# 國 立 交 通 大 學

應用數學系

數學建模與科學計算碩士班

碩士論文

# 網路中謠言散播中心的研究

# Search for Rumor Center

研究生：虞沛鐸

指導教授：傅恆霖 教授

中 華 民 國 一 百 零 二 年 五 月

# 網路中謠言散播中心的研究

# Search for Rumor Center


研究生：虞沛鐸　　　　Student: Pei-Duo Yu

指導教授：傅恆霖 教授　　　　Advisor: Hung-Lin Fu

國立交通大學

應用數學系

數學建模與科學計算碩士班

碩士論文

A Thesis Submitted to Institute of Mathematical Modeling and Scientific Computing
Department of Applied Mathematics
College of Science
National Chiao Tung University
For the Degree of Master
May 2013
Hsinchu, Taiwan, Republic of China


中 華 民 國 一 百 零 二 年 五 月

# 網路中謠言散撥中心的研究

**學生: 虞沛鐸　指導教授: 傅恆霖 教授**

國立交通大學
應用數學系數學建模與科學計算碩士班

## 摘要

在現今的世界裡，互連的網路結構中充滿著風險：一個單點發生問題，可能因爲互連的結構而被傳播至整個網路，進而放大其負面影響。從本質上來看，可以用謠言在網路中傳播的模型，來描述這種形式的傳播，在有傳播的模型之後，我們進一步的希望可以找到問題的源頭，並且在有限的資訊下(只知道有問題的點的互連關係)，加以控制以及預防這類型的風險發生。

在這篇論文當中，我們以樹狀網路爲主，利用從著名的傳染病模型“SIR-模型”簡化而來的“謠言散佈模型”來研究問題的源頭可能出現在哪裡。這篇論文將會包含一些有限樹狀網路與無限樹狀網路的新結果。

# Search for Rumor Center

**Student: Pei-Duo Yu  Advisor: Hung-Lin Fu**

*Department of Applied Mathematics*
*Institute of Mathematical Modeling and Scientific Computing*
*National Chiao Tung University*
*Hsinchu, Taiwan 30050*

## Abstract

In the modern world there are many network risks which share a common structure: an isolated risk is amplified because it is spread by the network. In essence, all of these types of spreading phenomenon can be modeled as a *rumor* spreading through a network, where the goal is to find the source of the rumor in order to control and prevent these network risks based on limited information about the network structure and the *rumor infected* nodes.

In this thesis, we shall use the so-called *Rumor Spread* model which is simplified from an epidemic model called *Susceptible-Infected-Recovered model* to study the *Rumor Center* in a tree-shaped network. Several new results are obtained on the cases where the network is defined on a *d-regular* tree either infinite or finite.

# 誌謝

首先，要感謝我的指導教授傅恆霖老師，從大二修過老師的離散數學之後就開始對這個領域產生興趣，也因位如此，剛進碩士班的時候希望找傅老師當指導教授，感謝老師能指導我完成論文，在這一年半的過程中，我學到很多作研究的人應該有的態度。

另外，要感謝學長姊們的幫忙，呂惠娟學姊、貓頭學長、連敏筠學姊、施智懷學長、軒軒學長，常常一有問題就跟他們討論，他們也很願意花時間來了解我的問題，並且幫忙我。

謝謝當初跟我一起找老師的精靈，從一開始到結束，我們總是一起討論課業上的問題，或是論文上的問題，還有謝謝陪我一起消磨時間的研究室同學們：阿沐、冠維、小關、黑輪、宇鎮、小胖、LULU、李昱融、黃敏琇。謝謝一起當微積分助教做事細心的陳珈惠，一起分攤了不少助教要做的事。

特別感謝符麥克老師，提供了我另外一個觀點來看這個問題，雖然還沒完整解決，但是是個有希望可以做出來的方向，感謝麥可老師花不少時間跟我討論這個問題，也很願意教我一些我沒學過的領域。

最後要感謝我的父母，讓我在沒有經濟壓力以及家庭壓力的情況下，安穩的完成這篇論文。

# Contents

# List of Figures

# Chapter 1

# Preliminaries

In this chapter, we will first introduce the background of this problem, and some notations and knowledge about graph theory. Then, a model we use throughout this thesis will be introduced.

## 1.1 Introduction

There are many networks in our daily life, to transmit electricity to everywhere in the city we need electrical power grid network. We have social network to describe the relation among people. Even in our body there is still some kinds of networks, for example, neural networks. Any two units in a given network can be connected together, so a single defect may have a great influence to the whole network. For instance, if there was a boy got the flu, then all his classmates will have a chance to be infected. Moreover, the flu may infect all over the school. In a city, if there is a telegraph pole losing its function then the city may confront with power failure. The priority of all these situations is to find out the first problematic unit, called the *source*, in the network. We need a better way to find out the source rather than to examine all problematic units. So, we need a model to describe all those situations and try to determine where is the source. Here we use the model called $Rumor-Spread$ *model* (RS-model for short), simplified from an epidemic model called SIR - model, where $S$ stands for susceptible, $I$ stands for infected and $R$ stands for recovered. In an RS - model there are only two kinds of vertices, infected and susceptible, the reason is that once a person knew the rumor, he can't get rid of the fact that he knew the rumor. We use the combinatorial counting to get the same result as in [5] in a relative easy way. Prior works on disease spreading are focus on the spreading of the epidemics or the lifetime of the disease. In [3], it showed that under some certain network topology, the relation between the ratio of infected vertices to cured vertices and the spectral radius of the network will decide the lifetime of the epidemic.

## 1.2 Preliminaries and Notations

A *cycle* is a graph with an equal number of vertices and edges whose vertices can be placed around a circle so that two vertices are adjacent if and only if they appear consecutively along the circle. *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 $v$.

A graph $G$ is *connected* if for any two vertices $u$, $v$ there is a u-v path in $G$. A *tree* is a connected simple graph without cycle. A *subtree* of a tree $T$ is a tree $T'$ with $V(T) \subseteq V(T')$ and $E(T) \subseteq E(T')$ . The *neighborhood* of a vertex $v$ is a set of all neighbors of v denoted as $N(v)$. The *degree* of a vertex $v$ is the number of its neighbors denoted as $d(v)$. A graph is said to be *k-regular* if all its vertices is of degree $k$. All degree 1 vertices in a tree are called *leaves*.

In this thesis, a *d-regular tree* is a tree where all its vertices are of $d - regular$, that is, the order and the size of this tree are infinite. A *rooted tree* $T$ is a tree with one vertex $r$ chosen as a *root* which can also denoted as $v_r$. The length of the simple path from the the root $v_r$ to a vertex $v$ is the *depth* of $v$ in $T$. The length of the simple path from the the vertex $v$ to a closest leaf to $v$ is the *height* of $v$. The *height* of a rooted tree is equal to its root's *height*. A *level* of a tree consists of all vertices at the same depth. For any vertex $v$ in a rooted tree with root $r$, a *parent* of $v$ is its neighbor on $r-v$ path ; the *children* are its other neighbors and *child*$(v)$ denote the set of all children of $v$. A *descendent* $u$ of $v$ is the vertex that $v$ contained in $r-u$ path. See more details about these notations and definitions in [1] and [8]. Given a tree $T$ with root $r$, a *branch* $T_v^r$ is the subtree with $v$ as a root contains all its descendent in $T_r$. For convenience, let $t_v^r$ denote the order of $T_v^r$. So for a $d-regular$ rooted tree with root $r$ and order $n$, we have $\sum_{i=1}^{d} t_i^r$.

## 1.3 Rumor Spreading(RS) Model

Here we use $SI$ - model (susceptible-infected) to describe a rumor spreading in a group of people. It is a discrete time model which is simplified from the well-known SIR model. There are two kinds of vertices in this model. Let $S(t)$ be a set of people who don't know rumor yet at time $t$ and $I(t)$ be a set of people who have known rumors at time $t$. For convenience, let $S_t$ denote $|S(t)|$ and $I_t$ denote $|I(t)|$. Using a fixed population, that is, $S_t + I_t = N$ and assume that each in time period only and exactly one rumor is spread so we have the following:

$$S_t + 1 = S_t + 1$$

$$I_t + 1 = I_t + 1$$

$$S_0 = N, I_0 = 0.$$

In a group of people, each person is represented by a vertex in a graph, and if any two of them know each other then we describe the relationship as an edge in the graph. Given a group of people, we have a graph $G$ corresponding to these people. Let $G_t$ be a subgraph of order $t$ of $G$. $G_t$ is composed of $t$ infected vertices which means the people who knew the rumor at time $t$. So, we call $G_1$ rumor source, that is, $G_1$ is the person where the rumor begin to be spread. Since in each time period exactly one rumor is spread, we have $|G_t + 1| = |G_t| + 1$ and $G_t + 1$ is developed from $G_t$ by adding a vertex $z$ and an edge to $G_t$.

For clearness, in our models presented in this thesis, we assume that the infected probability of all vertices that may be infected in next time period is equally possible, that is

$$P_{t+1}(z) = \frac{1}{\sum_{v \in V(G_t)} d(v) - 2(t-1)}, \tag{1.1}$$

and the probability of being the source are equivalent for all vertices.

## 1.4 Known Results

Prior study focused on the act of an epidemic's propagation in a network. [3] showed the lifetime of an epidemic in a network is related to ratio of cure rate to infection rate and the spectral radius of the graph(network). In [5], it showed the detection probability goes to zero as $n$ gets larger when $G$ is a *2-regular* graph. And the detection probability is 1/4 when $G$ is a *3-regular* graph. Moreover, it computed the detection probability when $G$ is not a regular tree, by computing the detection probability of the $BFS$-spanning tree of $G$ to approximate the detection probability of $G$. There are some numerical heuristic results shown in [5]. In [4], it provided the same result as in [5], but in a relatively simple method: the combinatorial counting.

