Rumor Source Detection in Unicyclic Graphs

Pei-Duo Yu†, Chee Wei Tan† and Hung-Lin Fu*
City University of Hong Kong†, National Chiao Tung University*
{peiduoyu,cheewtan}@cityu.edu.hk, hlfu@math.nctu.edu.tw

***Abstract*—Detecting information source in viral spreading has important applications such as to root out the culprit of a rumor spreading in online social networks. In particular, given a snapshot observation of the network topology of nodes having the rumor, how to accurately identify the initial source of the spreading? In the seminal work by Shah and Zaman in 2011, this problem was formulated as a maximum likelihood estimation problem and solved using a rumour centrality approach for graphs that are cycle-free. This however is optimal only for degree-regular trees, and even the special case of a single cycle is an open problem. In this paper, we address the maximum likelihood estimation problem by an generalized rumor centrality for spreading in graphs with cycles. We derive analytical characterization of the optimal solution and a polynomial-time algorithm to solve the problem.**

I. Introduction

Networks represent a fundamental medium for the spreading and diffusion of information. Network nodes are said to be infected when they possess this information, and network topologies govern how the spreading processes increase the susceptibility of other nodes to be infected leading to the successive spread of information from a few initial nodes to a much larger set. An example of such a viral spreading phenomenon is rumor spreading in online social networks. From a cybersecurity enforcement viewpoint, this begs the question of detecting and rooting out malicious information sources in a reliable and efficient manner [1]–[5]. In particular, given a snapshot observation of the infected nodes, who is the culprit source of the rumor spreading?

In a recent seminal work in [6], Shah and Zaman formulated this as a maximum likelihood estimation problem, and proposed *rumor centrality*, a form of network centrality, to solve this problem for degree-regular tree graphs *assuming that the underlying graph has no cycle*. This means that the resultant observation graph is cycle free. The infected node with the most number of ways to spread to other nodes is the *rumor center* that coincides with the maximum likelihood estimate. This rumor centrality approach was subsequently extended to various problem settings, e.g., extension in [4], [5], [7] to random trees, extension in [5], [8] to constrained observations, extension in [9] to multiple source detection, extension in [10] to detection with multiple snapshot observations, and extension in [11] to prove that when the number of infected vertices is large enough, the probability of the source vertex not in a confidence set is less than a given error ϵ. In [12], [13], the authors established its equivalence to the graph theoretic *centroid*.

There is however a key limitation in the rumor centrality approach. A main modeling assumption in all the aforementioned work is that the underlying graph (i.e., number of susceptible nodes) is cycle-free. This is never true in general of practical real-world networks where inter-connections are diverse and the presence of cycles cannot be ignored, and makes this constrained maximum likelihood estimation problem a much harder combinatorial problem. In essence, the cycle effects allow the dynamical spreading process to spread thrugh multiple alternate path thus increasing the likelihood that nodes near the cycle to be infected. Hence, the number of cycles and their location can significantly shape spreading and therefore the estimation performance. To be exact, existing algorithms in the literature, e.g., [6]–[10], are no longer optimal *even with the presence of a single cycle* in degree-regular pseudo-tree graph.

Rumor source detection over graphs with cycles is clearly more challenging, but is more realistic and also significantly generalizes all previous work [6]–[10] that mostly use the breadth-first search heuristic for graphs with cycles. Thus, another key issue is the design of optimal inference algorithms that work for a variety of network topologies. Finding the maximum likelihood estimate as opposed to a suboptimal heuristic to detect the source is important. A focus in this paper is thus to *extend* the rumor centrality and propose optimal algorithm design for pseudo-tree graph, leveraging the theorem derived in this paper, that have practical computational complexity.

A. Our Contributions

The main contributions are summarized as follows:

- This paper considers a more general underlying network, i.e., network with cycles. We analyze the graph with a single cycle, and characterize the optimal solution of the maximum likelihood estimation problem.
- For a degree-regular graph with cycles say G_n, we prove that under certain condition, the rumor center of G_n is equal to the rumor center any spanning tree of G_n.
- We propose a polynomial-time algorithm to find the *rumor center* on the graph with a single cycle.

