Network observability for source localization in graphs with unobserved edges

Sabina Zejnilović*†, Dieter Mitsche‡, João Gomes† and Bruno Sinopoli*
*Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA
†Institute for Systems and Robotics, Instituto Superior Técnico, University of Lisbon, Portugal
‡Laboratoire J.A. Dieudonné, Université de Nice Sophia-Antipolis, Nice, France

***Abstract*—Localizing a source of diffusion is a crucial task in various applications such as epidemics quarantine and identification of trendsetters in social networks. We analyze the problem of selecting the minimum number of observed nodes that would lead to unambiguous source localization, i.e. achieve network observability, when both infection times of all the nodes, as well as the network structure cannot be fully observed. Under a simple propagation scenario, we model the assumption that, while the structure of local communities is well known, the connections between different communities are often unobserved. We present a necessary and sufficient condition for the minimum number of observed nodes in networks where all components have either a tree, a grid, a cycle or a complete graph structure. Additionally, we provide a sufficient condition for the selection of observed nodes when the components are of arbitrary structure. Through simulation, we illustrate the performance of the proposed bound.**

***Index Terms*—network theory, graphs, source localization, observability**

I. Introduction

Network diffusion is used to describe various phenomena, such as spreading of epidemics in human populations and propagation of information in social networks. In each of these examples, localizing the source of diffusion is an important task that needs to be performed in order to either curb infections, restrict further damage, or identify trendsetters. Estimating the source of rumors or epidemics was first addressed in [1], followed by a growing number of works [2]–[6].

Having access to all the nodes in the network, due to network size, limited resources and privacy issues is not always feasible. In that case, source localization is performed based on the observations of a subset of nodes, denoted as the observers [2], [3], [6]. Consequently, choosing the most informative subset of nodes becomes an important issue and several strategies have been explored in the literature. The performance of high-degree nodes is compared to randomly selected nodes through simulation in [2]. Selection strategies, based on different centrality measures, are experimentally evaluated in [6]. In [7], the problem of finding the smallest subset of observed nodes to achieve correct source localization, under a simple deterministic propagation model, is cast as the problem of finding the smallest resolving set of a graph. Selecting a sufficient number of observers in incompletely observed tree networks is presented in [8]. In this paper, we further analyze the problem of selecting the minimum number of observers in networks where some edges are not known.

In many cases, complete knowledge of network topology is not a realistic assumption. Individuals may not be willing to disclose all their social connections, and not all information is propagated through monitored social network sites. Based on infection times of nodes, the complete network structure can be inferred [9], or missing network data can be learned [10]. Often, local connections within communities are well known, while the connections between them are not always observed. This may happen when diseases spread from one community to another through random contact, rather than a known friendship connection, or when novel information is spread through weak, rather than strong, social ties [11]. Therefore, in our model with deterministic propagation we assume that complete knowledge of local network components is available, while inter-component edges are the ones that are not observed. We classify the network as being observable for the source localization problem if the choice of observed nodes is such that the source can be unambiguously identified [7]. We present a necessary and sufficient condition for the minimum number and placement of observed nodes that would make the network observable, if each of the network components is either a tree, a grid, a cycle or a complete graph. Furthermore, when the components are of arbitrary structure, we present a sufficient condition for network observability and through simulation we illustrate the performance of the bound.

II. Model setup and problem statement

We assume a widely studied Susceptible-Infected propagation model, where once a node is infected, it remains infected [1], [2], [4]. A single infected node at time 0 initiates the network diffusion. We adopt a simple model of diffusion where once a node is infected at $t-1$, in the next time instant t, where t is a discrete time index, it will infect all of its neighbors, with probability 1. The time of infection of a node corresponds to its distance to the source. Assuming that resources are limited, only a subset of nodes is monitored. The source is then identified based on the infection times of

This research was supported by Fundação para a Ciência e a Tecnologia (project PEst-OE/EEI/LA0009/2013 and a PhD grant from the Carnegie Mellon-Portugal program) and EU FP7 project MORPH (grant agreement no. 288704)

