散佈謠言的源頭的機率。我們發現：對於路徑的圖形，機率會隨著時間增加趨近到0，其關係為$t^{-1/2}$；對於正則樹，機率有一個明確的範圍。當$d = 3$，其機率值會隨著時間增加趨近到$1/4$，此結果已利用隨機圖模型得到 [6]。

中 華 民 國 一 〇 一 年 六 月

A Mathematical Model for Finding the Culprit Who Spreads Rumors

Student: Zi-Hui Lee Advisor: Hung-Lin Fu

Department of Applied Mathematics
National Chiao Tung University

June, 2012

Abstract

In this thesis, we introduce a rumor spreading model based on the common susceptible-infected (SI) model which is a well known epidemiological model. We describe the maximum likelihood estimators of graphs and we evaluate the detection probabilities of finding the rumor source in d-regular trees. We observe that: For paths, the detection probability of finding the rumor source scales as $t^{-1/2}$, which approaches 0 as t approaches infinity. For regular trees, we find an explicit bound of the detection probabilities of finding the source in d-regular trees. As a consequence, for $d = 3$, the detection probability approaches $1/4$, this result has been obtained earlier by using a random graph model [6].

誌謝

在交大的兩年，謝謝你們。

謝謝我的指導教授傅恆霖老師，謝謝老師讓我成為他的學生，跟著老師的一年半我學到了許多，尤其是態度，老師總是期許我們要多讀點書，多運動等等，所有的一切我會謹記在心。謝謝與我同師門的學長姐：惠蘭學姐、貓頭學長、明輝學長、惠娟學姐、施智懷學長、敏筠學姐、軒軒學長、小泡學姐還有瑩晏學姐。謝謝你們在這兩年給我的支持、建議與鼓勵。因為過程中有你們的幫助我才能夠還算順利的完成我的碩士論文。

謝謝系上的老師們指導我完成學業，讓我順利的走到這。特別感謝的還有：謝謝資訊技術服務中心的蔡錫鈞主任提供我解決問題的方法與建議；謝謝系上的資訊人員詹宗智還有麻將學長教我使用系上的機器做運算；謝謝游森朋教授抽空看我的問題；謝謝陳秋媛老師的提點，讓我的論文內容更加嚴謹。

謝謝我的好朋友們：米奇、小雨、憶妏、沙發、惠雯、吉米、連、DK、嗨派，有事情就幫助我，有玩的就找我。還要謝謝所有我在這裡認識的人，因為有你們我的碩士生活才能如此多采多姿。

最後一定要謝謝我的父母，因為有他們的期許我才一路走到。在外地念書不能常常回家，謝謝家人的體諒以及時不時打電話來關心問候，我愛我的爸爸媽媽。

Contents

List of Figures

1 Introduction

Social network, internet and electrical power grid network are the common networks everywhere in our life. We are surrounded by all kinds of networks and are very easily subjected to the influence of network risks. Although every network has its different structure, the common phenomenon is: an isolated risk is enlarged because it can be spread by the network.

For example, in an electrical power grid network, an isolated failure can lead to a rolling blackout. Computer viruses utilize the internet to infect millions of computers everyday. The malicious rumors or misinformation can be spread in the social networks quickly and the person concerned will be deeply hurt and offended. In all of these situations, power network operator, internet service provider or victim of a malicious rumor would like to infer the source of risks as quickly as possible and then blockade the spread of risks. All of these situations can be modeled as rumors spreading through networks, where the goal is to find the source in order to control and prevent these network risks based on limited information about the network structure and the rumor infected vertices.

Prior work on rumor spreading has primarily focused on infectious diseases in populations. The standard model of infectious diseases is known as the *susceptible-infected-recovered* (SIR) model [1]. In this model, there are three types of vertices: (i) susceptible vertices that are capable of being infected; (ii) infected vertices that can spread the virus further; and (iii) recovered vertices that are cured and can't be infected anymore. Research in this model has focused on the structure of the network and rates of infection/cure [2, 3, 4, 5]. However, there has no idea of identifying the source of an epidemic. Now, a mathematical model has been developed to identify the rumor source in a network based on rumor infected vertices [6]. But, not much is known if the network is getting more complicate. In this thesis, we set forth to study the networks defined on d-regular trees.

1.1 Preliminaries

A **graph** G is a triple consisting of a **vertex set** $V(G)$, and **edge set** $E(G)$, and a relation that associates with each edge two vertices called its **endpoints**. The **order** of a graph G is the number of vertices in G and the **size** of a graph G is the number of edges in G. A **loop** is an edge whose endpoints are equal, and **multiple edges** are edges having the same pair of endpoints. A **simple graph** G is a graph having no loops or multiple edges. We consider a simple graph with a **countably infinite** vertex set.

A **subgraph** of graph G is a graph H such that $V(H) \subseteq V(G)$ and $E(H) \subseteq E(G)$ and the assignment of endpoints to edges in H is the same as in G. We then write $H \subseteq G$ and say that "G **contains** H". A **path** is a simple graph whose vertices can be ordered so that two vertices are adjacent if and only if they are consecutive in the list. A graph G is **connected** if each pair of vertices in G belongs to a path.