# Chapter 2

# Rumor Source Estimator for Tree

In this chapter, we will introduce a maximum likelihood estimator to estimate the rumor source in a tree shape network. For any given tree $T$, let $G_n \subseteq T$ be the infected subtree in T. The $ML-estimator$ for the rumor source is the vertex $v$ with the maximum probability $P(G_n|v)$.

## 2.1 The Probability $P(G_n|v)$ in Regular Tree

Let $T$ be a tree represents a network. According to the assumption of an $RS-model$, let $G_n$ be the subtree of $T$ at time $n$, that is, $|G_n| = n$ and $G_n$ represents the $n$ infected vertex.

$P(G_n|v^*)$ is the probability that view $v^*$ as the source and spreads rumor to all vertices in $G_n$ within $n$ steps. Let $\sigma_i$ be the possible infecting order starts from the source, and $S(v^*, G_n)$ be the collection of all $\sigma_i$ where $v^*$ is viewed as the source in $G_n$. Then, we have

$$P(G_n|v^*) = \sum_{\sigma_i \in S(v^*,G_n)} P(\sigma_i|v^*). \tag{2.1}$$

For simplicity, we explain how to calculate $P(G_4|v_1)$ by an example as shown in Figure 2.1. The vertices colored black are the infected vertices. Suppose $v_1$ is the source, then all possible infecting order $\sigma_i$ are :

$\sigma_1 = v_1 \to v_2 \to v_3 \to v_4 \qquad \sigma_2 = v_1 \to v_2 \to v_4 \to v_3 \qquad \sigma_3 = v_1 \to v_3 \to v_2 \to v_4$

$\sigma_4 = v_1 \to v_3 \to v_4 \to v_2 \qquad \sigma_5 = v_1 \to v_4 \to v_3 \to v_2 \qquad \sigma_6 = v_1 \to v_4 \to v_2 \to v_3$

(Note that, if we see $v_3$ as the source then $v_3 \to v_2 \to v_1 \to v_4$ is not a possible infecting order, since there is no edge between $v_3$ and $v_2$.)

Let's calculate the probability of $\sigma_1$, recall that the infected probability of $(t+1)_{th}$ vertex is $P_{t+1}(z) = \frac{1}{\sum_{v \in V(G_t)} d(v) - 2(t-1)}$.

Therefore, $P(\sigma_1|v_1) = \frac{1}{3} \cdot \frac{1}{4} \cdot \frac{1}{5}$, and also we know that $P(\sigma_i|v_1) = \frac{1}{3} \cdot \frac{1}{4} \cdot \frac{1}{5}$ for $i = 1, 2, 3, 4, 5, 6$.

In general, we have

$$P(\sigma_i|v_1) = \prod_{k=1}^{n-1} \frac{1}{\sum_{v_i \in V(G_k)} d(v_i) - 2(k-1)}, \tag{2.2}$$

where $G_k$ is a subgraph of $G_n$ and it represents the infected subgraph at $k_{th}$ time step along with the infecting order $\sigma_i$ .

For $d - regular$ tree,

$$P(\sigma_i|v_1) = \prod_{k=1}^{n-1} \frac{1}{dk - 2(k-1)}. \tag{2.3}$$

Note that if the rumor has not been spread to leaves of the tree yet, then $P(\sigma_i|v_1) = P(\sigma_j|v_1)$ for all $\sigma_i, \sigma_j \in S(v, G_n)$. Now, combining (2.1) and (2.3), then we have

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

We can now compute each $P(G_4|v_i)$ by finding out the value $|S(v_i, G_n)|$. Since they are of the same value $\frac{1}{3 \cdot 4 \cdot 5} = \frac{1}{60}$.

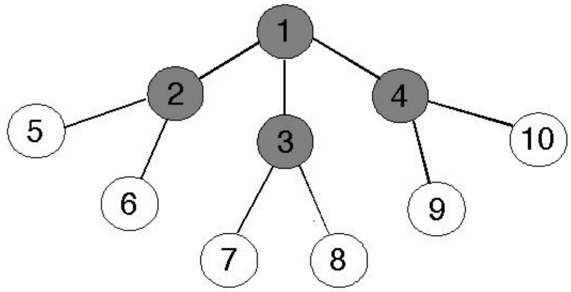



Figure 2.1: The example of how to calculate $P(G_4|v_1)$

We have:

$$P(G_n|v_1) = 3! \cdot \frac{1}{60}$$

$$P(G_n|v_2) = P(G_n|v_3) = P(G_n|v_4) = 1 \cdot 2! \cdot \frac{1}{60}$$

therefore, by the definition, the ML estimator is $v_1$.

In conclusion, $P(G_n|v)$ is proportional to $|S(v, G_n)|$, that is, if the rumor didn't meet the end vertex in the tree yet, then the only fact that affect $P(G_n|v)$ is the number of distinct ways that $v$ can spread the rumor to all vertices in $G_n$.

Let $R(v, G_n) = |S(v, G_n)|$ be called *rumor centrality* which is a important information about the vertex $v$, we call $v$ the *rumor center* of $G_n$ if $v$ has the maximum *rumor centrality*.

## 2.2 Rumor Centrality $R(v, G_n)$

Since the $P(G_n|v)$ is proportional to $R(v, G_n)$, we are able to determine which vertex is the ML estimator after calculating all $R(v_i, G_n)$ where $v_i \in G_n$. The following will introduce a formula represents the rumor centrality.

Let $G_n$ be a rooted tree with root $v_r$ and suppose $G_1 = v_r$, then in $G_2$, the second infected vertex may be any child of $v_r$. There are $d(v)$ vertices in $child(v)$ say $u_i$ where $i = 1, 2, 3 \ldots d(v)$. So we have:

$$R(v, G_n) = \frac{(n-1)!}{t_{u_1}^v! \cdot t_{u_2}^v! \cdot \ldots \cdot t_{u_{d(v)}}^v!} \cdot \prod_{i=1}^{d(v)} R(u_i, T_{u_i}^v). \tag{2.4}$$

We can rewrite (2.4) by the recursion from the root $v_r$ to all leaves of $G_n$, then we get a simpler form :

$$R(v, G_n) = n! \cdot \prod_{u \in G_n} \frac{1}{t_u^v}. \tag{2.5}$$

Now, consider two adjacent vertices $u$,$v$ in $G_n$ and a vertex $w \in G_n - \{u, v\}$, then we have $t_u^v = n - t_v^u$ and $t_w^v = t_w^u$, where $t_u^v$ is defined in preliminary.
By the above two facts, we conclude that

$$\frac{P(u|G_n)}{P(v|G_n)} = \frac{R(u, G_n)}{R(v, G_n)} = \frac{t_u^v}{n - t_v^u}. \tag{2.6}$$

**Theorem 2.2.1.** [5] *Given an $n$ vertices tree $G_n$. $v \in G_n$ is a rumor center if and only if*

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

*for all $u \in G_n - \{v\}$.*

In $G_n$, we can define the *weight* [4] of a vertex $v$ in $G_n$, it is defined as $weight(v) = max_{u \in child(v)}\{t_c^v\}$ . The vertex of $G_n$ with the minimum *weight* is called the *mass center* of $G_n$. More results on *mass center* of a tree can be found in [9]. Moreover, the *distance centrality* [5] of $v \in G_n$ is defined as $D(v, G_n) = \sum_{j \in G_n} d(v, j)$. The vertex in $G_n$ with minimum distance centrality is called *distance center*.

**Theorem 2.2.2.** *Let $G_n$ be defined as in theorem 2.2.1 and $v$ is a vertex in $G_n$, then the following statements are equivalent:*

*1. $v$ is a distance center of $G_n$.* [5]

*2. $v$ is a rumor center of $G_n$.* [5]

*3. $v$ is a mass center of $G_n$.*

**Proof.** Given $G_n$ be a tree of size $n$, and $v \in G_n$.

($1 \Rightarrow 2$) We prove it by contraposition argument, suppose $v$ is not a rumor center, by (2.2.1) there is a branch of $v$, say $T_u^v$, with order $> n/2$ and $u$ is adjacent to $v$. Now, we need a relationship between $\sum_{s \in G_n} d(v,s)$ and $\sum_{s \in G_n} d(u,s)$ described as following.

$$\sum_{s \in G_n} d(v,s) = \sum_{s \in G_n} d(u,s) + (t_u^v - 1) - (t_v^u - 1).$$

We have $\sum_{s \in G_n} d(v,s) > \sum_{s \in G_n} d(u,s)$, since $t_u^v > t_u^v$. This implies $v$ is not a *distance center*.

($2 \Rightarrow 3$) To prove this, we need a fact: If all $v$'s branches are of order $\leq n/2$, then $v$ is a mass center. Again, by contraposition argument, suppose $v$ is not a mass center, then there exists a branch of $v$ whose order $> n/2$, that is, $v$ is not a rumor center by (2.2.1).