II. Preliminaries of Rumor Centrality

We model an online social network of nodes by an undirected graph $G = (V, E)$, where the set of vertices V represents the nodes in the underlying network, and the set of edges E represents the links between the nodes. Following [6], we use the Susceptible-Infectious (SI) model in [14] to

TABLE I
TABLE OF NOTATION

Notation	Remark
G	The underlying graph (network of susceptible nodes)
G_n	A subgraph of G of n nodes infected by rumor
$v^\star$	Actual rumor source
$\hat{v}$	Maximum likelihood estimator for $v^\star$
v_c	*Rumor center* of G_n
v_l	A vertex in the cycle being infected last
$P(v\|G_n)$	Probability that $v = v^\star$, when G_n is observed
$P(\hat{v}\|G_n)$	Probability that $\hat{v} = v^\star$, i.e., correct detection probability

model rumor spreading. Nodes that possess the rumor are called *infected nodes* and otherwise they are *susceptible nodes*. The spreading is initiated by a single node $v^\star \in V$ that we call the rumor source. Once a node is infected (i.e., possesses the rumor), it stays infected and can in turn infect its susceptible neighbors. A rumor can spread from node i to node j if and only if there is an edge between them (i.e., $(i,j) \in E$). Let τ_{ij} be the spreading time from i to j, which are random variables that are independently and exponentially distributed with parameter λ (without loss of generality, let $\lambda = 1$). Hence, we have a random spreading model over an underlying graph G. Let G_n be a subgraph of order n of G, that models a snapshot observation of the spreading when there are n infected nodes, i.e., $|G_n| = n$. We shall assume that for each $v \in G_n$, v is not a leaf node of G, i.e., for each infected node $v \in G_n$, there are always susceptible neighbours of v in G. Clearly, G_1 is the *actual rumor source*, i.e., $v^\star$. The rumor source detection problem is thus to find $v^\star$ given this observation of G_n.

First, we review the maximum likelihood estimation problem of the rumor source in a tree network. The maximum likelihood estimator for the rumor source is the vertex v with the maximum probability $P(G_n|v)$ [6]. We focus on characterizing $P(G_n|v)$ for degree-regular tree networks.

Definition II.1. For a given G_n over the underlying graph G, $\hat{v}$ is an maximum likelihood estimator for the source in G_n, i.e., $P(\hat{v}|G_n) = \max_{v_i \in G_n} P(v_i|G_n)$.

By Bayes' theorem [15], $P(G_n|v)$ is the probability that v is the *actual rumor source culprit* that leads to observing G_n, and we have

$$P(v|G_n) = \frac{P(G_n|v)}{\sum_{v_i \in G_n} P(G_n|v_i)} \propto P(G_n|v).$$

Now, let σ_i be the possible spreading order sequence starting from v, and let $M(v, G_n)$ be the collection of all σ_i when v is the source in G_n. Then, we have

$$P(G_n|v) = \sum_{\sigma_i \in M(v,G_n)} P(\sigma_i|v). \tag{1}$$

In particular, for a d-regular tree, we have [6]:

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

Now, if G_n contains no cycle, then $P(\sigma_i|v) = P(\sigma_j|v)$ for all $\sigma_i, \sigma_j \in M(v, G_n)$. By combining (1) and (2), we have

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

which means that $P(G_n|v)$ is proportional to $|M(v, G_n)|$. This quantity $|M(v, G_n)|$ denoted by $R(v, G_n)$ is the *rumor centrality* in [6] that is crucial to solving the maximum likelihood estimation for degree-regular trees. In particular, the vertex having the maximum rumor centrality is called the *rumor center*.

Definition II.2. Let G_n be a tree with n vertices, for any $u, v \in G_n$, let T_v^u be the subtree rooted at v by removing the edge (u, v) from G_n. If G_n is a general graph, and T_q is a spanning tree of G_n, then $T_{q,v}^u$ is the subtree rooted at v by removing the edge (u, v) from T_q.

