

Optimal Detection of Influential Spreaders in Online Social Networks

Chee Wei Tan[†], Pei-Duo Yu[†], Liang Zheng^{*}, Chun-Kiu Lai[†], Wenyi Zhang[‡] and Hung-Lin Fu^{*}
 City University of Hong Kong[†], Princeton University^{*},
 The University of Science and Technology of China[‡], National Chiao Tung University^{*}
 {cheewtan,peteryu,cklai24-c}@cityu.edu.hk, liangz@princeton.edu, wenyizha@ustc.edu.cn,
 hlfu@math.nctu.edu.tw

Abstract—The wide availability of digital data in online social networks such as the Facebook offers an interesting question on estimating the influence of users based on the user interaction over time. For example, Facebook users use the Facebook “Like” button to endorse a digital object (e.g., a post or picture) posted by other user, and this online interaction indicates a level of influence among users who share similar opinions or disposition. A useful application is that a business entity with a Facebook presence can find a number of Facebook users to spread the word of new commercial products or to endorse some form of digital marketing message. In this paper, we study this estimation problem as finding a set of seeding nodes that are influential in spread in a graph. For online social network applications, this graph is modeled by Facebook users’ interactions such as the *Like*’s activity in the Facebook News Feed. This past record of snapshot observations is used in a maximum-likelihood estimation problem to identify a group of users whose role is instead to motivate a forward-engineering perspective of spreading over time to maximize the reach of a digital content. This means that this group of influential spreaders are *most likely* to maximize the reach a new digital object when it shows up on the Facebook News Feed. Lastly, we describe the performance of our algorithm using synthetic dataset and a software implementation using the Facebook Graph.

I. INTRODUCTION

The wide availability of digital data in online social networks and the enormous user pool offers an interesting question on estimating the influence of users based on the user interaction over time. Both this online user interaction and the online social network itself give rise to an interaction network that represents a fundamental medium for spreading and diffusion of various types of behavior and information. A prominent example is the online social network Facebook whereby the *News Feed* of a Facebook user constantly provides digital contents (e.g., user status updates, photos, videos, links) that are viewable by other Facebook users. In addition, the *News Feed* captures every online interaction from commenting on photos to clicking the Facebook *Like* endorsement button that again are viewable by other Facebook users. This digital interaction provides an impetus to spread information like a

word-of-mouth engine. Here, the spreading process increases the susceptibility of other users to the same; this results in the successive spread of a digital content from a few users to many more.

It is reasonable to expect that a particular Facebook user’s digital content that has garnered many other Facebook users to act upon it, whether by the Facebook *Like* endorsement button or other digital mechanism, is *likely* to attract a similar level of attention when there are future postings of digital contents of similar nature. This is likely to happen since the digital interaction (e.g., the Facebook “*Like*” button) captures the desire to share similar opinions or disposition, and this often comes from Facebook users who are either socially close to the particular Facebook user (as a Facebook Friend) or somebody else connected to the Facebook Friend. At the same time, it captures the level of connectivity in the online social network itself, i.e., those Facebook users are *likely* to have a substantial number of Facebook Friends. This can find useful application such as when the particular Facebook user is a business entity that maintains a Facebook presence and wants to find (from its past record of News Feed) a number of other Facebook users who are deemed influential enough to spread the word of new commercial products or to endorse some form of digital marketing message. This form of spreading is especially useful for viral marketing [1].

In this paper, we study how this digital interaction in online social networks can be used to maximize the reach of digital contents to as many users as possible. The motivation of this work is similar to prior work that analyzes the effect of viral spreading and the maximization of influence [2]–[5]. The goal in our paper is to find a set of users (seeding nodes in a graph) who can maximize the reach of a digital message, e.g., to generate the most number of Facebook *Like* endorsement for advertising purpose. In this paper, Facebook users are modelled by nodes in a graph, and their associations of online interaction activity (such as clicking the Facebook *Like* button) are modeled as edges in the graph. Snapshot observations of past records in the Facebook News Feed that captures the online interaction activity is used to find this group of users (i.e., the influential spreaders). Thus, the statistical inference of influential spreaders motivates a forward-engineering perspective of spreading over time (that are in turn recorded by the Facebook News Feed as past data that can be reused).