In a graph G, the **contraction** of edge e with endpoints u, v is the replacement of u and v with a single vertex whose incident edges are the edges other than e that were incident to u or v. The resulting graph $G \cdot e$ has one less edge e than G, and the new vertex $\{u, v\}$(or $\{v, u\}$).

When u and v are the endpoints of an edge e and writing $e = uv$ (or

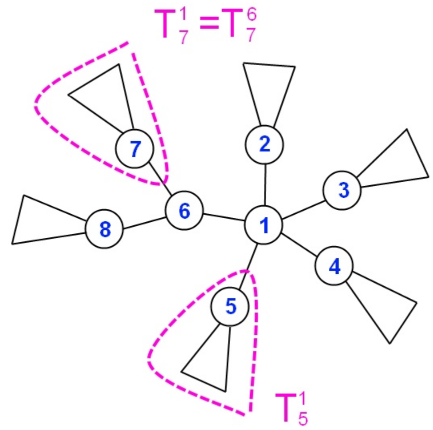

Figure 1: Illustration of subtree T_u^v

$e = vu$), they are **adjacent** and are **neighbors**. We write $u \leftrightarrow v$ for "u is adjacent to v". The **neighborhood** of v, written $N_G(v)$ or $N(v)$, is the set of vertices adjacent to v. The **degree** of vertex v in a graph G, written $d_G(v)$ or $d(v)$, is the number of edges incident to v. And G is **regular** if the degrees of all vertices are the same. It is k**-regular** if the common degree is k.

A graph with no cycles is acyclic. A tree is a connected acyclic graph. In this thesis, A d**-regular tree** is a tree where every vertex has d neighbors. Therefore, it is an infinite graph. A **rooted tree** T_r is a tree with one vertex r chosen as **root**. For each vertex v, Let $p(v)$ be the unique path from r to v. The parent of v is its neighbor on $p(v)$; its **children** are its other neighbors. The set $child(v)$ is the set of all children of v.

A **branch** of a tree is a subtree T_v^r of T_r induced by v and all its descendants. In addition, t_v^r denotes the number of vertices in T_v^r. A d-regular tree T can be decomposed into d subtrees $(T_{v_1}^v, T_{v_2}^v, \cdots T_{v_d}^v)$, such that these trees are branches of T_v and $\sum_{u\in N(v)} t_u^v = t_v^v - 1$. A simple example T is shown in Figure 1. T_7^1 is a branch of T_1; T_7^6 is a branch of T_6. Clearly, $T_7^1 = T_7^6$. And T can be decomposed into 6 subtrees $(T_2^1, T_3^1, T_4^1, T_5^1, T_6^1)$, such that these trees are branches of T_1.

These definitions about graph theory are cited from the book "Introduction to Graph Theory" written by Douglas B. West. If this part is not sufficient, please refer to the reference [8] for details.

1.2 Rumor spreading (RS) model

We consider a discrete time susceptible-infected (SI) model. Let $S(t)$ be a set of people who don't know rumors yet at time t and let $I(t)$ be a set of people who have known rumors at time t in which the number of people of $S(t)$ and $I(t)$ are denoted S_t and I_t, respectively. Using a fixed population, $S_t + I_t = N$. For convenience of study, we shall assume

$$\begin{cases} S_{t+1} = S_t - 1, \\ I_{t+1} = I_t + 1, \\ S_0 = N,\ I_0 = 0. \end{cases}$$

Let each person be a vertex and the relationship between two people be an edge, and these form the vertex set and edge set of a graph G. Let G_t be a subgraph of order t of G. This graph is compose of t infected vertices which are people who have known rumors at time t. The graph G_1 is a vertex which is called rumor source and in each discrete time-step $t+1$, $t > 0$, G_{t+1} develops from G_t by adding a vertex z with an edge. We assume that every vertex is chosen with the following probability distribution:

$$\mathbf{P}_{t+1}(z) = \frac{1}{\sum_{v\in V(G_t)} d(v) - 2(t-1)} \text{ for } v \in V(G_t). \tag{1}$$

2 Rumor Source Estimator

Consider a network G and a subgraph G_n of G that is a graph with n infected vertices. The *Maximum Likelihood(ML) estimator* is the vertex v which has the maximum of $\mathbf{P}(G_n|v)$. where $\mathbf{P}(G_n|v)$ is the probability of observing G_n under the RS model if v is the source.

2.1 The ML estimator of G_n in a regular tree

Continue from the paragraph above, if we want to know that what's the ML estimator of G_n, we would like to evaluate $\mathbf{P}(G_n|v)$ for all $v \in G_n$ and then choose v so that $\mathbf{P}(G_n|v)$ is maximum. In general, evaluating every vertex's $\mathbf{P}(G_n|v)$ is complicated. Let us consider a simple example as shown in Figure 2 with $n = 4$. First, we suppose that the vertex 1 is the source, and we would like to calculate $\mathbf{P}(G_4|1)$. Then there are six ways or vertex infection orders that a rumor can be spread to every vertex in G_4 with vertex 1 as the source: $(1, 2, 3, 4)$, $(1, 2, 4, 3)$, $(1, 3, 2, 4)$, $(1, 3, 4, 2)$, $(1, 4, 2, 3)$ and $(1, 4, 3, 2)$. However, if we suppose that the vertex 2 is the source, infection order $(2, 3, 1, 4)$ is not possible, since $2 \not\leftrightarrow 3$ (2 and 3 are not adjacent in G_4). Therefore, in general to evaluate $\mathbf{P}(G_n|v)$, we need to find all such *possible n-permutations of* $V(G_n)$ and their corresponding probabilities.