($3 \Rightarrow 1$) Suppose $v$ is a mass center, then all it's branch is of order $\leq n/2$. This implies $v$ is a rumor center. Let $u \in G_n$, if $u$ is adjacent to $v$, then $\sum_{s \in G_n} d(v,s) < \sum_{s \in G_n} d(u,s)$ and we finish the proof. If $u$ is not adjacent to $v$, then we can partition all vertices in $G_n$ into three sets. The first one is $T_v^u$, the second one is $T_u^v$ and the last one contains all vertices not in $T_v^u$ and $T_u^v$ say $R$. Let $l$ denote $d(u,v)$. Now, consider

$\sum_{s \in G_n} d(v,s) - \sum_{s \in G_n} d(u,s)$

$= (\sum_{s \in T_v^u} d(v,s) + \sum_{s \in T_u^v} d(v,s) + \sum_{s \in R} d(v,s)) - (\sum_{s \in T_v^u} d(u,s) + \sum_{s \in T_u^v} d(u,s) + \sum_{s \in R} d(u,s))$.

Since $v$ is the rumor center, we have :

(1) $|R| + t_u^v \leq n/2$, and $t_v^u > n/2$

(2) $(\sum_{s\in T_u^v} d(v,s) + \sum_{s\in T_v^u} d(v,s)) - (\sum_{s\in T_u^v} d(u,s) + \sum_{s\in T_v^u} d(u,s)) = l \cdot (t_u^v - t_v^u)$

(3) $|\sum_{s\in R} d(v,s) - \sum_{s\in R} d(u,s)| \leq l \cdot |R|$.

Combine these three properties, we conclude that $\sum_{s\in G_n} d(v,s) - \sum_{s\in G_n} d(u,s) < 0$, for any $u \in G_n$, that is, $v$ is the *distance center*. □

In [5], it shows that there are at most two rumor centers occur when the maximum size branch of rumor center is of size n/2. In fact, a tree has a either exactly one or exactly two median, see [9] for reference. Note that the definition of *median* in [8] is similar to distance center.

Now we know how to find the rumor center in a given tree $G_n$, [5] gave an algorithm called $message-passing\ algorithm$ for calculating each vertex's rumor centrality $R(v, G_n)$ with time complexity $O(n)$. In chapter 3, we will discuss the correctness of our estimator.

# Chapter 3

# The Estimator for Rumor Source in Infinite Trees

In chapter 2, we introduced the idea of "the most possible source", which we call it a rumor center. When we want to find out the source in $G_n$, the first vertex we should consider is the rumor center $v$. So in the following, we will show how to compute the probability $P(source = v|G_n)$.

## 3.1 The Detection Probability $P(source = v|G_n)$

Let $G$ be a tree and $G_n$ is a subtree of $G$ at time $n$, $v$ is the rumor center of $G_n$. $P(source = v|G_n)$ is the probability of the event that the rumor center is exactly the rumor source when given $G_n$. We have,

$$P(source = v|G_n) = \frac{P(G_n|source = v) \cdot P(source = v)}{\sum_{i \in G_n}[P(G_n|source = i) \cdot P(source = i)]}. \tag{3.1}$$

The probability of each vertex to be the source are equal, that is

$$P(source = i) = P(source = j) \ \ \forall\ i, j \in G_n,$$

and also we have $P(G_n|source = v) \propto R(v, G_n)$.

Thus,

$$P(source = v|G_n) = \frac{R(v, G_n)}{\sum_{i \in G_n} R(i, G_n)}. \tag{3.2}$$

With (3.2) we can derive several results about the detection probability.

**Theorem 3.1.1.** [5] [4] *Let $G$ and $G_n$ be defined as above, and $v$ be the rumor center in $G_n$. Then the detection probability*

$$P(source = v|G_n) \leq \frac{1}{2}.$$

**Proof.** Given $G_n$ be the subtree of $G$ with size $n$, and let $v$ be the rumor center of $G_n$. Obesrve that, once the rumor has been spread from $v$ to one of its neighbors say $u$, the number of spreading ways in those remaining vertices is equal to the rumor is originate from $u$ and been spread to $v$ first. So, after the rumor spread from $v$ to $u$, we contract the edge $(u, v)$ and we view it as a new graph say $G_{u,v}$. The new vertex from the contraction called $v'$. Then we conclude that:

$$R(v, G_n) = \sum_{u \in N(v)} R(v', G_{u,v}).$$

So, from (3.2) we have

$$P(source = v|G_n) = \frac{R(v,G_n)}{\sum_{i \in G_n} R(i,G_n)} \leq \frac{R(v,G_n)}{R(v,G_n)+\sum_{u \in N(v)} R(v',G_{u,v})} \leq 1/2.$$

□

No matter how large the size is or what shape of $G_n$ is, this trivial upper bound won't change. The following will introduce another bound which is derived from the full $d-regular$ tree. Before we introducing this bound, there are two assumptions we need as following:

1. If the size $n$ is fixed, then the full tree has the maximum detection probability.

2. Let $G_n$ be a full tree, then for any $k > n$ the detection probability of $G_n$ is larger than it of $G_k$.

Let $T_c(l, d)$ denote the $d-regular$ complete tree with $l$ levels and $v'$ is its rumor center. With above assumptions, we have the following result:

**Theorem 3.1.2.** *Let $G$ be a $d-regular$ tree, and let $G_n$ and $v$ be defined as in (3.1.1). Let $l = \lfloor log_{d-1}(\frac{n-1}{d}(d-2)+1)\rfloor + 1$ and $n' = |T_c(l,d)|$ where $T_c(l,d)$ is the complete tree with $l$ levels. Then*

$$P(source = v|G_n) \leq \frac{1}{1 + \sum_{s=1}^{l-1}[d(d-1)^{s-1}\prod_{i=1}^{s}\frac{(d-1)^{l-i}-1}{n'd-2n'-(d-1)^{l-i}+1}]}.$$

**Proof.** Let $G$, $G_n$ and $v$ as defined above, since $T_c(l,d)$ is the maximum size complete tree satisfies $|T_c(l,d)| \leq |G_n|$ and $v'$ is the rumor center in $T_c(l,d)$. Then from the assumption above, we have :

$$P(source = v|G_n) \leq P(source = v'|T_c(l,d))$$

Now, consider $P(source = v'|T_c(l,d)) = \frac{R(v',T_c(l,d))}{\sum_{i\in T_c(l,d)} R(i,T_c(l,d))}$, our goal is to find out the ratio between $v'$ and other vertices in $T_c(l,d)$ . By the symmetric structure of the complete tree, we have $R(v_i, G_n) = R(v_j, G_n)$, if $v_i$ and $v_j$ are in the same level of the complete tree. Suppose $R(v', T_c(l,d)) = 1$, and let $v_s$ be the vertex with $depth = s$ , then by property (2.6) we have

$$R(v_{s-1},T_c(l,d)) : R(v_s,T_c(l,d)) = \left(n' - \frac{(d-1)^{l-s}-1}{(d-2)}\right) : \left(\frac{(d-1)^{l-s}-1}{(d-2)}\right), \tag{3.3}$$

for $s = 1,2,3,..l-1$, where $\frac{(d-1)^{l-s}-1}{(d-2)}$ is the number of vertices in $T_{v_s}^{v'} \subseteq T_c(l,d)$. We can rewrite (3.3) as

$$\frac{R(v_{s-1},T_c(l,d))}{R(v_s,T_c(l,d))} = \frac{(d-1)^{l-s}-1}{n'd-2n'-(d-1)^{l-s}+1}. \tag{3.4}$$

By the recursive relation in (3.4) and the assumption $R(v',T_c(l,d)) = 1$ we have

$$\sum_{i\in T_c(l,d)} R(i,T_c(l,d)) = 1 + \sum_{s=1}^{l-1}[d(d-1)^{s-1}\prod_{i=1}^{s}\frac{(d-1)^{l-i}-1}{n'd-2n'-(d-1)^{l-i}+1}],$$

where $d(d-1)^{s-1}$ is the number of vertex in each level.

Thus,

$$P(v' = source|T_c(l,d)) = \frac{1}{1 + \sum_{s=1}^{l-1} [d(d-1)^{s-1} \prod_{i=1}^{s} \frac{(d-1)^{l-i} - 1}{n'd - 2n' - (d-1)^{l-i} + 1}]},$$

and we complete the proof. □

Note that, when $n$ is little enough such that $l = 2$ then this bound is exactly 1/2 and this bound will decrease as $n$ getting larger. The following is the tendency of the upper bound decreasing as $n$ getting larger.

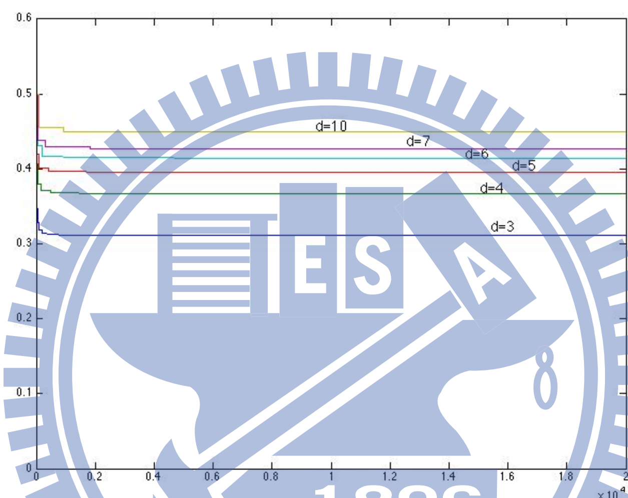



Figure 3.1: The Tendency of The Upper Bound With different $d$

## 3.2 The Value of $P(source = v|G_n)$

In this section, we will compute the detection probability of infinity order $d-regular$ tree and show two results which are the same as in [5] but in different methods.