With Definition II.2, if G_n is a tree, then we have [6]:

$$|M(v,G_n)| = n! \cdot \prod_{u \in G_n} \frac{1}{|T_u^v|}, \tag{3}$$

moreover, if $u, v \in G_n$ are adjacent, then, from [6], we have

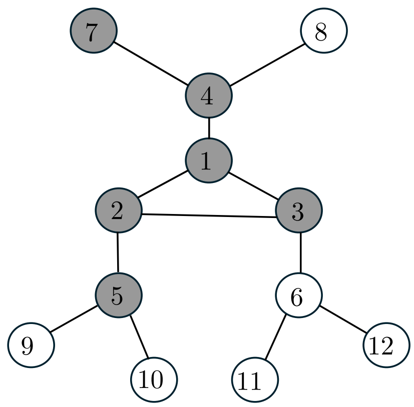
$$\frac{|M(v,G_n)|}{|M(u,G_n)|} = \frac{T_v^u}{T_u^v}. \tag{4}$$

We study a *generalized rumor centrality* for G_n with cycles (and thus an *generalized rumor center*) in this paper, but for brevity, we use *rumor center* to denote the vertex with the maximum rumor centrality.

III. PSEUDO-TREES WITH A CYCLE

Let us consider a case when G is an infinite regular graph with only one cycle, i.e., a pseudo-tree. We denote the cycle as C_h where h is the size (number of vertices on the cycle) of the cycle. Here, we call those vertices on C_h *cycle vertices*. Assume v is a cycle vertex, then we define t_v to be the subtree rooted at v in G_n. Take Figure 1 for example, t_{v_1} is the subtree that contains v_1, v_4 and v_7. In this section, we study how a cycle affects the probability $P(v|G_n)$ when G_n contains the cycle C_h. To generalize the analysis in [6], we should intuitively assume that the probability of being infected is proportional to the number of infected neighborhoods. With this assumption, the analysis in [6] will not change, but we can consider the case that two infected nodes have a common susceptible neighborhood, i.e., there is a cycle in G_n.

Fig. 1. G_6 is an infected subgraph with a single cycle C_3 contains three cycle vertices v_1, v_2 and v_3. We can partition G_6 into three subtrees say $t_1 = \{v_1, v_4, v_7\}$, $t_2 = \{v_2, v_5\}$ and $t_3 = \{v_3\}$.



A. *Impact of a Single Cycle On* $P(G_n|v)$

σ_i	Spreading Order	$P(\sigma_i\|G_6)$
σ_1	$v_4 \to v_1 \to v_3 \to v_2 \to v_5 \to v_7$	$\frac{2}{1200}$
σ_2	$v_4 \to v_1 \to v_2 \to v_5 \to v_3 \to v_7$	$\frac{2}{1800}$
σ_3	$v_4 \to v_1 \to v_2 \to v_5 \to v_7 \to v_3$	$\frac{2}{2520}$

Example 1. Consider the infected subgraph $G_6 \subset G$ as shown in Figure 1, where $G_6 = \{v_1, v_2, v_3, v_4, v_5, v_7\}$ and there is a 3-cycle in G_6. Consider a spreading order $\sigma_1 \in M(v_4, G_6)$, where $\sigma_1 : v_4 \to v_1 \to v_2 \to v_3 \to v_5 \to v_7$. We have $P(\sigma_1|v_4) = (1/3) \cdot (1/4) \cdot (2/5) \cdot (1/4) \cdot (1/5) = 2/1200$. Note that when v_1 and v_2 are infected, v_3 has two infected neighborhoods which implies that the probability of v_3 be infected in the next time period is twice higher than v_5, v_7 and v_8. In particular, there are three possible values for $P(\sigma_i|v_4)$ as shown in Table III-A, for all $\sigma_i \in M(v_4, G_6)$. Moreover, we observe that the denominators are different in Table III-A, due to sharing a common neighbor in the presence of a cycle. We call this property *cycle effect*.