Let $S(v, G_n)$ be the set of all possible n-permutations of $V(G_n)$ starting with vertex v where $v \in V(G_n)$, and $\sigma = (v_1 = v, v_2, \cdots, v_n)$ for each $\sigma \in S(v, G_n)$. The probability $\mathbf{P}(\sigma|v)$ is the probability of observing G_n under the RS model if v is the source in the infected order σ. Let $G_k(\sigma)$ be the set of k vertices which are in the subgraph of G_n with order $(v_1 = v, v_2, \cdots, v_k)$

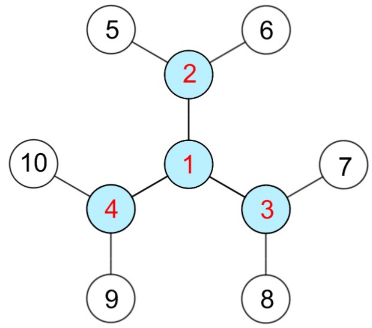

Figure 2: Network with 4 infected vertices.

for $1 \leq k \leq n$. Then we have

$$\mathbf{P}(\sigma|v) = \prod_{k=1}^{n-1} \frac{1}{\sum_{v_i \in G_k(\sigma)} d(v_i) - 2(k-1)}. \tag{2}$$

For d-regular trees,

$$\mathbf{P}(\sigma|v) = \prod_{k=1}^{n-1} \frac{1}{dk - 2(k-1)}. \tag{3}$$

From equation (3), we can see that every permutation σ has the same probability and is independent of the source. Specifically, for any source v and permutation σ, $\mathbf{P}(\sigma|v)$ is a constant. Thus

$$\begin{aligned}\mathbf{P}(G_n|v) &= \sum_{\sigma \in S(v,G_n)} \mathbf{P}(\sigma|v) \\ &= |S(v,G_n)| \cdot \prod_{k=1}^{n-1} \frac{1}{dk - 2(k-1)} \\ &\propto |S(v,G_n)|.\end{aligned}$$

It follows that $\mathbf{P}(G_n|v)$ is proportional to $|S(v,G_n)|$. Let $R(v,G_n)$ be the number of distinct ways to spread a rumor to every vertex in G_n with v as the source. Clearly, $|S(v,G_n)| = R(v,G_n)$. Then

$$\mathbf{P}(G_n|v) \propto R(v,G_n). \tag{4}$$

In summary, the ML estimator of G_n in a regular tree can be obtained by finding the maximum of $R(v, G_n)$ for all v. For the above example as shown in Figure 2 the estimation of vertices are

$$
\begin{aligned}
\mathbf{P}(G_4|1) &= \sum_{\sigma \in S(1,G_4)} \mathbf{P}(\sigma|1) \\
&= \mathbf{P}((1,2,3,4)|1) + \mathbf{P}((1,2,4,3)|1) \\
&+ \mathbf{P}((1,3,2,4)|1) + \mathbf{P}((1,3,4,2)|1) \\
&+ \mathbf{P}((1,4,2,3)|1) + \mathbf{P}((1,4,3,2)|1) \\
&= 6 \cdot \left(\frac{1}{3} \cdot \frac{1}{4} \cdot \frac{1}{5}\right) = \frac{1}{10}.
\end{aligned}
$$

$$
\begin{aligned}
\mathbf{P}(G_4|2) &= \sum_{\sigma \in S(2,G_4)} \mathbf{P}(\sigma|2) \\
&= \mathbf{P}((2,1,3,4)|2) + \mathbf{P}((2,1,4,3)|2) \\
&= 2 \cdot \left(\frac{1}{3} \cdot \frac{1}{4} \cdot \frac{1}{5}\right) = \frac{1}{30}.
\end{aligned}
$$

$$
\begin{aligned}
\mathbf{P}(G_4|3) &= \sum_{\sigma \in S(3,G_4)} \mathbf{P}(\sigma|3) \\
&= \mathbf{P}((3,1,2,4)|3) + \mathbf{P}((3,1,4,2)|3) \\
&= 2 \cdot \left(\frac{1}{3} \cdot \frac{1}{4} \cdot \frac{1}{5}\right) = \frac{1}{30}.
\end{aligned}
$$

$$
\begin{aligned}
\mathbf{P}(G_4|4) &= \sum_{\sigma \in S(4,G_4)} \mathbf{P}(\sigma|4) \\
&= \mathbf{P}((4,1,2,3)|4) + \mathbf{P}((4,1,3,2)|4) \\
&= 2 \cdot \left(\frac{1}{3} \cdot \frac{1}{4} \cdot \frac{1}{5}\right) = \frac{1}{30}.
\end{aligned}
$$

Clearly, we can now evaluate $R(v, G_4)$ for all v. $R(1, G_4) = 6$, $R(2, G_4) = R(3, G_4) = R(4, G_4) = 2$. Thus vertex 1 is the probable rumor source in this example.