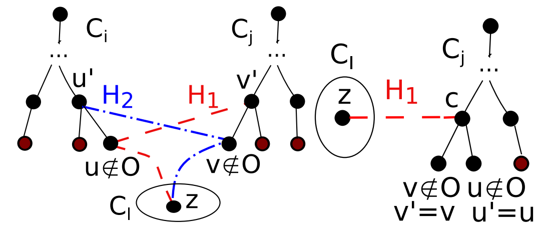
978-1-4799-7088-9/14/$31.00 ©2014 IEEE

observed nodes. A network of n nodes is represented using a graph $G = \{V, E\}$, where $V = \{1, \ldots, n\}$ is the set of nodes representing the vertices and $E \subseteq \binom{V}{2}$ is the set of edges. In undirected graphs, there is an edge between nodes i and j if they can communicate directly. Since we assume the knowledge of the structure of local communities, but not of the connections between communities, the observed network is a disconnected graph F that comprises k components, C_i, for $i = 1, \ldots k$. Since we are interested only in the first time that a node gets informed or infected, we assume the number of unobserved inter-component edges is $k-1$, making the graph connected, yet there are no cycles between the components. We denote as $\mathcal{F}(k)$ the class of such observed graphs with k components and $k-1$ missing edges. Then, with $\mathcal{H}(F)$, we denote the class of all the possible graphs that can be constructed by adding $k-1$ edges between the components of the observed graph such that the resulting graph is connected. If the observed graph is $F \in \mathcal{F}(k)$, then the true network structure can be any graph from the class $\mathcal{H}(F)$.

A path $i - j$ is a sequence of all different nodes starting from node i and ending with j. A tree is a connected graph without any cycles and a forest is a disjoint union of trees. The distance between two nodes i and j in a connected graph H is the number of edges in the shortest path between them, denoted as $d_H(i, j)$. If $S \subseteq V$ denotes the set of nodes $\{s_1, \ldots, s_r\}$, then $\boldsymbol{d}_H(i, S)$ is the r-vector of distances $[d_H(i, s_1), \ldots, d_H(i, s_r)]$. A resolving set of nodes of a graph H is a set S, such that $\boldsymbol{d}_H(u, S) \neq \boldsymbol{d}_H(v, S)$ for any two different nodes $u, v \in V$ [12], i.e. any two nodes have distinct distance signatures. Finding a resolving set of an arbitrary graph is an NP-hard problem and can be approximated by a greedy algorithm within a factor of $\mathcal{O}(\log n)$ [12]. Under deterministic diffusion, the network is observable if the observers form a resolving set [7]. A forest is observable if all nodes have different distances to the set of observers across all possible trees that correspond to the forest [8]. The number of such trees explodes exponentially with the number of unobserved edges. We now extend this concept to general networks. By a generalized resolving set of a disconnected graph F, we denote a set of nodes O such that for any two different nodes u and v, and any two graphs $H_1, H_2 \in \mathcal{H}(F)$, $\boldsymbol{d}_{H_1}(u, O) \neq \boldsymbol{d}_{H_2}(v, O)$. Hence, a network $F \in \mathcal{F}(k)$ is observable if the observers form a generalized resolving set. As resources available for observing nodes are often limited, we are interested in finding the smallest set of observers that make the network observable.

III. Minimum number of observers

Let S_i denote a minimum cardinality resolving set of a component C_i, $i = 1, \ldots, k$. A leaf is a node of degree 1. Let $L(C_i)$ denote the set of all leaves of C_i, and $K(C_i)$ the set of nodes of at least degree 3 that are connected by paths to one or more leaves, when C_i is a tree. While in [8], it was shown that the forest is observable if all of its leaves are observed, here we present both the necessary and sufficient condition.



(a) Both components have a leaf which is not in O.

(b) All leaves from C_i are in O, C_j is not a path.

Fig. 1: Trees H_1, H_2 where u and v are not distinct

Theorem 1. *Let $F \in \mathcal{F}(k)$ be a graph where each component is a tree. For F to be observable the necessary and sufficient number of observers is $\min_j \sum_{i=1, i\neq j}^{k} |L(C_i)| + |S_j|$, unless all components are isolated nodes, in which case $k-1$ nodes are needed. We may assume without loss of generality, that the minimum is attained for $j = k$. Then the set $O = \cup_{i=1}^{k-1} L(C_i) \cup S_k$ is a minimum cardinality generalized resolving set of F.*

Proof. We first prove the sufficiency claim. Since all the leaves of a tree are its resolving set [12], any two nodes u and v from the same component are distinct. Let us assume now that u is in C_i and v in C_j, for $i \neq j$. If $u \in O$, then it is distinguishable from v, since $d_{H_1}(u, u) = 0 \neq d_{H_2}(u, v)$. Let p be a node in C_i and q in C_j, such that $p - q$ is a path that connects C_i and C_j in H_2. Observe that $d_{H_2}(p, q) \geq 1$. If $u \notin O$ and $u = p$, then for all $r \in O$ in C_i we have $d_{H_2}(r, v) = d_{H_2}(r, p) + d_{H_2}(p, q) + d_{H_2}(q, v) > d_{H_2}(r, p) = d_{H_1}(r, u)$. For the remaining case, $u \notin O$ and $u \neq p$, let r be a leaf in $L(C_i)$ such that u is on the path $r - p$. Such a leaf clearly exists [8]. Then $d_{H_2}(r, v) = d_{H_2}(r, u) + d_{H_2}(u, p) + d_{H_2}(p, v) > d_{H_1}(r, u)$. Thus, the two distance vectors are never equal.