The research has been supported by the Doctoral Program of Higher Education (SRFDP) and Research Grants Council Earmarked Research Grants (RGC ERG) Joint Research Scheme through Specialized Research Fund 20133402140001, National Natural Science Foundation of China through Grant 61379003, the Research Grants Council of Hong Kong under Project M-CityU107/13.

The main contribution of this paper are summarized as follows:

- We provide a problem formulation to this statistical inference of influential spreaders in the online social network as a maximum-likelihood estimation problem over a graph whose solution is the subset of users (nodes in the graph) who are *most likely* to maximize the spread of a digital object to as many users as possible. This problem formulation is based on the rumor source detection problem in the literature using a network centrality approach.
- We provide a graph convexity characterization of the optimal solution, i.e., *most influential user*, using this network centrality approach by showing that this *most influential user* is equivalent to the mass center (also known as the centroid) in graph theory.
- We propose a message passing algorithm to rank the users (ranked in terms of the network centrality) when the graph is a tree network to identify a subset of influential users. This algorithm can also be extended to a general graph by first applying the breadth-first-search (BFS) algorithm to yield a BFS tree graph.
- We describe the performance of our algorithm using synthetic dataset and a software implementation using the Facebook Graph.

The rest of this paper is organized as follows. We describe the related work of the rumor source detection problem in the literature in II. In Section III, we describe a basic spreading model. In Section IV, we describe the graph convexity results and a message passing algorithm for ranking the nodes in the graph. In Section V, we evaluate the performance of our message passing algorithm on a synthetic data-set and describe the software implementation using Facebook graph. Finally, we conclude the paper in Section VI.

II. RELATED WORK

Epidemic-like spreading is an important network science topic extensively studied in the literature [6]. In particular, rampant spreading of malicious information has been identified as a major cyber-security challenge in the networks [7]. Indeed, detection of these malicious information sources has many important applications, e.g., rooting out a computer virus or rumor spreading in the Internet or an online social network respectively. Given a snapshot observation of the nodes in the network possessing the malicious information, how to reliably identify the source of the spreading? This is in general a challenging combinatorial hard problem that is complicated by the dynamics of the spreading and the underlying network. Adding to the computational barrier is the network size and its topology that influence the design of a detection algorithm with practical computational complexity.

In the recent seminal work in [8], [9], Shah and Zaman formulated this as a maximum likelihood estimation problem assuming the Susceptible-Infectious (SI) model in [10] for spreading. For ease of exposition, we shall call the nodes possessing the malicious information to be *infected nodes* and those that do not to be *susceptible nodes*. The authors

in [8], [9] proposed the notion of *rumor centrality* to solve this problem exactly for degree-regular tree graphs. The idea of rumor centrality is to assign a number to each infected node and to find the node with the maximum value called the *rumor center*, which is the optimal solution of the maximum likelihood estimation. In particular, the rumor center can be obtained using message passing algorithms [11]. The detection performance can be quantified asymptotically, i.e., when the number of infected nodes becomes very large.

For general graph topology, even though the maximum likelihood estimation is still an open problem, it has been shown that suboptimal heuristics based on the rumor centrality (e.g., see the breadth-first search heuristic in [8], [9]) perform reasonably well. Since work in [8], [9], there has been various extensions. For example, the generalization in [9] to random trees, extensions in [12] with suspect sets, extension in [13] to multiple source detection and extension in [14], [15] to detection with multiple snapshot observations.

In this paper, we consider a new probabilistic approach to the rumor source detection using the SI model similar to that in [8], [9]. The idea is to provide a probabilistic characterization to the *rumor boundary* of the observed graph data, which in turn leads to a probabilistic approach of source estimation. Furthermore, for a tree graph, we show that the source estimation can be efficiently computed using a distributed message passing algorithm. This provides an alternative to the discrete approach of rumor centrality approach in [8], [9].