2.2 Rumor centrality

In this paragraph, we shall explain how to calculate the rumor centrality for each vertex in a given graph G_n. Moreover, the rumor center of G_n is a vertex with the maximum of rumor centrality.

Let G_n be a rooted tree with root v and assume v has a rumor. The next infected vertex must be one of the children of v. For all $u \in child(v)$, there are $R(u, T_u^v)$ ways to spread a rumor in the branch T_u^v with u as the source. Thus

$$R(v, G_n) = (n-1)! \prod_{u \in child(v)} \frac{R(u, T_u^v)}{t_u^v!}. \tag{5}$$

To understand the above expression, the number of ways to permute $n-1$ steps from different subtrees is a permutation of the multiset $\{t_{v_1}^v \cdot 1, t_{v_2}^v \cdot 2, \cdots t_{v_d}^v \cdot d\}$. If we continue this recursion (5) until we reach the leaves of the tree, we obtain

$$\begin{aligned} R(v, G_n) &= (n-1)! \prod_{u \in child(v)} \frac{R(u, T_u^v)}{t_u^v!} \\ &= (n-1)! \prod_{u \in child(v)} \left(\frac{(t_u^v - 1)!}{t_u^v!} \cdot \prod_{w \in child(u)} \frac{R(w, T_w^v)}{t_w^v!} \right) \\ &= (n-1)! \prod_{u \in child(v)} \frac{1}{t_u^v} \cdot \prod_{w \in child(u)} \frac{R(w, T_w^v)}{t_w^v!} \\ &= (n-1)! \prod_{u \in G_n - v} \frac{1}{t_u^v}. \end{aligned}$$

Since a leaf vertex l has one vertex so $R(l, T_l^v) = 1$. By the fact that $t_v^v = n$, then

$$R(v, G_n) = n! \prod_{u \in G_n} \frac{1}{t_u^v}. \tag{6}$$

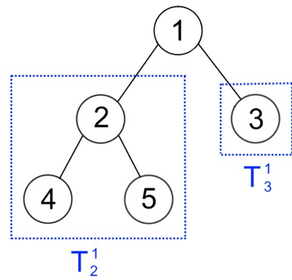

Figure 3: Network of calculating rumor centrality.

Let us consider an example as shown in Figure 3.

$$R(1, G_5) = \frac{5!}{5 \cdot 3 \cdot 1 \cdot 1 \cdot 1} = 8.$$

Indeed, there are 8 possible n-permutations of the network in Figure 4 with vertex 1 as the source. They are

$$(1, 2, 4, 5, 3), (1, 2, 4, 3, 5), (1, 2, 3, 4, 5), (1, 3, 2, 4, 5)$$
$$(1, 2, 5, 4, 3), (1, 2, 5, 3, 4), (1, 2, 3, 5, 4), (1, 3, 2, 5, 4)$$

In order to find the rumor center of a given graph G_n, we have to find the rumor centrality of every vertex in G_n respectively. In fact, the rumor centrality of v can be deduced from rumor centrality of its neighbors. To this end, consider two adjacent vertices u and v in G_n, we have

$$t_u^v = n - t_v^u. \tag{7}$$

And $t_w^v = t_w^u$ for each $w \in G_n - \{u, v\}$. Thus,

$$\frac{R(u, G_n)}{R(v, G_n)} = \frac{t_u^v}{n - t_u^v}. \tag{8}$$

For example in Figure 4, we have

$$\frac{R(2, G_5)}{R(1, G_5)} = \frac{12}{8} = \frac{t_2^1}{5 - t_2^1} = \frac{3}{2}.$$

It shows that for each vertex u, we can calculate its rumor centrality by using its neighbor's rumor centrality and t_u^v.

The following is an important property of the rumor center:

Theorem 2.2.1. *[6] Given an n vertices tree, vertex v is the rumor center if and only if*

$$t_u^v \leq \frac{n}{2}$$

for all $u \neq v$. Furthermore, a tree can have at most 2 rumor centers.

It plays a crucial role in establishing our main results.

3 Main Result

In this section, we shall explain the behavior of the detection probabilities of finding the rumor source in different graphs. Let $E_t(G)$ be the event of correct rumor source detection under the ML rumor source estimator after time t on a graph G. If the graph G considered is prescribed, then we use E_t to denote $E_t(G)$. Then $\mathbf{P}(E_t)$ is the correct detection probability of finding the rumor source in a given graph G.

3.1 Detection probabilities of d-regular trees, $d \leq 3$

In this paragraph, we shall present shorter proofs by using combinatorial methods than the original proofs in reference [6]. We first consider the detection probability of finding the source in a 2-regular tree. By (4), rumor center is the ML estimator of G_n where G_n is a subtree of a d-regular tree.