Example 1 reveals some interesting properties of cycle effect due to a single cycle:

- $P(\sigma_i|v)$ increases with how soon the last cycle vertex appears in σ_i (as ordered from left to right of σ_i). The last cycle vertex on σ_1 is v_2, and is v_3 on σ_2 and σ_3.
- When there is a cycle in G_n, then $P(G_n|v)$ is no longer proportional to $|M(v, G_n)|$.
- For each σ_i, there are actually two corresponding permitted spreading order due to the cycle.

This means $P(\sigma_i|v)$ is no longer a constant for each i, and is dependent on the position of the last cycle vertex in each spreading order. We proceed to compute $P(\sigma_i|v)$ as follows. For brevity of notation, let v_l denote the last cycle vertex. Let $D_c(v)$ denote the shortest distance (number of hops) from v to C_h, for example, $D_c(v_4) = 1$ and $D_c(v_1) = 0$ in Figure 1. Remark: For each $\sigma \in |M(v, G_n)|$, v_l can be any vertex on the cycle except the vertex v' with the distance $d(v, v') = D_c(v)$.

We define

$$\begin{aligned} M_v^{v_l}(G_n, k) &= \{\sigma | v_l \text{ is on the } k\text{th position of } \sigma\}; \\ P_v^{v_l}(G_n, k) &= P(\sigma|v), \text{ for } \sigma \in M_v^{v_l}(G_n, k); \\ m_v^{v_l}(G_n, k) &= |M_v^{v_l}(G_n, k)|, \end{aligned}$$

where $M_v^{v_l}(G_n, k)$ is a set collects all the spreading orders starting from v and v_l is on the kth position, and $m_v^{v_l}(G_n, k)$ denote the number of elements in $M_v^{v_l}(G_n, k)$. In addition, $M_v^{v_l}(G_n, k)$ partition $M(v, G_n)$ into several parts according to the position of v_l. We have

$$|M(v, G_n)| = 2 \cdot \sum_{k=D_c(v)+h}^{n-t_{v_l}+1} m_v^{v_l}(G_n, k) \tag{5}$$

since the position of v_l on the spreading order will be ranging from $D_c(v)+h$ to $n-t_{v_l}+1$. For example, in Table III-A, we can see that $v_l = v_2$ is the 4th element on σ_1 and $v_l = v_3$ is the 6th element on σ_3. Here, 4 comes from $D_c(v_4) + h = 1 + 3$, and 6 comes from $n - t_{v_l} + 1 = 6 - 1 + 1$. Finally, the multiplication with 2 comes from the third property.

Now we can rewrite $P(G_n|v)$ for G_n with a cycle as:

$$P(G_n|v) = \sum_{k=D_c(v)+h}^{n-t_{v_l}+1} m_v^{v_l}(G_n, k) \cdot P_v^{v_l}(G_n, k), \tag{6}$$

and our goal is to find the vertex $\hat{v}$ that achieves

$$P(G_n|\hat{v}) = \max_{v_i \in G_n} P(G_n|v_i). \tag{7}$$

Since $P(G_n|v)$ is not proportional to $|M(v, G_n)|$, we should compute $P(G_n|v)$ by considering each part $m_v^{v_l}(G_n, k)$ and their corresponding probability $P_v^{v_l}(G_n, k)$. Let $z_d(i) = (i-1)(d-2)$, then

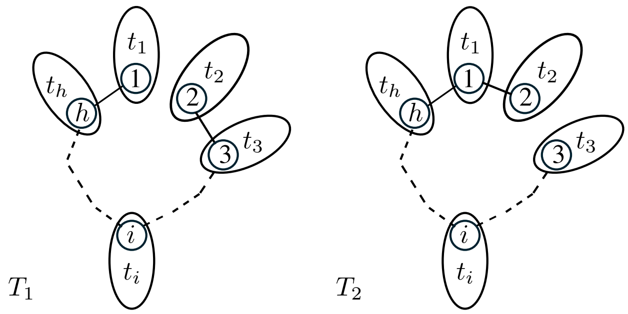
$$P_v^{v_l}(G_n, k) = 2 \cdot \prod_{i=1}^{k-1} \frac{1}{d + z_d(i)} \cdot \prod_{i=k-1}^{n-2} \frac{1}{d + z_d(i) - 1}. \tag{8}$$