The optimal solution leads to a network centrality approach that orders the nodes by influence. There are other related work. In [16], the authors proposed a message passing algorithm for computing a proposed harmonic influence centrality that measures the influence of nodes on the average opinion in networks. Our work is also closely related to the rumour source detection problem such as in [17] for susceptible-infectious-susceptible spreading model. The authors in [18] analysed a probabilistic characterization of the rumor graph boundary that can be used for statistical inference

III. SYSTEM MODEL

In this section, we describe a basic spreading model known as the Susceptible-Infectious (SI) model. This model has also been used for the rumor detection problem in [8], [9].

A. Rumor Spreading Model

We consider an infinite network modeled as an undirected graph $G = (V, E)$, where $V = \{v_1, v_2, \dots\}$ is a countably infinite set of nodes and E is the set of edges of the form (i, j) for nodes v_i and v_j in V . The degree of a node v_i is the number of its neighbors denoted by d_i . We assume that initially at time $T = 0$ there is only one rumor source $v^* \in V$.

The rumor spreading model process is modeled by the well-known susceptible-infected (SI) model, which is a variant of the susceptible-infected-recovered (SIR) model for infectious disease spreading [10]. In this model, there are two types of nodes: (i) susceptible nodes that are capable of being infected; and (ii) infected nodes that can spread the rumor to their immediate neighbors. Note that spreading occurs in a

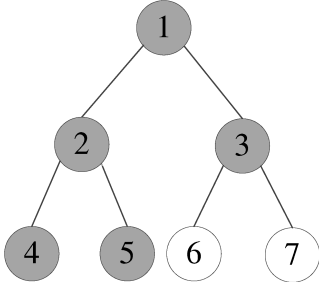


Fig. 1. An example tree network of rumor graphs and its underlying graphs. The grey and white circles in this figure represent respectively the infected nodes and susceptible nodes. Note that the underlying graph can be infinity large, which is omitted due to the limited page space.

cascading manner, i.e., once a susceptible node gets infected by its neighbor, it retains the rumor forever and in turn may infect its other susceptible neighbors, i.e., when $(i, j) \in E$. Let τ_{ij} be the spreading time for an infected node v_i to infect its susceptible neighbor v_j for all $(i, j) \in E$, where τ_{ij} are mutually independent and have exponential distribution with parameter λ . Without loss of generality, we assume $\lambda = 1$.

B. Rumor Source Estimator

Let us suppose that the rumor originates from a node $v^* \in V$ at time $t = 0$ and spreads in the network G . Then, at time $t = T$, we observe the network G and find N infected nodes, which collectively constitutes a rumor graph, denoted by G_N . Let \mathcal{I} be the set of infected nodes, and $N = |\mathcal{I}|$ represents the cardinality of \mathcal{I} , i.e., the number of infected nodes in G_N . Obviously, G_N is a connected subgraph of G . In this paper, we regard G as the underlying graph and G_N as the rumor graph. Our goal is to infer the rumor source based on the knowledge of network topology and the snapshot observation of the infected nodes. In the following, we construct an estimator to identify a node \hat{v} as the rumor source based on the observation of G_N . The ML estimator of v^* that maximizes the correct detection probability is given by

$$\hat{v} \in \arg \max_{v \in G_N} P(G_N | v), \quad (1)$$

where $P(G_N | v)$ is the probability of observing G_N supposing that v is the rumor source. Note that ties are broken uniformly at random for (1). Note that the solution of (1) may not be unique.

IV. EQUIVALENCE OF RUMOR CENTRALITY WITH MASS CENTER

For a tree graph G_n , finding the rumor center, i.e., $\max_i R(v_i, G_n)$ where $v_i \in G_n$, can be performed by a message-passing algorithm that recursively computes the rumor centrality for each node starting from the leaves [8]. In this section, we give a new characterization of the rumor center by showing its equivalence to the *mass center* (see, e.g., [19]) in graph theory.