Theorem 3.1.1. *Suppose a rumor has spread in a 2-regular tree. Then we have that*

$$\mathbf{P}(E_t) = O\left(\frac{1}{\sqrt{t}}\right).$$

Proof:

Consider a path which compose of $2n$ (respectively $2n+1$) infected vertices. By Theorem (2.2.1) and (6), v is a unique rumor center in G_{2n+1} and we have

$$R\left(v, G_{2n+1}\right) = \binom{2n}{n},$$

and there are two vertices v, v' both are rumor centers in G_{2n} then we have

$$R\left(v, G_{2n}\right) = R\left(v', G_{2n}\right) = \binom{2n-1}{n-1}.$$

Every graph G_n in a 2-regular tree is a path. The correct detection probability of finding the rumor source in a path is

$$\mathbf{P}(\text{source} = v | G_n) = \frac{\mathbf{P}(G_n | \text{source} = v)\mathbf{P}(\text{source} = v)}{\sum_{i \in G_n} \mathbf{P}(G_n | \text{source} = i)\mathbf{P}(\text{source} = i)}.$$

The probabilities of vertices which are rumor sources in a d-regular tree are equal possible. Thus

$$\mathbf{P}(\text{source} = v | G_n) = \frac{\mathbf{P}(G_n | \text{source} = v)}{\sum_{i \in G_n} \mathbf{P}(G_n | \text{source} = i)}.$$

By (4),

$$\mathbf{P}(\text{source} = v | G_n) = \frac{R(v, G_n)}{\sum_{i \in G_n} R(i, G_n)} \tag{9}$$

Hence

$$\begin{aligned}\mathbf{P}(\text{source} = v | G_{2n+1}) &= \frac{\binom{2n}{n}}{2\left(\binom{2n}{0} + \binom{2n}{1} + \cdots + \binom{2n}{n-1}\right) + \binom{2n}{n}} \\ &= \binom{2n}{n} \cdot \frac{1}{2^{2n}}, \text{ and}\end{aligned}$$

$$\begin{aligned}\mathbf{P}(\text{source} = v | G_{2n}) &= \frac{\binom{2n-1}{n-1}}{2\left(\binom{2n-1}{0} + \binom{2n-1}{1} + \cdots + \binom{2n-1}{n-1}\right)} \\ &= \binom{2n-1}{n-1} \cdot \frac{1}{2^{2n-1}} = \binom{2n}{n} \cdot \frac{1}{2^{2n}}.\end{aligned}$$

By Stirling's formula, $\binom{2n}{n} \cdot \frac{1}{2^{2n}} \sim \frac{\sqrt{4\pi n}\left(\frac{2n}{e}\right)^{2n}}{2^{2n}\sqrt{2\pi n}\left(\frac{n}{e}\right)^{n}\sqrt{2\pi n}\left(\frac{n}{e}\right)^{n}} = \frac{1}{\sqrt{\pi n}}.$

Thus

$$\mathrm{P}(E_t) = \begin{cases} \mathbf{P}(\text{source} = v \big| \, G_t) = \sqrt{\dfrac{2}{\pi t}} & \text{if } t \text{ is } \textit{even}, \\ \dfrac{1}{2}\mathbf{P}(\text{source} = v \big| \, G_t) + \dfrac{1}{2}\mathbf{P}(\text{source} = v' \big| \, G_t) = \sqrt{\dfrac{2}{\pi(t-1)}} & \text{if } t \text{ is } \textit{odd}. \end{cases} \tag{10}$$

It follows that the path detection probability scales as $t^{-1/2}$, which approaches 0 as t approaches infinity. ■

Now, we consider the detection probability of finding the source in a d-regular tree, $d \geq 3$.

Theorem 3.1.2. *Suppose a rumor has spread in a regular tree. Then we have that*

$$0 < \mathbf{P}(E_t) \leq \frac{1}{2}.$$

Proof:

Consider a graph G_n in a regular tree. We can regard the number of ways to spread a rumor to every vertex in G_n with v as the source as the sum of ways that rumor can be spread from v through u, where u is the neighbor of v. Hence

$$R(v, G_n) = \sum_{u \in N(v)} R(\{v, u\}, G_n \cdot vu). \tag{11}$$

Given any two adjacent vertices u and v, we have

$$R(u, G_n) = \sum_{w \in N(u)} R(\{u, w\}, G_n \cdot uw) \text{ and}$$
$$R(v, G_n) = \sum_{u \in N(v)} R(\{v, u\}, G_n \cdot vu).$$

Since u is adjacent to v, there is at least one common term in above two sums on the RHS. Thus we have

$$\begin{aligned} R(v, G_n) &\leq \sum_{i \in G_n, i \neq v} R(i, G_n) \\ \Rightarrow R(v, G_n) &\leq \frac{1}{2} \sum_{i \in G_n} R(i, G_n). \end{aligned} \tag{12}$$

Since $R(v, G_n)$ is positive, we have

$$0 < \frac{R(v, G_n)}{\sum\limits_{i \in G_n} R(i, G_n)} \leq \frac{1}{2}.$$

This concludes that for every graph which composes of infected vertices, the detection probability is greater than 0 and less than 1/2. ■

For d-regular trees with $d > 2$, Theorem (3.1.2) states that the event that positive detection probability happens is independent of the order of the graph. In what follows, we shall evaluate the explicit detection probability of finding the source in a d-regular tree.