The first factor in (8) is the probability that k vertices are infected where the kth infected vertex is v_l, and the second factor is the probability of that all remaining $n-k$ vertices are infected thereafter. The -1 in the denominator of the second factor and multiply by 2 in the beginning are due to the common neighbor in a cycle. Note that multiply by 2 in the beginning makes no difference when computing $P(G_n|v)$ for each $v \in G_n$. From (6), we can conclude that the number of spreading orders and the position of v_l will affect $P(G_n|v)$.

B. *Computing* $|M(v, G_n)|$ *for* v *on* G_n

In this section, we introduce how to compute $|M(v, G_n)|$. To compute $|M(v, G_n)|$, we can leverage the message-passing algorithm in [6] if G_n is a tree. Observe that for each infected vertex in G_n, it is infected by one of its infected neighbors (even it has two infected neighbors), so the actual infecting route is a spanning tree of G_n instead of a graph with cycle. Hence, the number of all spreading orders on a graph G_n with a cycle can be computed as

Fig. 2. C_h is constructed by $v_1, v_2, ..., v_h$, and t_i is a subtree rooted at v_i.



$$|M(v, G_n)| = \sum_{1 \le i \le h} |M(v, T_i)|, \tag{9}$$

where T_i is the spanning tree of G_n, for $i = 1, 2, ..., h$. If G_n contains a C_h, then the time complexity of computing $M(v, G_n)$ for $v \in G_n$ is $O(hn)$. Since G_n has h spanning trees, and for each spanning tree we can use the message-passing algorithm in [6] within $O(n)$ time.

C. Rumor Center on uni-cycle G_n

In this section, we are going to find the *rumor center* v_c in G_n. Instead of computing $|M(v, G_n)|$ for all $v \in G_n$ in each spanning tree, we will leverage some analytical results to help us to find v_c. Let $t_i = t_{v_i}$ be defined as above. In the following, with a slightly abuse, we denote the subtree size $|t_i|$ as t_i.

Theorem 1. *Let G_n be a regular graph with only one cycle C_h and $v_1, v_2, ..., v_h$ be the cycle vertices. If there exists a vertex $v \in G_n$ such that each connected component of $G_n \backslash \{v\}$ is of size less or equal to $n/2$, then $v_c = v$. Moreover, if v and v' are connected in G_n, and the connected component which contains v' is of size equal to $n/2$, then both v and v' are rumor center of G_n.*

Remark: v is the rumor center of G_n does not guarantee that each connected component of $G_n \backslash \{v\}$ is of size less or equal to $n/2$. (If G_n is a tree, then the statement above is true. [6]) The following lemma describe the case when the condition in Theorem 1 is not satisfied for all vertex in G_n.

Lemma 1. *From Theorem 1, if no such vertex exists, then the rumor center of G_n will be on the cycle, moreover, $t_i \le n/2$ for $i = 1, 2, ..., h$, i.e., there is a connected component of $G_n \backslash \{v_c\}$ with size larger than $n/2$.*

We can find the v_c of G_n by Theorem 1, but if the condition in Theorem 1 is not satisfied, then we can still know that v_c will on the cycle by Lemma 1. In the following, we will introduce how to find v_c on G_n if v_c is a cycle vertex. We denote T_j as the spanning tree of G_n which is constructed by $G_n \backslash (v_j, v_{j+1})$, for $j = 1, 2, ..., h-1$ and $T_h = G_n \backslash (v_h, v_1)$. Note that (v_h, v_1) and $(v_j, v_j + 1)$ for $j = 1, 2, ..., h-1$ are cycle edges of C_h. Given a cycle vertex v_i and assume that $|M(v_i, T_p)| = r$, where r and p are integers and $1 \le p \le h$. Then for $1 \le q \le h$, we have