Let G_n be a rooted tree with the root at v_r where $v_r \in G_n$. For any vertex v in the rooted tree G_n , a *parent* of v is its neighbor on the path connecting the vertex v and v_r . The children of v are its other neighbors, and we let $\text{child}(v)$ denote

the set of children nodes of the vertex v . If v is a leaf, $\text{child}(v)$ is an empty set. A branch T_v^r of this rooted tree is a subtree with its root at v and we let t_v^r denote the order of T_v^r , i.e., the size of T_v^r in terms of the maximum number of children of v allowed.

Now, suppose that v_r is the rumor source culprit and the spreading has initiated, i.e., $G_1 = v_r$. Then, in G_2 , this second infected vertex may be any child of v_r . Since there are $d(v)$ vertices in $\text{child}(v)$ for any of this vertex, say u_i , where $i = 1, 2, 3, \dots, d(v)$, we thus have [8]:

$$R(v, G_n) = \frac{(n-1)!}{t_{u_1}^v! \cdot t_{u_2}^v! \cdot \dots \cdot t_{u_{d(v)}}^v!} \cdot \prod_{i=1}^{d(v)} R(u_i, T_{u_i}^v). \quad (2)$$

This can be expanded recursively from the root v_r to all the leaves of G_n to yield [8]:

$$R(v, G_n) = n! \cdot \prod_{u \in G_n} \frac{1}{t_u^v}. \quad (3)$$

Now, consider two adjacent vertices u and 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 the order of a subtree T_u^v with v being the rumor source and u as the root containing all the children of the tree. By using this recursion, it can be established that:

$$\frac{P(u|G_n)}{P(v|G_n)} = \frac{R(u, G_n)}{R(v, G_n)} = \frac{t_u^v}{n - t_u^v}, \quad (4)$$

which leads to the following result (see Proposition 1 in [8]).

Theorem 1: Given a tree G_n with n vertices, $v \in G_n$ is a rumor center if and only if

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

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

In words, this result characterizes the rumor center in terms of the sizes of its local subtrees.

Next, let us introduce a graph-theoretic notion of G_n that is useful to provide another perspective of the rumor center by relating it to Theorem 1 above. Let us define the weight of a vertex v in G_n by

$$\text{weight}(v) = \max_{c \in \text{child}(v)} t_c^v.$$

The vertex of G_n with the minimum weight is called the *mass center* of G_n [19]. Let us also define the *distance centrality* of $v \in G_n$ as $D(v, G_n) = \sum_{j \in G_n} d(v, j)$, where $d(v, j)$ is the distance (in terms of hop) between vertices v and j [8]. The vertex in G_n with the minimum distance centrality is called the *distance center*. We have the following result.

Theorem 2: Let G_n be a general tree graph and v is a vertex in G_n . Then, the following statements are equivalent:

- 1) The vertex v is a rumor center of G_n and also a distance center of G_n (proved in [8]).
- 2) The vertex v is a mass center (or centroid) of G_n .

It has been established in graph theory that a tree has either exactly one, or exactly two mass centers joined by an edge (see, e.g., [19]). This implies that, by using Theorem 2, there are *at most two* rumor centers and this scenario with two

Algorithm 1 Message Passing Algorithm

Choose a root node $v \in G_N$.

for $v_i \in G_N$ **do**

if v_i is a leaf **then**

 Calculate its probability according to (??).

else if v_i is not a leaf or the root **then**

 Pass the product of the values received from all its children nodes to its parent node.

else

 The product of the values received from the root's children nodes is the ML for this node.

end if

end for

Algorithm 2 Message Passing Algorithm to computer Mass Center

Let $T(V, E) = G_n$. For all $v \in G_n$, initialize $\text{weight}(v) = 1$.

while Order of $T > 1$ **do**

 Select a leaf a with minimum weight; $V \leftarrow V \setminus \{a\}$.

if $b \in V$ is adjacent to a **then**

$\text{weight}(b) \leftarrow \text{weight}(b) + \text{weight}(a)$;

$E \leftarrow E \setminus \{a, b\}$.

end if

end while

rumor centers happens only when the maximum branch size is exactly $n/2$. Furthermore, that the mass center and the distance center coincides has been pointed out in [19].