**Theorem 3.2.1.** [5] [4] *Suppose $G$ is a $2-regular$ tree, and $G_n \subseteq G$ is the infected subtree of $G$,let $v$ be the rumor center in $G_n$ then*

$$\lim_{n \longrightarrow \infty} P(source = v|G_n) = 0.$$

This theorem shows that, the detection probability in a $2-regular$ tree goes to zero as the infected vertices increase.

**Proof.** Without loss of generality, suppose the order of the infected subtree is odd and equal to $2t+1$ for some $t \in N$.

Given $G$, $G_{2t+1}$ as defined above. By theorem 2.2.2, there is a rumor center say $v$ and with $R(v, G_{2t+1}) = \binom{2t}{t}$. Then, we have

$$P(source = v|G_{2t+1}) = \frac{R(v, G_{2t+1})}{\sum_{i \in G_{2t+1}} R(i, G_{2t+1})}$$
$$= \frac{\binom{2t}{t}}{\binom{2t}{0} + \binom{2t}{1} + \binom{2t}{2} + \ldots + \binom{2t}{2t}}$$
$$= \frac{(2t)!}{t! \cdot t!} \cdot \frac{1}{2^{2t}}.$$

By Stirling Formula, we have $\lim_{t \longrightarrow \infty} \frac{t!}{\sqrt{2\pi t} t^t e^{-t}} = 1$.

Now, consider $P(source = v|G_{2t+1})$ when $t$ is getting larger.

Let $H_i^v = \{G_i \subseteq G|\ G_i\ is\ a\ 2-regular\ graph\ and\ v\ is\ the\ source\ of\ G_i\}$. So, the detection probability when $t$ is getting larger is :

$$\lim_{t \longrightarrow \infty} \sum_{G_{2t+1} \in H_{2t+1}^v} [P(G_{2t+1}|v) \cdot P(source = v|G_{2t+1})] = \lim_{t \longrightarrow \infty} P(source = v|G_{2t+1})$$
$$= \lim_{t \longrightarrow \infty} \frac{\sqrt{4\pi t} \cdot (2t)^{2t} \cdot e^{-2t}}{(\sqrt{2\pi t} \cdot t^t \cdot e^{-t})^2} \cdot \frac{1}{2^{2t}}$$
$$= \lim_{t \longrightarrow \infty} \frac{1}{\sqrt{\pi t}} = 0.$$

□

The following is going to find out the detection probability in $d-regular$ tree with infinity order when $d > 2$ by the property in Theorem 2.2.1.

Let $A_d$ and $B_d$ be two sets such that

$$A_d = \{(a_1, a_2, ..., a_d) | 0 \leq a_i \leq \frac{n}{2}, \sum_{i=1}^{d} a_i = n-1\} \quad \textit{and}$$

$$B_d = \{(b_1, b_2, ..., b_d) | b_i \in N \cup \{0\}, \sum_{i=1}^{d} b_i = n-1\}.$$

Note that, these two sets are corresponding to the orders of branch's sizes of a vertex. And if its branch's size is in $A_d$ then it is the rumor center. We have $|B_d| = \binom{n-1+d-1}{d-1}$ and let $S_k = \{(s_1, s_2, ..., s_d) | s_k > \frac{n}{2}, 0 \leq s_j < \frac{n}{2}, \sum_{i=1}^{d} s_i = n-1\}$, so $|S_i| = \binom{d+\lceil \frac{n}{2} \rceil - 3}{d-1}$ and

$$\begin{aligned} |A_d| &= |B_d| - \sum_{k=1}^{d} |S_k| \\ &= \binom{n+d-2}{d-1} - d \cdot \binom{\lceil \frac{n}{2} \rceil + d - 3}{d-1} \end{aligned}$$

since $S_i \cap S_j = \phi$ for $i \neq j$.

Given a $d-regular$ tree $G$ and let $G_n \subseteq G$, suppose $v$ is the source in $G_n$ with the order of branch's size $(t^v_{u_1}, t^v_{u_2}, ..., t^v_{u_3})$. Recall (2.4), to compute all spreading ways that $v$ can spread to every vertices in $G_n$ we should compute the spreading way in each branch of $v$ first. Suppose $u \in N(v)$ and now $u$ has the rumor, then there are

$$\frac{-1}{d-3} \prod_{i=0}^{t^v_u} ((d-2)(i-1)+1), \text{ for } d \geq 4$$

spreading ways in $T^v_u$, since there are $(d-1)$ choice when $u$ is going to spread the rumor, and once $u$ spread the rumor to one of its child there are $(2d-3)$ and so on. Note that, when $d = 3$, $\frac{-1}{d-3}$ is meaningless so we discuss it separately later in next theorem. Thus, given $(t^v_{u_1}, t^v_{u_2}, ..., t^v_{u_d})$ of $v$, the number of spreading ways is

$$\frac{(n-1)!}{t^v_{u_1}! \cdot t^v_{u_2}! \cdot \ldots \cdot t^v_{u_{d(v)}}!} \cdot \prod_{k=1}^{d} [\frac{-1}{d-3} \prod_{i=0}^{t^v_{u_k}} ((d-2)(i-1)+1)], \text{ for } d \geq 4.$$

The corrected detection occurs when rumor center equals to source, so it occurs when the ordered pair of $v$'s branches belongs to $|A_d|$. Let $P_d(C_t)$ be the probability of corrected detection at time $t$ in a subtree of $d-regular$ tree, then we have

$$P_d(C_t) = \frac{(n-1)! \sum_{(t^v_{u_1},t^v_{u_2},\dots,t^v_{u_d})\in A_d} (\prod_{k=1}^{d} \frac{\frac{-1}{d-3}\prod_{i=0}^{t^v_{u_k}}((d-2)(i-1)+1)}{t^v_{u_k}!})}{(n-1)! \sum_{(t^v_{u_1},t^v_{u_2},\dots,t^v_{u_d})\in B_d} (\prod_{k=1}^{d} \frac{\frac{-1}{d-3}\prod_{i=0}^{t^v_{u_k}}((d-2)(i-1)+1)}{t^v_{u_k}!})}, \tag{3.5}$$

for $d \geq 4$.

**Theorem 3.2.2.** [5] [4] *Let $G$ be a $3-regular$ tree, then*

$$\lim_{t\longrightarrow\infty} P_3(C_t) = \tfrac{1}{4}.$$

**Proof.** Let $A_3$ and $B_3$ as defined above. To use (3.5), we should avoid the case when $i=0$, that is one of the branch size of $v$ is 0. Let $A_3^1$ be the subset of $A_3$ with at least a zero in $(t^v_{u_1}, t^v_{u_2}, t^v_{u_3})$ and $A_3^2 = A_3 \setminus A_3^1$. Also $B_3^1$ and $B_3^2$ are defined repectively. Let $z_d(i) = (d-2)(i-1)+1$ .Then by (3.5), we have

$$P_3(C_t) = \frac{\sum_{(t^v_{u_1},t^v_{u_2},t^v_{u_3})\in A_3^1} (\prod_{k=1}^{2} \frac{\prod_{i=1}^{t^v_{u_k}}(z_d(i))}{t^v_{u_k}!}) + \sum_{(t^v_{u_1},t^v_{u_2},t^v_{u_3})\in A_3^2} (\prod_{k=1}^{3} \frac{\prod_{i=1}^{t^v_{u_k}}(z_d(i))}{t^v_{u_k}!})}{\sum_{(t^v_{u_1},t^v_{u_2},t^v_{u_3})\in B_3^1} (\prod_{k=1}^{2} \frac{\prod_{i=1}^{t^v_{u_k}}(z_d(i))}{t^v_{u_k}!}) + \sum_{(t^v_{u_1},t^v_{u_2},t^v_{u_3})\in B_3^2} (\prod_{k=1}^{3} \frac{\prod_{i=1}^{t^v_{u_k}}(z_d(i))}{t^v_{u_k}!})}$$

$$= \frac{\sum_{(t^v_{u_1},t^v_{u_2},t^v_{u_3})\in A_3^1} 1 + \sum_{(t^v_{u_1},t^v_{u_2},t^v_{u_3})\in A_3^2} 1}{\sum_{(t^v_{u_1},t^v_{u_2},t^v_{u_3})\in B_3^1} 1 + \sum_{(t^v_{u_1},t^v_{u_2},t^v_{u_3})\in B_3^2} 1} = \frac{|A_3^1| + |A_3^2|}{|B_3^1| + |B_3^2|}.$$

For $t$ is even, we have

$$\frac{|A_3^1| + |A_3^2|}{|B_3^1| + |B_3^2|} = \frac{6 + \binom{t-2}{2} - 3\cdot\binom{\lceil\frac{t}{2}\rceil-2}{3-1}}{3 + 3(t-2) + \binom{t-2}{2}} = \frac{n^2+10n}{4t^2+4t}.$$

For $t$ is odd, we have

$$\frac{|A_3^1|+|A_3^2|}{|B_3^1|+|B_3^2|}=\frac{3+\binom{t-2}{2}-3\cdot\binom{\lceil\frac{t}{2}\rceil-2}{3-1}}{3+3(t-2)+\binom{t-2}{2}}=\frac{n^2+4n+3}{4t^2+4t}.$$

Thus, $\lim_{t\longrightarrow\infty} P_3(C_t)=\frac{1}{4}$. □

**Theorem 3.2.3.** *Let $G$ be a $d-regular$ tree, then*

$$\lim_{t\longrightarrow\infty} P_d(C_t)=1-\frac{d}{2}+\frac{(d-2)\cdot\Gamma\left(\frac{d}{d-2}\right)}{2^{\frac{d}{d-2}}\cdot\Gamma\left(\frac{1}{d-2}\right)\Gamma\left(\frac{d-1}{d-2}\right)}.$$

**Proof.** See the proof in Appendix C.

**Corollary 3.2.4.** *Let $G$ be a $d-regular$ tree, then*

$$\lim_{d\longrightarrow\infty}\lim_{t\longrightarrow\infty} P_d(C_t)=1-ln2.$$

# Chapter 4

# Trees with End Vertices

In this chapter, we will discuss about the case that $G$ is a tree not of infinte order. Suppose $G_n \subseteq G$, then there may exists a vertex called *end vertex* which can only accept the rumor but cannot spread it, since it has only one neighbor.