$$\frac{|M(v_i, T_q)|}{|M(v_i, T_p)|} = \frac{\prod_{j \in C_h, j \ne i} T^i_{p,j}}{\prod_{k \in C_h, k \ne i} T^i_{q,k}}, \tag{10}$$

where $T^i_{p,j}$ is the subtree T^i_j of the spanning tree T_p. From [6], we know that the ratio of $|M(v_i, T_p)|/|M(v_j, T_p)|$ is proportional to their branch size in T_p if v_i and v_j are adjacent. Now, for the same vertex v_i, but in different spanning tree say T_p and T_x, we can also derive the ratio $|M(v_i, T_p)|/|M(v_i, T_q)|$. Hence, if we assume $|M(v_1, T_1)| = r$, then we can derive $|M(v, T)|$ for all $v \in C_h$ and T is a spanning tree of G_n in terms of r and t_i, where t_i is shown in Figure 2. We give an algorithm in the following to find v_c with time complexity $O(n + h^2)$ in worst case.

Algorithm 1 Algorithm to compute v_c for G_n with single cycle

---

Require: G_n

Step 1: If there is a vertex v_c satisfies the condition in Theorem 1, then v_c is the rumor center. If no such vertex exists, then go to Step 2.

Step 2: Given a spanning tree T^* and a cycle vertex v' on G_n, compute $|M(v', T^*)|$.

Step 3: Construct a table (e.g. Table II) of $|M(v, T)|$, where v is a cycle vertex and T is a spanning tree of G_n, starting from $|M(v', T^*)|$. For each row, i.e., on the same spanning tree T, if v_i and v_j are adjacent then we have $|M(v_i, T)| : |M(v_j, T)| = T^j_i : T^i_j$. For each column, i.e., on the same vertex v but different spanning tree, we can use the ratio in (10) to complete the column.

Step 4: Sum over each column, the vertex with $\max_{v \in G_n} \sum_{i=1}^{h} |M(v, T_i)|$ is v_c of G_n.

Take C_3 for example, $|M(v_1, T_1)|/|M(v_1, T_2)| = t_3/(t_2 + t_3)$. Assume $|M(v_1, T_1)| = r$, then table II shows the ratio of $|M(v_i, T_j)|$, for all $1 \le i, j \le 3$. With Table II, we can conclude that

$$|M(v_1, G_n)| = \tfrac{2(t_2+t_3)}{t_3} r;$$
$$|M(v_2, G_n)| = \tfrac{2t_2(t_2+t_3)}{t_1 t_3} r;$$
$$|M(v_3, G_n)| = \tfrac{2(t_2+t_3)}{t_1} r,$$

which implies

$$|M(v_1, G_n)| : |M(v_2, G_n)| : |M(v_3, G_n)| = t_1 : t_2 : t_3.$$

By Theorem 1 and the result above, we can conclude that : If G_n contains a C_3, then v_c of G_n is either a vertex satisfies the condition in Theorem 1 or a vertex on C_3 and satisfies $t_i = \max_{1 \le j \le 3} t_j$.

We can use the same approach to find the rumor center for any regular graph with a single cycle, Table III shows the ratio of $|M(v_i, T_j)|$ between any $1 \le i, j \le 4$ when G_n contains a C_4. From Table III, we have

TABLE II
TABLE OF $|M(v_i, T_i)|$, WHEN G_n CONTAINS A C_3

	$\|M(v_1,T_i)\|$	$\|M(v_2,T_i)\|$	$\|M(v_3,T_i)\|$
T_1	r	$\frac{t_2(t_2+t_3)}{t_1(t_1+t_3)}r$	$\frac{t_2+t_3}{t_1}r$
T_2	$\frac{t_2+t_3}{t_3}r$	$\frac{t_2(t_2+t_3)}{t_3(t_1+t_3)}r$	$\frac{t_2+t_3}{t_1+t_2}r$
T_3	$\frac{t_2}{t_3}r$	$\frac{t_2(t_2+t_3)}{t_1t_3}r$	$\frac{t_2(t_2+t_3)}{t_1(t_1+t_2)}r$