Now, a practical implication of Theorem 2 is that this rumor center for a given tree G_n can be found using alternative algorithms (such as those proposed in [19], [20]) and the related ranking algorithms in [21], [22] that are based on the notion of mass center to calculate $R(v, G_n)$. All these alternative algorithms have computational time complexity $O(n)$ similarly to the message passing algorithm in [8].

A. Message Passing Algorithm to Rank Centrality

In order to look for the source of a tree G_N , we need to first calculate the MLs for all the infected nodes. We assume that all the nodes have a full knowledge of its distance to all the other nodes in the tree and its own degree. We propose a distributed message passing algorithm to compute the ML estimator of a node $v \in G_N$ by passing from the leaf node their probabilities calculated in (??) to their parents. The parent nodes then pass the product of all the probabilities collected from their children nodes and pass the result to their parent nodes. This process iterates until all the children nodes of the root pass the result to it. This procedure is summarized in Algorithm 1.

To rank the rumor centrality of nodes in G_N , we can either rank the nodes by rumor centrality, distance centrality or mass centrality.

Theorem 3: Let G_n be a general tree with n vertices, $u, v \in G_n$ are two adjacent nodes where u, v are not rumor center. Then, the following statements are equivalent:

- 1) $R(v, G_n) \geq R(u, G_n)$.

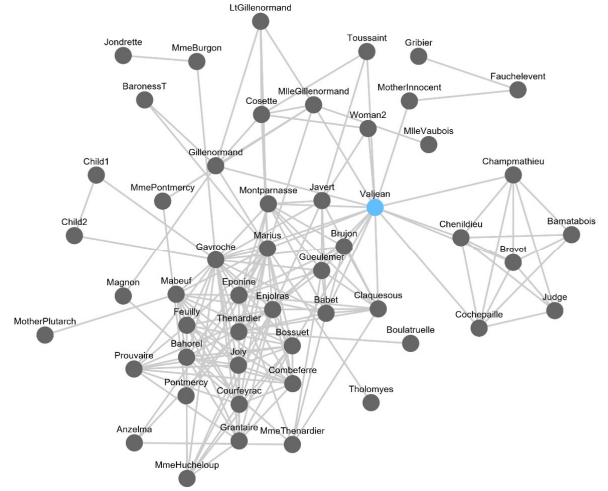


Fig. 2. The Les Misérables data-set from *The Stanford GraphBase: A Platform for Combinatorial Computing* [23] that depicts the network of fictional characters in Victor Hugo's 1862 novel Les Misérables. The message-passing algorithm identifies Jean Valjean, the protagonist of this novel, as the user with the maximum mass centrality.

- 2) $D(v, G_n) \leq D(u, G_n)$.
- 3) $\text{weight}(v) \leq \text{weight}(u)$.

The result of the above theorem also lead to the result of Theorem 2. And now we have three ways to compare the centrality of any two adjacent nodes. Suppose we knew the rumor centrality of each node in G_n and we found the rumor center by Algorithm 1. The following algorithm help us to pick top-k rumor centrality for a given input integer k . Let K denote the set of nodes with top-k rumor centrality, and $CHS = \{u | u \in \text{child}(K)\}$.

Algorithm 3 Algorithm for top-k Centrality Nodes

Input k, G_n with rumor center v_1 and rumor centrality of each node

Set $K = \{v_1\}$, $CHS = \phi$

for $i = 1..k - 1$ **do**

$CHS = CHS + \{u | u \in \text{Child}(v_i)\}$

 set $v_{i+1} = \max\{CHS\}$

$K = K + \{v_{i+1}\}$

end for

For any given integer k , the above algorithm takes $O(k^2 D)$ times to pick out the top-k nodes, where D is the maximum degree in G_n .

V. NUMERICAL SIMULATION

Each node is a fictional character in Victor Hugo's 1862 novel Les Misérables. There is an edge between two characters if they appear in the same chapter.