Our goal is to calculate the detection probability of finding the rumor source. Recall the Theorem (2.2.1), v is the unique rumor center if and only if $t_u^v < \frac{n}{2}$ for all $u \neq v$. So we consider the number of vertices of branches of T_v, and these branches are rooted trees with root u, $u \in N(v)$. Without loss of generality, we assume $t_u^v \geq 1$ for all $u \in N(v)$. Let A_d and B_d be two sets such that

$$A_d = \{(a_1, a_2, \cdots, a_d) |\, 1 \leq a_i < \frac{n}{2}, \sum_{i=1}^{d} a_i = n - 1\}, \text{ and}$$

$$B_d = \{(b_1, b_2, \cdots, b_d) |\, b_i \in \mathbb{N}, \sum_{i=1}^{d} b_i = n - 1\}.$$

Clearly, $A_d \subseteq B_d$. Moreover,

$$|B_d| = \binom{n-1-d+d-1}{d-1} = \binom{n-2}{d-1}.$$

Now, let $S_i = \{(x_1, x_2, \cdots, x_d) \in B_d : x_i \geq \frac{n}{2}\}$. By principle of Inclusion and Exclusion,

$$|A_d| = |B_d| - \big|\bigcup_{i=1}^{d} S_i\big| = |B_d| - \sum_{i=1}^{d} |S_i|. \tag{13}$$

Let $x_j = y_j + 1, j \neq i$ and $x_i = \lceil \frac{n}{2} \rceil + y_i$. $\sum_{i=1}^{d} y_i = (n-1) - (d-1) - \lceil \frac{n}{2} \rceil = \lfloor \frac{n}{2} \rfloor - d$. Then we have $|S_i| = \binom{\lfloor \frac{n}{2} \rfloor - d + d - 1}{d-1} = \binom{\lfloor \frac{n}{2} \rfloor - 1}{d-1}$. Hence

$$|A_d| = \binom{n-2}{d-1} - d \cdot \binom{\lfloor \frac{n}{2} \rfloor - 1}{d-1}.$$

For any vertex in a d-regular tree, say v, let $(t^v_{v_1}, t^v_{v_2}, \cdots, t^v_{v_d})$ denote the orders of branches $(T^v_{v_1}, T^v_{v_2}, \cdots, T^v_{v_d})$ which is the decomposition of G_n in a d-regular tree in which $\sum_{v_i \in N(v)} v_i = n - 1$. Consider $(a_1, a_2, \cdots, a_d) \in A_d$, v is a unique rumor center of any graph that the orders of branches of it is satisfied $t^v_{v_1} = a_1, t^v_{v_2} = a_2, \cdots, t^v_{v_d} = a_d$.

We want to calculate the total number of ways to spread a rumor to n vertices from v. In addition, the order of breaches of this form graph G_n is $(t^v_{v_1}, t^v_{v_2}, \cdots, t^v_{v_d})$. First, assume that the vertex v had spread a rumor to u where $u \in N(v)$, and then it can be spread to u's descendants only. There are $d - 1$ choice of the next infected vertices since u has $d - 1$ children. Now, there are two vertices (u and a child of u) have rumors, then there are $2d - 3$ choice of the next infected vertices. Therefore, the number of ways to spread a rumor to m vertices in that graph T_u is

$$\prod_{i=1}^{m} \left((d-2)(i-1) + 1 \right). \tag{14}$$

Hence, given $(t^v_{v_1}, t^v_{v_2}, \cdots, t^v_{v_d})$, the total number of ways to spread a rumor is

$$(n-1)! \prod_{k=1}^{d} \frac{\prod_{i=1}^{t^v_{v_k}} ((d-2)(i-1)+1)}{t^v_{v_k}!}. \tag{15}$$

Note that the number of ways to permute $n-1$ steps from different subtrees is a permutation of the multiset $\{t^v_{v_1} \cdot 1, t^v_{v_2} \cdot 2, \cdots t^v_{v_d} \cdot d\}$.

Let $P_d(n)$ be the ratio of the number of ways to spread a rumor to n vertices such that $(t^v_{v_1}, t^v_{v_2}, \cdots, t^v_{v_d}) \in A_n$ to the number of ways to spread a rumor to n vertices such that $(t^v_{v_1}, t^v_{v_2}, \cdots, t^v_{v_d}) \in B_n$. Thus $P_d(t) \approx \mathbf{P}(E_t)$. We have

$$P_d(n) = \frac{\displaystyle\sum_{(t^v_{v_1}, t^v_{v_2}, \cdots, t^v_{v_d}) \in A_d} \left(\prod_{k=1}^{d} \frac{\prod_{k=1}^{t^v_{v_k}} ((d-2)(i-1)+1)}{t^v_{v_k}!} \right)}{\displaystyle\sum_{(t^v_{v_1}, t^v_{v_2}, \cdots, t^v_{v_d}) \in B_d} \left(\prod_{k=1}^{d} \frac{\prod_{k=1}^{t^v_{u}} ((d-2)(i-1)+1)}{t^v_{v_k}!} \right)}. \tag{16}$$

Theorem 3.1.3. *If G is a 3-regular graph, then we have that*

$$\lim_{t\to\infty} \mathrm{P}(E_t) = \frac{1}{4}.$$

Proof:

Let the source be v. By (16),

$$P_3(n) = \frac{\displaystyle\sum_{(t^v_{v_1}, t^v_{v_2}, t^v_{v_3}) \in A_3} \left(\prod_{k=1}^{d} \frac{\prod_{i=1}^{t^v_{v_k}} i}{t^v_{v_k}!} \right)}{\displaystyle\sum_{(t^v_{v_1}, t^v_{v_2}, t^v_{v_3}) \in B_3} \left(\prod_{k=1}^{d} \frac{\prod_{i=1}^{t^v_{v_k}} i}{t^v_{v_i}!} \right)} = \frac{\displaystyle\sum_{(t^v_{v_1}, t^v_{v_2}, t^v_{v_3}) \in A_3} 1}{\displaystyle\sum_{(t^v_{v_1}, t^v_{v_2}, t^v_{v_3}) \in B_3} 1}$$

$$= \frac{|A_3|}{|B_3|} = \frac{\binom{n-2}{2} - 3 \cdot \binom{\lfloor \frac{n}{2} \rfloor - 1}{2}}{\binom{n-2}{2}} = 1 - 3\frac{(\lfloor \frac{n}{2} \rfloor - 1)(\lfloor \frac{n}{2} \rfloor - 2)}{(n-2)(n-3)}.$$

This implies that

$$P_3(n) = \begin{cases} 1 - \dfrac{3}{4}\dfrac{(n-4)}{(n-3)} & \text{if } n \text{ is } \textit{even}, \\ 1 - \dfrac{3}{4}\dfrac{(n-5)}{(n-2)} & \text{if } n \text{ is } \textit{odd}. \end{cases} \tag{17}$$

Hence, the proof follows. ■

3.2 Detection probabilities of d-regular trees, $d \geq 4$

In what follows, we use (16) to calculate the detection probability of finding the source in a d-regular tree, $d \geq 4$. Let v be the source and let $w_{dn} = \prod_{i=1}^{n} ((d-2)(i-1)+1)$ (14). We can rewrite it by (13).

$$P_d(n) = 1 - \frac{\sum_{i=1}^{d} \sum_{(x_1,x_2,\cdots,x_d)\in S_i} \left(\frac{w_{dx_1}}{x_1!} \frac{w_{dx_2}}{x_2!} \cdots \frac{w_{dx_d}}{x_d!} \right)}{\sum_{(t^v_{v_1},t^v_{v_2},\cdots,t^v_{v_d})\in B_d} \left(\frac{w_{dt^v_{v_1}}}{t^v_{v_1}!} \frac{w_{dt^v_{v_2}}}{t^v_{v_2}!} \cdots \frac{w_{dt^v_{v_d}}}{t^v_{v_d}!} \right)}.$$

Let $x_d \geq \frac{n}{2}$,

$$P_d(n) = 1 - \frac{d \sum_{(x_1,x_2,\cdots,x_d)\in S_d} \left(\frac{w_{dx_1}}{x_1!} \frac{w_{dx_2}}{x_2!} \cdots \frac{w_{dx_d}}{x_d!} \right)}{\sum_{(t^v_{v_1},t^v_{v_2},\cdots,t^v_{v_d})\in B_d} \left(\frac{w_{dt^v_{v_1}}}{t^v_{v_1}!} \frac{w_{dt^v_{v_2}}}{t^v_{v_2}!} \cdots \frac{w_{dt^v_{v_d}}}{t^v_{v_d}!} \right)}.$$

Let $f(x)$ be the exponential generating function for the sequence $\{w_{dn}\}_{n=1}^{\infty}$. And we have

$$(1-ax)^{-\frac{1}{a}} = \sum_{n=0}^{\infty} \binom{-\frac{1}{a}}{n} (-ax)^n = 1 + \sum_{n=1}^{\infty} \frac{\prod_{i=1}^{n}(a(i-1)+1)}{n!} x^n. \tag{18}$$

From (18), we immediately know $f(x) = (1-ax)^{-\frac{1}{a}} - 1$ where $a = d-2$. Let $F_k(x) = (f(x))^k$. We have

$$\begin{aligned}
F_k(x) &= \left(w_{d1}\frac{x}{1!} + w_{d2}\frac{x^2}{2!} + \cdots\right)^k = \left((1-ax)^{-\frac{1}{a}} - 1\right)^k \\
&= \sum_{l=0}^{k} \binom{k}{l} (1-ax)^{-\frac{l}{a}} (-1)^{k-l} \\
&= (-1)^k + \sum_{l=1}^{k} (-1)^{k-l} \binom{k}{l} \sum_{n=0}^{\infty} \binom{-\frac{l}{a}}{n} (-ax)^n \\
&= (-1)^k + \sum_{l=1}^{k} (-1)^{k-l} \binom{k}{l} \left(1 + \sum_{n=1}^{\infty} \binom{-\frac{l}{a}}{n} (-ax)^n\right) \\
&= \sum_{l=0}^{k} (-1)^{k-l} \binom{k}{l} + \sum_{l=1}^{k} (-1)^{k-l} \binom{k}{l} \sum_{n=1}^{\infty} \frac{\prod_{i=1}^{n} (l + a(i-1))}{n!} x^n \\
&= \sum_{n=1}^{\infty} \left(\sum_{l=1}^{k} (-1)^{k-l} \binom{k}{l} \frac{\prod_{i=1}^{n} (l + a(i-1))}{n!} \right) x^n.
\end{aligned}$$