## 4.1 The Influence of End Vertices On The Probability $P(v = source|G_n)$

We begin this section with a simple example. Suppose $G$ is a finite order $3-$*regular* tree and $G_5 \subseteq G$ which is shown in figure 4.1. In $G_5$, it's easy to find the rumor center is $v_1$. Consider $P(G_5|v)$, for the spreading order $\sigma : v_1 \to v_2 \to v_5 \to v_3 \to v_4$, we have $P(\sigma|v_1) = \frac{1}{3} \cdot \frac{1}{4} \cdot \frac{1}{3} \cdot \frac{1}{4}$. Note that if $v_5$ is not the end vertex then $P(\sigma|v_1) = \frac{1}{3} \cdot \frac{1}{4} \cdot \frac{1}{5} \cdot \frac{1}{6}$. This shows that the time at when the rumor spread to $v_5$ will has influence on the probability $P(\sigma|v_1)$. Now we list out all the spreading ways in the following and sort it according to the position of $v_5$ in each $\sigma_i$:

$\sigma_1 : v_1 \to v_2 \to v_5 \to v_3 \to v_4$ $\quad$ $\sigma_2 : v_1 \to v_2 \to v_5 \to v_4 \to v_3$

$\sigma_3 : v_1 \to v_3 \to v_2 \to v_5 \to v_4$ $\quad$ $\sigma_4 : v_1 \to v_4 \to v_2 \to v_5 \to v_3$

$\sigma_5 : v_1 \to v_2 \to v_3 \to v_5 \to v_4$ $\quad$ $\sigma_6 : v_1 \to v_2 \to v_4 \to v_5 \to v_3$

$\sigma_7 : v_1 \to v_2 \to v_3 \to v_4 \to v_5$ $\quad$ $\sigma_8 : v_1 \to v_2 \to v_4 \to v_3 \to v_5$

$\sigma_9 : v_1 \to v_3 \to v_2 \to v_4 \to v_5$ $\quad$ $\sigma_{10} : v_1 \to v_3 \to v_4 \to v_2 \to v_5$

$\sigma_{11} : v_1 \to v_4 \to v_2 \to v_3 \to v_5 \qquad \sigma_{12} : v_1 \to v_4 \to v_3 \to v_2 \to v_5.$

According to the above, we conclude that $P(\sigma_1|v) = P(\sigma_2|v) = \frac{1}{144}$, $P(\sigma_3|v) = P(\sigma_4|v) = P(\sigma_5|v) = P(\sigma_6|v) = \frac{1}{240}$, and $P(\sigma_7|v) = P(\sigma_8|v) = P(\sigma_9|v) = P(\sigma_{10}|v) = P(\sigma_{11}|v) = P(\sigma_{12}|v) = \frac{1}{360}$. So, $P(G_5|v_1) = \frac{34}{720}$. We also have $P(G_5|v_4) = P(G_5|v_3) = \frac{7}{720}$ by symmetric, and $P(G_5|v_2) = \frac{40}{720}$. Note that although $v_1$ is the rumor center we have $P(G_5|v_1) < P(G_5|v_2)$. So we will guess $v_2$ as the source, the example shows that the earlier the end vertex appears in $\sigma_i$ the larger $P(\sigma_i|v)$ it will be.

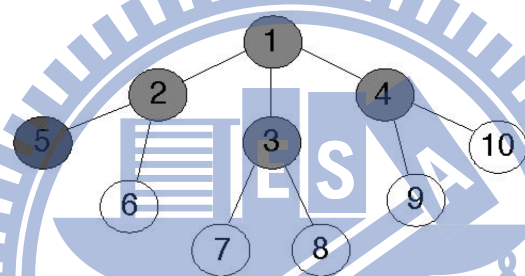



Figure 4.1: In $G_4$, $v_5$ is an end vertex

From this example we conclude that in a finite network, not only the spreading way has influence on the probability of being a candidate of the source but also the distance from the end vertex. Suppose there are two adjacency vertices say $v_1$ and $v_2$ in a finite tree $G_n$ which has an end vertex $v_e$, with $R(v_1, G_n) > R(v_2, G_n)$ and $d(v_2, v_e) = m$. If $v_1$ lies on the path from $v_2$ to $v_e$, then we have $P(G_n|v_1) > P(G_n|v_2)$. Let $\Omega^k_{v_1}$ be the set of all permutation (each permutation represents a spreading way) that started from $v_1$ and $v_e$ is the $(k+1)_{th}$ element in the permutation, then $|\Omega^k_{v_1}| \geq |\Omega^k_{v_2}|$. Since for any $k \in N$, there are $k - m$ choice before the rumor be spread to $v_e$ for $v_1$, but there are only $k-m-1$ choice for $v_2$, in addition, if $v_2$ wants to spread the rumor to $v_e$, the rumor must pass through $v_1$, conversely then not. From the observation, we have the following theorem.

**Theorem 4.1.1.** *Let $G$ be a tree with finite order and $G_n \subseteq G$ is a subtree of $G$ with an end vertex $v_e \in G_n$, then the vertex $v*$ with maximum probability $P(G_n|v)$ is located on the path from rumor center to $v_e$.*

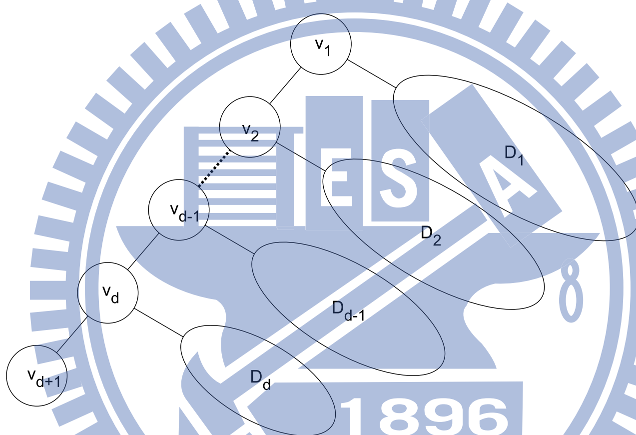



Figure 4.2: $G_n$ with an end vertex $v_{d+1}$

**Proof.** Let $G$, $G_n$ as defined above. $v_r$ is the rumor center in $G_n$ and $v_e$ is the end vertex in $G_n$. Suppose $d(v_r, v_e) = d$. Then, the path $P$ from $v_r$ to $v_e$ is shown in figure 4.2, where $v_1 = v_r$ and $v_{d+1} = v_e$. Define $D_i = \{\ v \in G_n|$ the path from $v$ to $v_e$ contains $v_i\ \}$. Given $i = 1, 2, ..., d$, let $v \in D_i$, then we have $R(v, G_n) \leq R(v_i, G_n)$ (Suppose not, then $v_1$ is not the rumor center, we get a contradiction.) So, each vertex in $D_i$ has less spreading ways than $v_i$ and farther from $v_e$ than $v_i$ does. By the observation above, we conclude that $P(G_n|v_i) \geq P(G_n|v)$. So, the vertex with the maximum probability to spreading the rumor all over $G_n$ is located on the path from $v_1$ to $v_e$. □

Note that, if all leaves are end vertices, that is $G_n = G$, then the probability $P(G_n|v)$ of each vertex in $G_n$ is $\frac{1}{n}$, since any vertex is able to spread the rumor to all vertices in $G$ in $n$ step when $G_n = G$.

## 4.2 Computing $P(G_n|v)$

In the previous section, we know that if there are end vertices in $G_n$ then the rumor center may not be the most possible vertex of being the source. And also for each spreading way the time when the rumor be spread to the end vertex will affect the probability of this spreading way. The following is going to find the vertex with maximum probability $P(G_n|v)$ in $G_n$, where $G_n$ is a path on a $d-regular$ graph $G$.

Suppose $G$ is a $d-regular$ tree with finite order and $G_n \subset G$ is a finite 2-regular subtree of $G_n$ with an end point. Without loss of generality, suppose $n$ is odd and $n = 2t + 1$ for some $t$. Label all vertices on $G_n$ as shown in figure 4.3.