Now let O be an arbitrary generalized resolving set. We will show that O has to be at least the size given by the sufficient condition. Let C_i and C_j be 2 components with at least 2 nodes, such that both have a leaf which is not in O. Let u be such a leaf in C_i with neighbor u' and v be a leaf in C_j with neighbor v', such that $u, v \notin O$. We can construct H_1 and H_2 as shown in Figure 1 (a) where u and v are not distinct. Hence, either all the leaves in C_i or C_j have to be in O. Wlog, let us assume $L(C_i) \in O$. Now we assume that from C_j only $|S_j| - 1$ nodes are selected. When C_j is not a path, there exists a node $c \in K(C_j)$ such that two of its associated leaves, u and v, are both not in O. Figure 1 (b) illustrates how tree H_1 is constructed where u and v are not distinct. When C_j is path and $|S_j| - 1 = 0$ nodes from C_j are in O, H_1 and H_2 can be constructed where two terminal nodes of C_j are not distinct. Thus, at least $|S_j|$ nodes have to be taken from C_j. Similarly, it can be shown that when C_i has only one node, it has to be in O, unless C_j also has only one node. In that case, it is necessary to include one of the nodes. Therefore, if there exists at least one component with 2 or more nodes, from all but one component all the leaves have to be taken, and from the remaining, at least a resolving set. If all k components

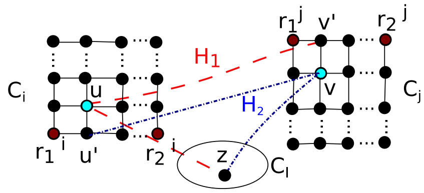


Fig. 2: Constructing H_1 and H_2 when components are grids

have only one node, $|O|$ should be $k-1$. □

A grid is a graph whose nodes correspond to the points in the plane with integer coordinates. Two nodes are connected by an edge whenever the corresponding points are at Euclidean distance 1. Corner nodes are nodes of degree two and consecutive corner nodes have the same value in one coordinate.

Theorem 2. *Let $F \in \mathcal{F}(k)$ be a graph where each component is a grid. For F to be observable the necessary and sufficient number of observers is $3k-1$. Let $O_i = \{r_1^i, r_2^i, r_3^i\}$ denote a set of 3 corner nodes in C_i. Then $O = \cup_{i=1}^{k-1} O_i \cup S_k$ is a minimum cardinality generalized resolving set of F.*

Proof. Let the corner node r_1^i be in position $(0,0)$, r_2^i at $(x_i, 0)$ and r_3^i at $(0, y_i)$. Since any two consecutive corner nodes form a resolving set of a grid [12], we follow the same reasoning as in Theorem 1 and analyze only the case $u \neq p$ to show sufficiency. We prove the claim by contradiction and assume $\boldsymbol{d}_{H_1}(u, O_i) = \boldsymbol{d}_{H_2}(v, O_i)$, for u in C_i and v in C_j. We then obtain the following equations:

$$\begin{aligned} x_u + y_u &= x_p + y_p + d_{H_2}(p,q) + d_{H_2}(q,v) \\ x_i - x_u + y_u &= x_i - x_p + y_p + d_{H_2}(p,q) + d_{H_2}(q,v) \\ x_u + y_i - y_u &= x_p + y_i - y_p + d_{H_2}(p,q) + d_{H_2}(q,v). \end{aligned} \quad (1)$$

The system of equations (1) has a single solution $x_u = x_p$, $y_u = y_p$, and $d_{H_2}(p,q) + d_{H_2}(q,v) = 0$ which contradicts $d_{H_2}(p,q) \geq 1$ and proves the claim.