Let $[x^n]F(x)$ be the coefficient of x^n in $F(x)$. Then,

$$[x^n]F_k(x) = \sum_{l=1}^{k} (-1)^{k-l} \binom{k}{l} \frac{\prod_{i=1}^{n} (l + a(i-1))}{n!}. \tag{19}$$

The coefficient of x^n in $F(x)$ is $\frac{w_{dn_1}}{n_1!}\frac{w_{dn_2}}{n_2!}\cdots\frac{w_{dn_k}}{n_k!}$ whose degree is $n = n_1 + n_2 + \cdots + n_k$. Therefore, Let $a = d-2$, the detection probability of finding the source in a d-regular tree is

$$P_d(n) = 1 - \frac{d \sum_{m \geq \frac{n}{2}}^{n-d} \frac{\prod_{i=1}^{m} (a(i-1)+1)}{m!} \left([x^{n-1-m}]F_{d-1}(x)\right)}{\left[x^{n-1}\right] F_d(x)}. \tag{20}$$

Clearly, this is too complicate to simplify the right side of (20). In Figure 5, we use computer to obtain the general behavior of this term for several d's and $n \leq 100,000$. As a matter of fact, we have $\frac{1}{2} \geq P_d(n) \geq P_{d'}(n) \geq \frac{1}{4}$ if $d \geq d' \geq 3$

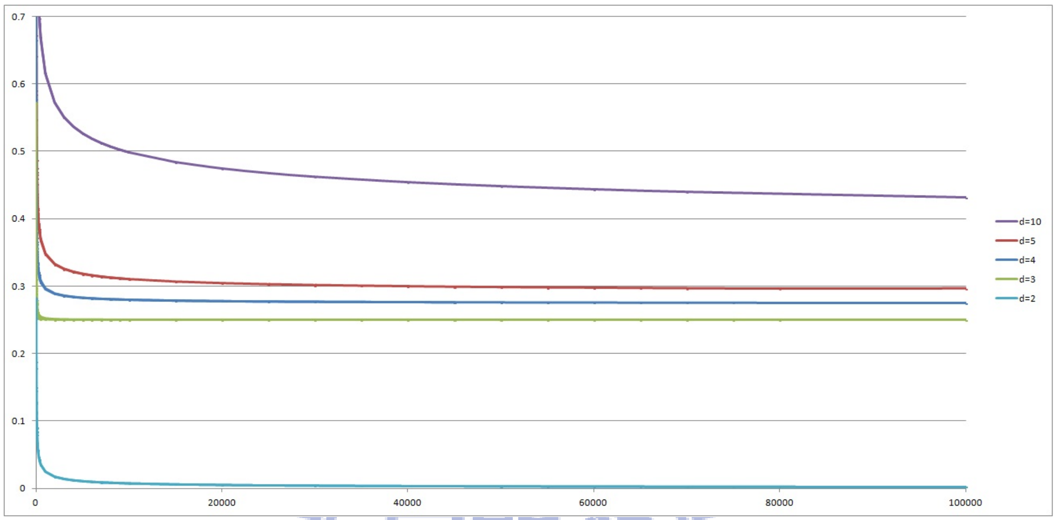

Figure 4: Detection probabilities of regular trees

4 Conclusion

In this thesis, we have obtained a mathematical model for finding the culprit who spreads rumors in a network defined on a d-regular tree (countably infinite graph). We are concerned with the detection probabilities of finding the culprit. By using this model, we are able to give a shorter and more explicit proof for the cases when $d \leq 3$. See [6] for a comparison. Furthermore, we can estimate the detection probabilities of finding the source in d-regular trees for $d > 3$ by an explicit formula, though it is quite complicate. It will be better if we can simplify the formula by using certain combinatorial identities. Moreover, if we can reduce the estimation error for general graphs, and generalize the estimator to networks with different rumor spreading rate, then we have a much better result than the known works in [6].

References

[1] N. T. J. Bailey, *The Mathematical Theory of Infectious Diseases and its Applications*, second edition, Griffin, London, 1975.

[2] A. Ganesh, L. Massoulie, and D. Towsley, "The effect of network topology on the spread of epidemics", Proc. 24th Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM), vol. 2, pp. 1455-1466, March 2005.

[3] C. Moore and M. E. J. Newman, "Epidemics and percolation in small-world networks", Phys. Rev. E, vol. 61, pp. 5678-5682, May 2000.

[4] M. E. J. Newman, "The spread of epidemic disease on networks", Phys. Rev. E, vol. 66, pp. 016128, July 2002.

[5] R. Pastor-Satorras and A. Vespignani, "Epidemic spreading in scale-free networks", Phys. Rev. Lett., vol. 86, pp. 3200-3203, April 2001.

[6] D. Shah and T. Zaman, "Rumors in a Network: Who's the Culprit?", IEEE Transactions on Information Theory, vol. 57, pp. 5163-5181, August 2011.

[7] J. H. Van Lint and R. M. Wilson, *A Course in Combinatorics*, second edition, Cambridge University Press, Cambridge, 2001.

[8] D.B. West, *Introduction to Graph Theory*, second edition, Prentice-Hall, Upper Saddle River, New Jersey, 2001.