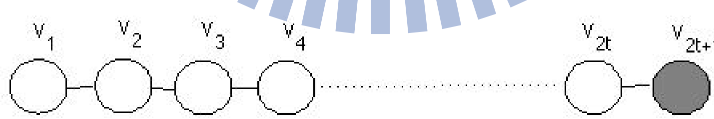



Figure 4.3: $G_n$ with an end vertex $v_{2t+1}$

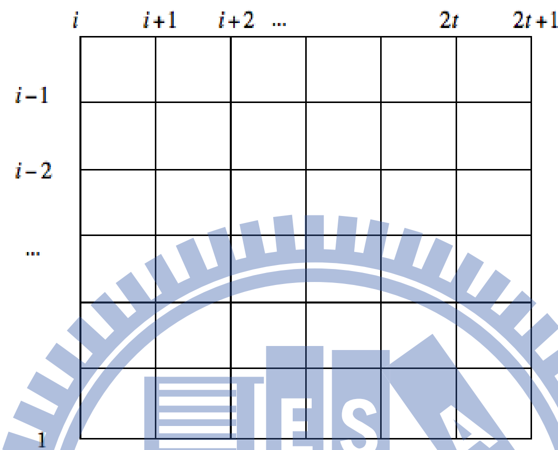
We can get

(4.1)
$$P(G_{2t+1}|v_i) = \sum_{m=0}^{i-1} \binom{m+(2t+i)}{m} \prod_{l=1}^{m+2t+1-i} \frac{1}{z_d(l)+d-1} \cdot \prod_{l=m+2t+1-i}^{2t-1} \frac{1}{z_d(l)+d-2}$$

for $i = 1, 2, ..., 2t$.

And $P(G_{2t+1}|v_{2t+1}) = \prod_{l=1}^{2t} \frac{1}{z_d(l)}$.

We can find that, to compute all spreading way of $v_i$ is equal to find all ways from $i$ to the lower right corner in the figure 4.4.




In (4.1) $m = 0$ means $i$ goes right to the end and then goes down to the end, that is, $v_i$ spread the rumor straight to the end vertex and then spread to the rest of vertices. So $m$ is the parameter related to the time when the rumor meets the end vertex. And the left $\prod$ is the probability of this sprading way when the rumor meet the end vertex and the right one is the probability after the rumor met the end vertex.

From (4.1) we can use computer to compute $P(G_{2t+1}|v_i)$ for all $i$, the following is the numerical result for $d = 2, 3$ and $n = 41$, where $d$ for $d - regular$ and $n = 2 \cdot t + 1$ is the size of $G_n$. In these figures, the $x - axis$ is the node $v_i$ where $i = 1, 2, ..., n$, and the $y - axis$ is the probability $P(v_i = source|G_n)$. See more figures in appendix.

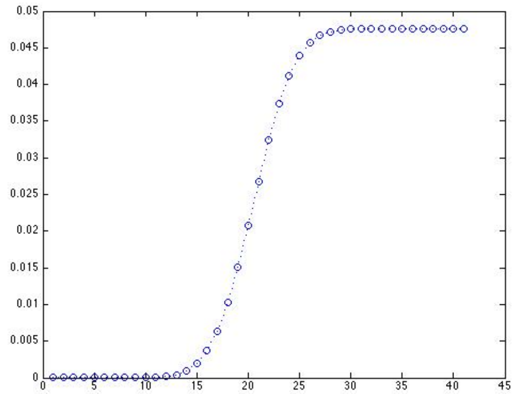


Figure 4.4: d=2,n=41

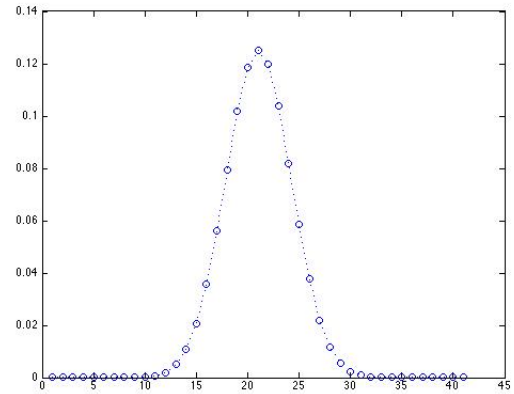


Figure 4.5: d=3,n=41

Finally we will introduce how to compute the $P(G_n|v)$ for each vertex in $G_n$ where $G_n$ is general subgraph of $G$ with an end vertex, $G$ is a finite $d-regular$ tree. When we compute the probability of a spreading way $\sigma$ we should compute the probability before the rumor meets the end vertex and after the rumor meets the end vertex separately just as the case above.

Let $G$ be a finite $d-regular$ tree and $G_n \subseteq G$ is a subtree of $G$ of order $n$ with an end vertex $v_e$. Given $v \in G_n$ and $v \neq v_e$. Let $m_v(t)$ be the number of permutation started from $v$ and $v_e$ is located on the $(t+1)_{th}$ element in the permutation. Then we have

$$(4.2) \qquad P(G_n|v) = \sum_{t=d(v_e,v)}^{n-1} m_v(t) \prod_{i=1}^{t} \frac{1}{d+(i-1)(d-2)} \cdot \prod_{i=t}^{n-2} \frac{1}{d+(i-1)(d-2)-1}$$

The difference between (4.1) and (4.2) is that $m_v(t)$ is easy to compute moreover we can write a formula for it when $G_n$ is a path, but in general tree $m_v(t)$ is unknown. To compute $m_v(t)$ is a hard problem, it is equal to list out all permutation started from $v$.

## 4.3 Conclusion and Future Work

In the infinite case, we can compute all $d \in N$, by counting the number of increasing trees in a $d - regular$ tree. In the finite case, we provided a theorem indicate the "most possible source" lies on the path from rumor center to the end vertex. To compute $m_v(t)$ for all possible $t$ and for all $v$ lies on the path, we have two algorithms. One is by brute force, that is, to list out all possible spreading ways from $v$ and then count $m_v(t)$ for each $t$. Another one is a two-step algorithm, see details in appendix B. The second one is more complicated. Since $m_v(t)$ is related to the shape or structure of $G_n$, it's hard to derive a general formula. The future study would be finding a efficient method to compute $m_v(t)$.

# Appendix A

# More Figures on Detection Probability

The following is the numerical results of $G_n$ which is a $2-regular$ graph with an end vertex. $d$ represents for $G$ is $d-regular$ and $n$ is the order of $G_n$.