Javert, Gavroche, Cosette

VI. CONCLUSION

In this paper, we formulated and studied an influence maximization problem to identify the users who are *most likely* to Like a given digital object when it shows up on the Facebook News Feed. Snapshot observations of past records

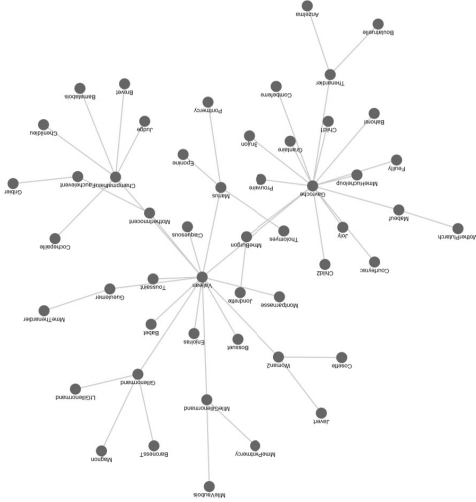


Fig. 3. An example tree network of rumor graphs and its underlying graphs.

in Facebook News Feed that link users by *Like*'s activity is first used to identify a group of users who share similar interests. A maximum-likelihood estimation problem is then solved to identify the users who are *most likely* to *Like* a given digital object when it shows up on the Facebook News Feed. This is motivated by a forward engineering approach to the rumor source detection problem for the SI model. The optimal solution of this problem leads to a network centrality approach that orders the nodes by influence. Lastly, we evaluate the performance of our software implementation in utilizing the Facebook Graph search functionalities to quantify the influence spread.

It can be of immense practical value to know the relative level of influence spread among these people. Given a particular post or picture, what is the expected number of Facebook users who *Like* it when it shows up on news feed?

APPENDIX

A. Proof of Theorem 2

Let G_n be a tree of size n and $v \in G_n$. Let us prove the direction (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 $\sum_{s \in G_n} d(u, s)$ as described in the 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 $t_u^v > t_v^u$. This implies that v is not a distance center.

Let us prove the direction (2 \Rightarrow 3) First, we need the following 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).

Let us prove the direction (3 \Rightarrow 1) Suppose v is a mass center, then each of all its branches is of order $\leq n/2$. This implies that 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) $|R| + t_v^u \leq n/2$, and $t_v^u > n/2$
- (2) $(\sum_{s \in T_v^u} d(v, s) + \sum_{s \in T_u^v} d(v, s)) - (\sum_{s \in T_v^u} d(u, s) + \sum_{s \in T_u^v} d(u, s)) = l \cdot (t_v^u - t_u^v)$
- (3) $|\sum_{s \in R} d(v, s) - \sum_{s \in R} d(u, s)| \leq l \cdot |R|$.

Combining 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. \square

B. Proof of Theorem 3

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

(1 \Rightarrow 2)

Suppose $R(v, G_n) \geq R(u, G_n)$, we have $D(v, G_n) = D(u, G_n) - t_v^u + t_u^v$ and $t_v^u \geq t_u^v$ so we conclude that $D(v, G_n) \leq D(u, G_n)$.

(2 \Rightarrow 3)

Suppose $D(v, G_n) \leq D(u, G_n)$, we have $D(v, G_n) - D(u, G_n) = t_u^v - t_v^u \leq 0$. This implies $t_u^v \leq t_v^u$.

Note that $weight(u) = t_u^v$, (if not, then there is an branch of u with size larger then t_u^v implies $t_u^v \geq t_v^u$, which is a contradiction.)

So we have $weight(u) = t_u^v \geq weight(v)$.

(3 \Rightarrow 1)

Suppose $weight(v) \leq weight(u)$, and note that $weight(u) = t_u^v$.

Since u is not the rumor center, we have $t_u^v > n/2$ and so $t_v^u \leq n/2$, this implies $R(v, G_n) \geq R(u, G_n)$.