TABLE III
TABLE OF $|M(v_i, T_i)|$, WHEN G_n CONTAINS A C_4

	$\|M(v_1,T_i)\|$	$\|M(v_2,T_i)\|$
T_1	r	$\frac{(t_2+t_3+t_4)(t_2+t_3)t_2}{(t_1+t_3+t_4)(t_1+t_4)t_1}r$
T_2	$\frac{(t_2+t_3+t_4)(t_2+t_3)}{(t_3+t_4)t_3}r$	$\frac{(t_2+t_3+t_4)(t_2+t_3)t_2}{(t_1+t_3+t_4)(t_3+t_4)t_3}r$
T_3	$\frac{(t_2+t_3+t_4)t_2}{t_3t_4}r$	$\frac{(t_2+t_3+t_4)(t_2+t_3)t_2}{(t_1+t_4)t_3t_4}r$
T_4	$\frac{(t_2+t_3)t_2}{(t_3+t_4)t_4}r$	$\frac{(t_2+t_3+t_4)(t_2+t_3)t_2}{(t_3+t_4)t_1t_4}r$
	$\|M(v_3,T_i)\|$	$\|M(v_4,T_i)\|$
T_1	$\frac{(t_2+t_3+t_4)(t_2+t_3)}{(t_1+t_4)t_1}$	$\frac{t_2+t_3+t_4}{t_1}r$
T_2	$\frac{(t_2+t_3+t_4)(t_2+t_3)}{(t_1+t_2+t_4)(t_1+t_2)}r$	$\frac{(t_2+t_3+t_4)(t_2+t_3)}{(t_1+t_2)t_3}r$
T_3	$\frac{(t_2+t_3+t_4)(t_2+t_3)t_2}{(t_1+t_2+t_4)(t_1+t_4)t_4}r$	$\frac{(t_2+t_3+t_4)t_2}{(t_1+t_2+t_3)t_3}r$
T_4	$\frac{(t_2+t_3+t_4)(t_2+t_3)t_2}{(t_1+t_2)t_1t_4}r$	$\frac{(t_2+t_3+t_4)(t_2+t_3)t_2}{(t_1+t_2+t_3)(t_1+t_2)t_1}r$

$$\frac{|M(v_i,G_n)|}{|M(v_j,G_n)|} = \frac{t_i(t_{i-1}+t_{i+1})(n-t_j)}{t_j(t_{j-1}+t_{j+1})(n-t_i)},$$

if v_i and v_j are adjacent.

IV. CONCLUSION

We proposed a generalized rumor centrality to analyze the rumor source detection problem for a degree-regular graph with a single cycle. We demonstrated the cycle effect on the likelihood $P(v|G_n)$ according to the particular location of a vertex v_l in the cycle being infected last. The likelihood is characterized by both rumor centrality and the location of v_l, hence, in this paper we focus on how to find the rumor center, i.e, the vertex with the maximum rumor centrality. If G_n satisfies the condition in Theorem 1, then we can find the rumor center of G_n by consider any spanning tree of G_n. If the condition in Theorem 1 is not satisfied, then we can use Algorithm 1 to find the rumor center with time complexity $O(n + h^2)$ where h is the size of the cycle.

REFERENCES

[1] A. Ganesh, L. Massoulie, and D. Towsley, "The effect of network topology on the spread of epidemics," *Proc. IEEE INFOCOM*, 2005.

[2] G. Fanti, P. Kairouz, S. Oh, K. Ramchandran, and P. Viswanath, "Hiding the rumor source," *IEEE Trans. Information Theory*, forthcoming, 2017.

[3] D. Easley and J. Kleinberg, *Networks, Crowds, and Markets: Reasoning About a Highly Connected World*. Cambridge University Press, 2010.