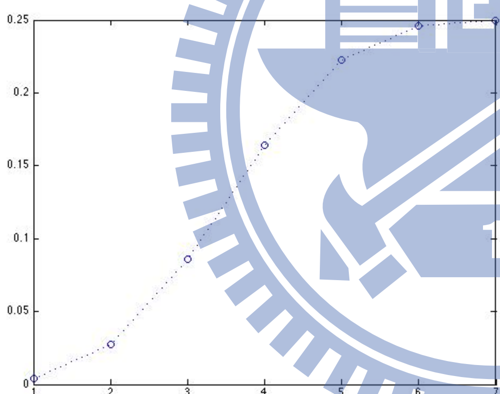


Figure A.1: d=10,n=7

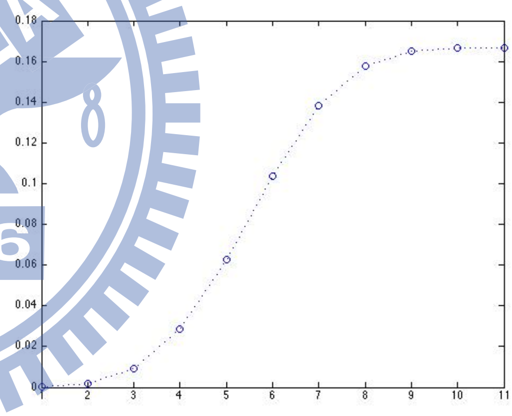


Figure A.2: d=10,n=11

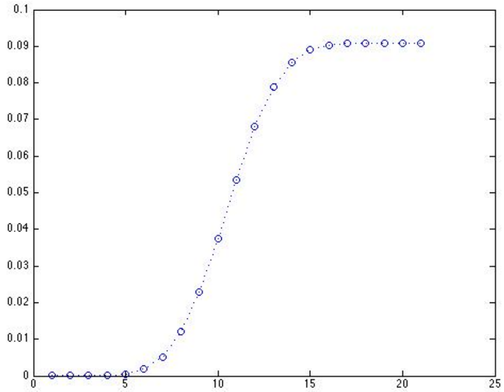


Figure A.3: d=2,n=21

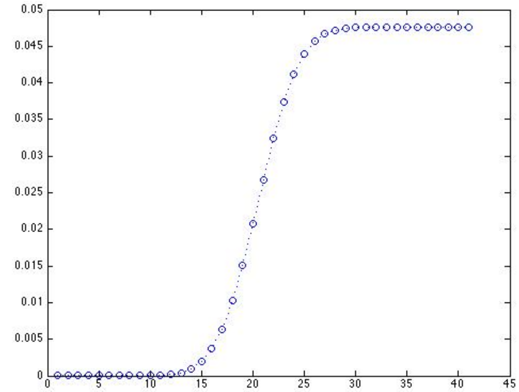


Figure A.4: d=2,n=41

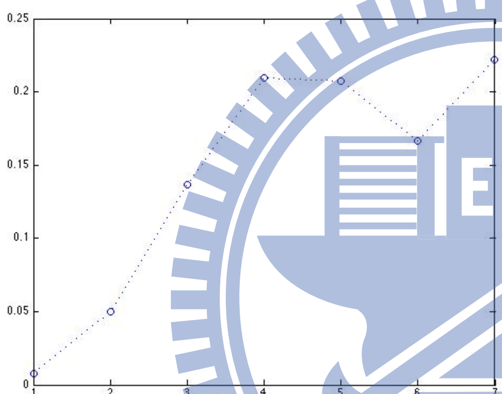


Figure A.5: d=3,n=7

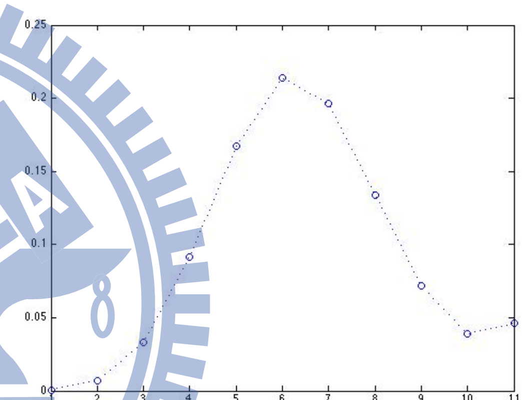


Figure A.6: d=3,n=11

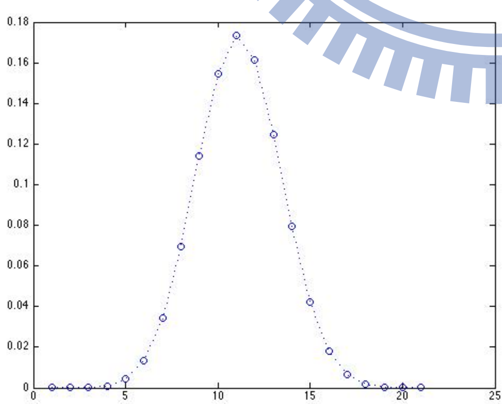


Figure A.7: d=3,n=21

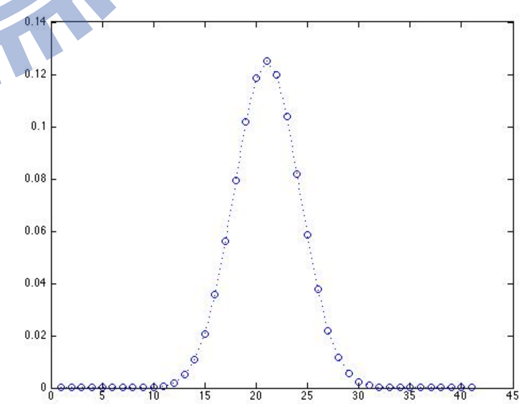


Figure A.8: d=3,n=41

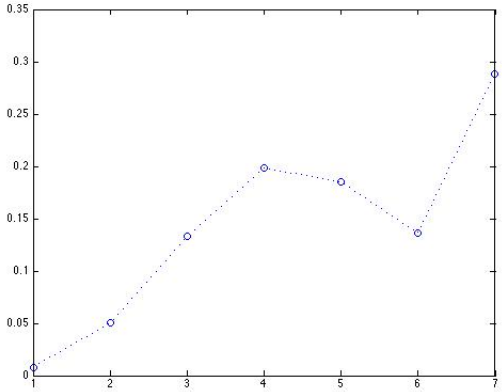


Figure A.9: d=5,n=7

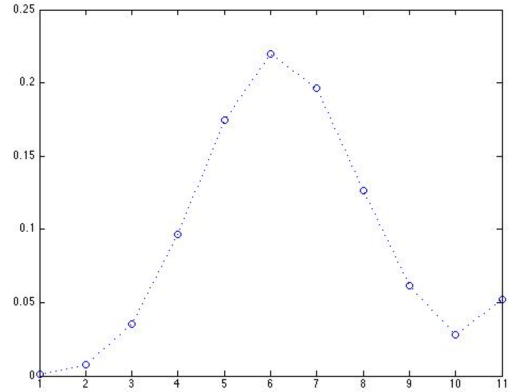


Figure A.10: d=5,n=11

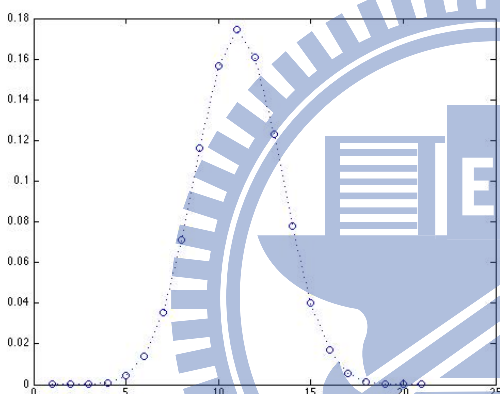


Figure A.11: d=5,n=21

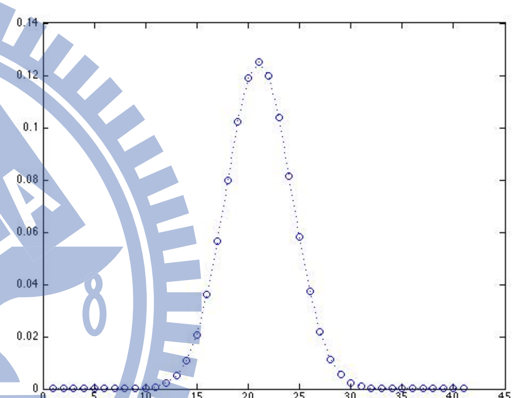


Figure A.12: d=5,n=41

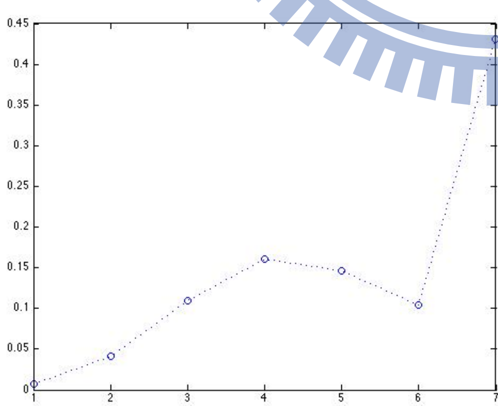


Figure A.13: d=10,n=7

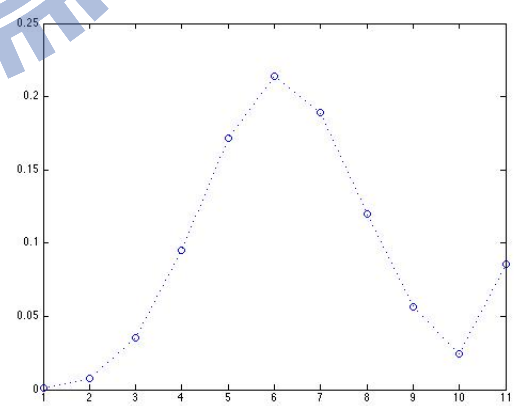


Figure A.14: d=10,n=11

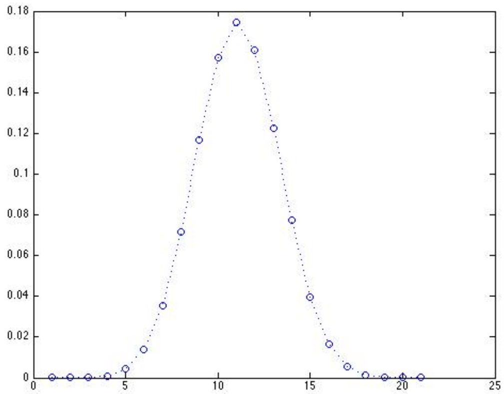



Figure A.15: d=10,n=21

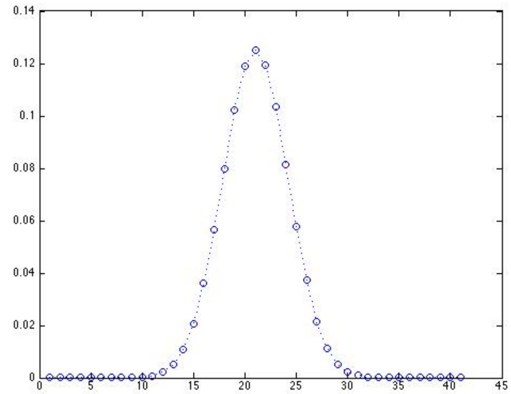



Figure A.16: d=10,n=41

We can observe that, when $n$ is getting larger, the probability $(y-axis)$ is getting smaller. When $d \geq 3$, the influence of end vertex $v_n$ is getting smaller as $n$ gets larger.

# Appendix B

# Two-Steps Method for Computing $m_v(t)$

Let $G$ be a tree, and $G_n$ is a subtree of order $n$ ($n-subtree$ for short). $P$ is the path from root $v$ to the end vertex $v_e$, and suppose $d(v, v_e) = d$. Let $m_v(t)$ be the number of permutations started from $v$ and $v_e$ is located on the $(t+1)_{th}$ element in these permutations. In [6], it provided an algorithm to enumerate the number of $n-subtree$ in a tree. We just need a part of its algorithm, which is to find the family of subtrees whose root are $v$.

First, we should contract all vertices lie on the path $P$ except $v_e$, then we get a new tree $G'$ and $v'$ represents those contracted vertices. Now we can apply the algorithm to find a $(t-d)-subtree$ with root $v'$. For each $(t-d)-subtree$ in $G'$, we can find the number of spreading ways started from $v'$ by using (2.5). The $(t+1)_th$ vertex must be $v_e$.

In the second step, after the rumor spread to $v_e$, we contract all these $t+1$ vertices. We get a new tree $G''$ from $G'$ and a new vertex $v''$ represents those vertices contracted. We can use (2.5) again to find out the number of spreading ways. After combining step one and step two, we can find out $m_v(t)$. See more details about the algorithm for listing all $n-subtree$ in a tree in [6].

# Appendix C

# Proof of Theorem 3.2.3

We consider rumor spreading on $d$-regular graphs. First, we fix some notations.

Let $\tilde{\mathcal{T}}_n$ denote the tree after the rumor has spread to $n$ nodes. We give the vertices labels from the set $\{1, 2, \ldots, n\}$ according to the first time a vertex learns the rumor. Thus, the vertex with label 1 is the source and it is the only node with outdegree $d$ whereas all other nodes have outdegree $d-1$ (if $\tilde{\mathcal{T}}_n$ is drawn as a rooted tree in the usual way).