REFERENCES

- [1] A. Sela, I. Ben-Gal, A. S. Pentland, and E. Shmueli, "Improving information spread through a scheduled seeding approach," *Proc. of IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, 2015.
- [2] D. Kempe, J. Kleinberg, and E. Tardos, "Maximizing the spread of influence through a social network," *Proc. of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, 2003.
- [3] M. E. J. Newman and M. Girvan, "Finding and evaluating community structure in networks," *Physics Review E*, vol. 69, 2004.
- [4] F. Chierichetti, J. Kleinberg, and A. Panconesi, "How to schedule a cascade in an arbitrary graph," *SIAM Journal on Computing*, vol. 43, no. 6, p. 19061920, 2014.
- [5] P. Shakarian, S. Eyre, and D. Paulo, "A scalable heuristic for viral marketing under the tipping model," *Social Network Analysis and Mining*, vol. 3, no. 4, p. 12251248, 2013.
- [6] A. Ganesh, L. Massoulie, and D. Towsley, "The effect of network topology on the spread of epidemics," *Proc. IEEE INFOCOM*, 2005.
- [7] S. Wasserman and K. Faust, *Social Network Analysis: Methods and Applications*. Cambridge Univ. Press, 1994.
- [8] D. Shah and T. Zaman, "Rumors in a network: who's the culprit?" *IEEE Trans. Inf. Theory*, vol. 57, no. 8, pp. 5163–5181, 2011.
- [9] —, "Rumor centrality: a universal source detector," *Proc. of 12th ACM SIGMETRICS/PERFORMANCE joint int. conf. on measurement and modeling of computer systems*, 2012.

- [10] N. T. J. Bailey, *The Mathematical Theory of Infectious Diseases and its Applications*, 2nd ed. Griffin, 1975.
- [11] D. J. C. Mackay, *Information Theory, Inference and Learning Algorithms*, 1st ed. Cambridge University Press, 2003.
- [12] W. Dong, W. Zhang, and C. W. Tan, "Rooting out the rumor culprit from suspects," *Proc. of IEEE ISIT*, 2013.
- [13] W. Luo, W. P. Tay, and M. Leng, "Identifying infection sources and regions in large networks," *IEEE Transactions on Signal Processing*, vol. 61, no. 11, pp. 2850–2865, 2013.
- [14] Z. Wang, W. Dong, W. Zhang, and C. W. Tan, "Rumor source detection with multiple observations: Fundamental limits and algorithms," *Proc. of ACM SIGMETRICS*, 2014.
- [15] —, "Rooting out rumor sources in online social networks: The value of diversity from multiple observations," *IEEE Journal of Selected Topics in Signal Processing*, vol. 9, no. 4, pp. 663–677, 2015.
- [16] L. Vassio, F. Fagnani, P. Frasca, and A. Ozdaglar, "Message passing optimization of harmonic influence centrality," *IEEE Transactions on Control of Network Systems*, vol. 1, no. 1, p. 109120, 2014.
- [17] Z. Wang, W. Zhang, and C. W. Tan, "On inferring rumor source for SIS model under multiple observations," *Proc. of IEEE Digital Signal Processing*, 2015.
- [18] L. Zheng and C. W. Tan, "A probabilistic characterization of the rumor graph boundary in rumor source detection," *Proc. of IEEE Digital Signal Processing*, 2015.
- [19] B. Zelinka, "Medians and peripherians of trees," *Archivum Mathematicum*, vol. 4, no. 2, pp. 87–95, 1968.
- [20] Y. Métiévier and N. Saheb, "Probabilistic analysis of an election algorithm in a tree," *Proc. of 19th International Colloquium on Trees in Algebra and Programming, Lecture Notes in Computer Science*, vol. 787, pp. 234–245, 1994.
- [21] S. Zaks, "Optimal distributed algorithms for sorting and ranking," *IEEE Transactions on Communications*, vol. C-34, no. 4, p. 376379, 1985.
- [22] O. Gerstel and S. Zaks, "A new characterization of tree medians with applications to distributed algorithms," *Networks*, vol. 24, no. 1, p. 2329, 1994.
- [23] D. E. Knuth, *The Stanford GraphBase: A Platform for Combinatorial Computing*. Addison-Wesley, Reading, MA, 1993.