[4] L. Vassio, F. Fagnani, P. Frasca, and A. Ozdaglar, "Message passing optimization of harmonic influence centrality," *IEEE Trans. Control of Network Systems*, vol. 1, no. 1, pp. 109–120, 2014.

[5] N. Karamchandani and M. Franceschetti, "Rumor source detection under probabilistic sampling," *Proc. IEEE ISIT*, 2013.

[6] D. Shah and T. Zaman, "Rumors in a network: Whos's the culprit?," *IEEE Trans. Information Theory*, vol. 57, pp. 5163–5181, 2011.

[7] D. Shah and T. Zaman, "Rumor centrality: a universal source detector," *Proc. of ACM SIGMETRICS*, 2012.

[8] W. Dong, W. Zhang, and C. W. Tan, "Rooting out the rumor culprit from suspects," *Proc. of IEEE ISIT*, 2013.

[9] W. Luo, W. P. Tay, and M. Leng, "Identifying infection sources and regions in large networks," *IEEE Trans. Signal Processing*, vol. 61, no. 11, pp. 2850–2865, 2013.

[10] 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.

[11] J. Khim and P. Loh, "Confidence sets for the source of a diffusion in regular trees," *IEEE Trans. Network Science and Engineering*, vol. 4, no. 1, pp. 27–40, 2017.

[12] C. W. Tan, P. D. Yu, C. K. Lai, W. Zhang, and H. L. Fu, "Optimal detection of influential spreaders in online social networks," *Proc. of CISS*, pp. 145–150, 2016.

[13] B. Zelinka, "Medians and peripherians of trees," *Arch. Math.*, vol. 4, no. 2, pp. 87–95, 1968.

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

[15] D. J. C. Mackay, *Information Theory, Inference and Learning Algorithms*. Cambridge University Press, first ed., 2003.

APPENDIX

A. Proof of Theorem 1

Proof. Let $v \in G_n$ and suppose each connected component of $G_n \backslash \{v\}$ is of size strictly less then $n/2$. Given $v' \in G_n$, then from (9) we have

$$\begin{aligned}|M(v', G_n)| &= |M(v', T_1)| + |M(v', T_2)| + ... + |M(v', T_h)| \\ &< |M(v, T_1)| + |M(v, T_2)| + ... + |M(v, T_h)| \\ &= |M(v, G_n)|.\end{aligned}$$

Since for $i = 1, ..., h$, we have $T^v_{i,v'} < n/2$ which implies

$$|M(v, T_i)| > |M(v', T_i)|,$$

[6] and we conclude that v is the rumor center. Here, $T^v_{i,v'}$ is defined in Definition II.2 where $T = T_i$. Now, suppose there is a vertex v' such that the connected component of $G_n \backslash \{v\}$ which contains v' is of size $n/2$. Then we have $|M(v, T_i)| = |M(v', T_i)|$, for $i = 1, ..., h$, which implies $|M(v', G_n)| = |M(v, G_n)|$. Hence, we can conclude that both v and v' are rumor centers of G_n. Note that this proof can be extended to multiple cycles. □

B. Proof of Lemma 1

Proof. Assume that for any $v \in G_n$, there is a connected component of $G_n \backslash \{v\}$ with size larger than $n/2$. To the contrary, suppose v' is not a cycle vertex and v' is the rumor center of G_n. By our assumption, we can find a vertex v'' in the largest connected component, i.e. its size $> n/2$, of $G_n \backslash \{v'\}$ such that $|M(v'', G_n)| > |M(v', G_n)|$. (This can be proved by using the same approach in the proof of Theorem 1.) This implies v' is not the rumor center, which is a contradiction. Hence, the rumor center must be a cycle vertex.

Now, suppose there is a integer $1 \le i' \le h$ such that $t_i > n/2$. Then we can find a vertex $v' \in t_{i'}$ such that all connected component of v' is of size less or equal to $n/2$. Again, this is contradict to our assumption. So, we can conclude that $t_i < n/2$ for all $i = 1, 2, ..., h$. □