Next, note that the $d$ subtrees of the source are $d-1$-ary increasing trees (increasing trees are labeled trees with label sequences from the root to any leaf increasing). Set

$$T_n = \text{number of } (d-1)\text{-ary increasing trees}, \qquad T(z) = \sum_{n\geq 1} T_n \frac{z^n}{n!}.$$

It is easy to derive that

$$T(z) = -1 + (1-(d-2)z)^{-1/(d-2)}. \tag{3.1}$$

Next, set

$$\tilde{T}_n = \text{number of possible } \tilde{\mathcal{T}}_n \qquad \textit{and} \qquad \tilde{T}(z) = \sum_{n\geq 1} 2\tilde{T}_n \frac{z^n}{n!}.$$

Then,

$$\tilde{T}_n = \frac{1}{2}\prod_{i=1}^{n} [2+(d-2)(i-1)] \qquad \textit{and} \qquad \tilde{T}(z) = -1 + (1-(d-2)z)^{-2/(d-2)} \tag{3.2}$$

We are interested in $C_n$, the event that the source of $\tilde{\mathcal{T}}_n$ is the rumor center. We have,

$$P(C_n) = 1 - dP(\text{size of one subtree of the source of } \tilde{\mathcal{T}}_n \geq n/2).$$

We fix one subtree of the source of $\tilde{\mathcal{T}}_n$ say $t_1^v$ and denote by $I$ its size. Let $j_i$ represent the size of the $i-th$ subtree of $\tilde{\mathcal{T}}_n$. Obviously,

$$\begin{aligned}
P(I=j) &= \frac{1}{\tilde{T}_n} \sum_{j+j_2+j_3+\ldots+j_d=n-1} \binom{n-1}{j, j_2, j_3, \ldots, j_d} T_j T_{j_2} T_{j_3} \ldots T_{j_d} \\
&= \frac{(n-1)!T_j}{j!\tilde{T}_n} \sum_{j_2+j_3+\ldots+j_d=n-1-j} \frac{T_{j_2}}{j_2!}\frac{T_{j_3}}{j_3!}\cdots\frac{T_{j_d}}{j_d!} \\
&= \frac{(n-1)!T_j}{j!\tilde{T}_n}[z^{n-1-j}](1+T(z))^{d-1} \\
&= \frac{(n-1)!T_j}{j!\tilde{T}_n}[z^{n-1-j}](1-(d-2)z)^{-\frac{d-1}{d-2}} \\
&= \frac{(n-1)!T_j}{j!\tilde{T}_n}(d-2)^{n-1-j}[z^{n-1-j}](1-z)^{-\frac{d-1}{d-2}} \\
&= \frac{(n-1)!T_j}{j!\tilde{T}_n}(d-2)^{n-1-j}\frac{(n-1-j)^{\frac{1}{d-2}}}{\Gamma(\frac{d-1}{d-2})}. \qquad (3.3)
\end{aligned}$$

In the sequel, we need the following standard lemma from analytic combinatorics.

**Theorem C.0.1** (Theorem VI.1 in [2]). *For $\alpha \in \mathbb{C} \setminus \mathbb{Z}_{\leq 0}$ set*

$$f(z) := (1-z)^{-\alpha}.$$

*Then, as $n \to \infty$,*

$$[z^n]f(z) \sim \frac{n^{\alpha-1}}{\Gamma(\alpha)}\left(1 + \sum_{k=1}^{\infty} \frac{e_k(\alpha)}{n^k}\right),$$

*where $e_k(\alpha)$ is a polynomial of degree $2k$.*

**Asymptotics.** We now turn to asymptotics. The following computation is based on $n \to \infty$.

First, applying Theorem C.0.1 to (3.1) gives

$$T_n \sim \frac{n! \cdot n^{\frac{2-d}{d-1}}}{\Gamma(\frac{1}{d-1})}(d-1)^n \qquad (n \to \infty).$$

Similarly, applying Theorem C.0.1 to (3.2) yields

$$\tilde{T}_n \sim \frac{n! \cdot n^{\frac{4-d}{d-2}}}{2\Gamma(\frac{2}{d-2})}(d-2)^n \qquad (n \to \infty).$$

By theorem 2.2.1, we need to compute

$$\sum_{n/2 \le j \le n-1} P(I=j),$$

where $P(I=j)$ is given by (3.3). Therefore, we again use Theorem C.0.1 and the expansions for $T_n$ and $\tilde{T}_n$ from above. This gives

$$\begin{aligned}
\sum_{n/2 \le j \le n-1} P(I=j) &\sim \frac{(n-1)!}{\tilde{T}_n} \sum_{n/2 \le j \le n-1} \frac{T_j}{j!\Gamma(\frac{d-1}{d-2})}(d-2)^{n-1-j}(n-1-j)^{\frac{1}{d-2}} \\
&\sim \frac{(n-1)! \cdot 2\Gamma(\frac{2}{d-2})}{n! \cdot n^{\frac{4-d}{d-2}}(d-2)^n} \sum_{n/2 \le j \le n-1} \frac{\frac{j! \cdot j^{\frac{3-d}{d-2}}}{\Gamma(\frac{1}{d-2})}(d-2)^j}{j!\Gamma(\frac{d-1}{d-2})}(d-2)^{n-1-j}(n-1-j)^{\frac{1}{d-2}} \\
&\sim \frac{2\Gamma\frac{2}{d-2}}{(d-2) \cdot n^{\frac{2}{d-2}}\Gamma(\frac{1}{d-2})\Gamma(\frac{d-1}{d-2})} \sum_{n/2 \le j \le n-1} j^{\frac{1}{d-2}-1}(n-1-j)^{\frac{1}{d-2}} \\
&\sim \frac{2\Gamma\frac{2}{d-2}}{(d-2) \cdot n^{\frac{2}{d-2}}\Gamma(\frac{1}{d-2})\Gamma(\frac{d-1}{d-2})} \cdot n^{\frac{2}{d-2}-1} \sum_{n/2 \le j \le n-1} \left(\frac{j}{n}\right)^{\frac{1}{d-2}-1} \left(\frac{n-1-j}{n}\right)^{\frac{1}{d-2}} \\
&\sim \frac{2\Gamma(\frac{2}{d-2})}{(d-2)\Gamma(\frac{1}{d-2})\Gamma(\frac{d-1}{d-2})} \int_{1/2}^{1} x^{\frac{1}{d-2}-1}(1-x)^{\frac{1}{d-2}} dx \qquad (n \to \infty).
\end{aligned}$$

**Lemma 1.**

$$\int_{1/2}^{1} x^{c-1}(1-x)^c dx = \frac{1}{2}\left(B(c,c+1) - \frac{1}{c \cdot 2^{2c}}\right)$$

**Proof.**

$$\int_0^1 x^{c-1}(1-x)^c dx = \int_0^{1/2} x^{c-1}(1-x)^c dx + \int_{1/2}^{1} x^{c-1}(1-x)^c dx$$

Now, consider the right part $\int_{1/2}^{1} x^{c-1}(1-x)^c dx$, and name it $R$ (the left part is $L$). By integration by part we have

$$R = \frac{1}{c}(1-x)^c x^c \,\Big|_{x=1/2}^{1} + L$$

So

$$L = \frac{\int_0^1 x^{c-1}(1-x)^c dx + \frac{1}{c \cdot 2^{2c}}}{2} = \frac{B(c, c+1) + \frac{1}{c \cdot 2^{2c}}}{2},$$

and

$$R = \frac{1}{2}\left(B(c, c+1) - \frac{1}{c \cdot 2^{2c}}\right).$$

□

Finally, combining everything yields for the detection probability of the source in $d$-regular trees is the following limit

$$\begin{aligned}\lim_{t \longrightarrow \infty} P_d(C_t) &= 1 - d \cdot \sum_{n/2 \le j \le n-1} P(I = j) \\ &= 1 - \frac{d}{2} + \frac{(d-2) \cdot \Gamma\left(\frac{d}{d-2}\right)}{2^{\frac{d}{d-2}} \cdot \Gamma\left(\frac{1}{d-2}\right)\Gamma\left(\frac{d-1}{d-2}\right)}\end{aligned}$$

# References

[1] Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest and Clifford Stein, Sedgewick, *Introduction to Algorithm 3th ed*, The MIT Press, Cambridge, Massachusetts London, England 2009.

[2] Philippe Flajolet, Robert Sedgewick, *Analytic Combinatorics*, Cambridge University Press, 2009.

[3] 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.

[4] Zi-Hui Lee, "A Mathematical Model for Finding the Culprit Who Spreads Rumors", 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.

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

[6] Kunihiro Wasa, Yusaku Kaneta, Takeaki Uno, and Hiroki Arimura, "Constant Time Enumeration of Bounded-Size Subtrees in Trees and Its Application", IST, Hokkaido University, Sapporo, Japan.

[7] Yang Wang, Deepayan Chakrabarti, Chenxi Wang, Christos Faloutsos, "Epidemic Spreading in Real Networks: An Eigenvalue Viewpoint", Carnegie Mellon University.

[8] D.B. West, *Introduction to Graph Theory* , second edition, Prentice Hall, Upper Saddle River, New Jersy, 2011.

[9] Bohdan Zelinka,"Medians and Peripherians of Trees', Archivum Mathematicum, Vol. 4 (1968), No.2, 87–95.