To prove necessity, let us assume there exist two components C_i and C_j, such that from each, only two nodes, r_1^i, r_2^i from C_i and r_1^j, r_2^j from C_j are in O. If on at least one component the selected corner nodes are not consecutive, we claim, and omit the proof due to space limitations that there exist nodes u and v that are not distinguishable by r_1^i and r_2^i. H_1 can be constructed by connecting r_1^i with any node from any other component, and u and v are still not distinct. Otherwise, when the corner nodes are consecutive, Figure 2 illustrates the choice of nodes u and v, H_1 and H_2, such that $\boldsymbol{d}_{H_1}(u, O) = \boldsymbol{d}_{H_2}(v, O)$. Therefore, at least 3 nodes from C_i or C_j have to be in O. Assuming that 3 nodes from C_i are in O, with the same arguments as in Theorem 1 it can be shown that at least $|S_j| = 2$ nodes from C_j have to be in O. In conclusion, for any two components, at least 3 nodes from one and 2 nodes from another have to be in O. □

Theorem 3. *Let $F \in \mathcal{F}(k)$ be a graph where each component is a cycle of size greater than 3. Let k_e denote the number of components that have an even number of nodes. For F to be observable the necessary and sufficient number of observers is $2k + k_e - 1$, if $k_e > 0$, and $2k$ otherwise. For a component C_i with an even number of nodes n_i, let $O_i = \{r_1^i, r_2^i, r_3^i\}$, where r_1^i, r_2^i are neighboring nodes in C_i and r_3^i is at a distance at least $\frac{n_i - 2}{2}$ from both of them. For an odd cycle C_i, let $O_i = \{r_1^i, r_2^i\}$, where r_1^i and r_2^i are nodes in C_i at distance $\frac{n_i-1}{2}$ from each other. If $k_e = 0$, a minimum cardinality generalized resolving set of F is $\cup_{i=1}^{k} O_k$, otherwise it is $O = \cup_{i=1}^{k-1} O_i \cup S_k$, assuming wlog that C_k is an even cycle.*

Proof. First, let us prove the claim of sufficiency. If the nodes u and v are not distinguishable by O_i, then $d_{H_1}(r,u) = d_{H_2}(r,v) = d_{H_1}(r,p) + d_{H_2}(p,v)$, for some H_1 and H_2 and

$$d_{H_1}(u,r) > d_{H_1}(p,r) \quad (2)$$

must hold for some $r \in O_i$. First, we analyze when both C_i and C_j are even cycles. When both p and u lie in the same semicycle, i.e. both on the shorter path $r_2^i - r_3^i$ or both on the shorter path $r_1^i - r_3^i$, either u or p is closer to r_3^i and the other one is closer to r_1^i. Then (2) cannot hold for both r_1^i and r_3^i. When u and p lie in different semi-cycles, then (2) cannot hold for both r_1^i and r_2^i. Similarly, we analyze when at least one component is an odd cycle with the main difference that the shorter path $r_1^i - r_2^i$ has length $\frac{n_i-1}{2}$, while the longer path $r_1^i - r_2^i$ is $\frac{n_i+1}{2}$. Using this we also obtain that condition (2) cannot hold for both r_1^i and r_2^i. Therefore, it is sufficient to select 2 from odd, and 3 from even cycles, except for the last component, from which it suffices to select 2 nodes.

Now, we prove the claim of necessity, observing that at least 2 nodes of each cycle have to be chosen, as otherwise the two neighbors of the chosen node r are not distinct within graph H_1 constructed by connecting r with one node of each other component. Let us assume there exist two even cycles from which only two nodes, r_1^i, r_2^i in C_i and r_1^j, r_2^j in C_j are in O. If on at least one component C_i, nodes r_1^i and r_2^i are exactly $\frac{n_i}{2}$ apart, then let u and v be two neighbors of r_1^i. Then u and v are not distinct and we can always construct H_1, as described previously, so that they remain as such. Otherwise, let us assume in both C_i and C_j, the nodes selected in O are not at distance $\frac{n_i}{2}$ ($\frac{n_j}{2}$, respectively). Then, let u (v) be a neighbor of r_1^i in C_i (r_1^j in C_j) that is on the longer path $r_1^i - r_2^i$ ($r_1^j - r_2^j$). We can construct H_1 by connecting u with r_1^j and u with some node z at any other component (if there are more components), and H_2 by connecting v with r_1^i and v with the same node z. Then again, u and v are not distinct. Hence, at least 3 nodes from even cycle C_i or C_j have to be in O. In conclusion, from all but one component with an even number of nodes, 3 nodes have to be chosen, and from the remaining ones, at least 2. □

Theorem 4. *Let $F \in \mathcal{F}(k)$ be a graph where each component is a complete graph of at least 3 nodes. For F to be observable the necessary and sufficient number of observers is $n-k$. A set consisting of all but one node on each component is a minimum cardinality generalized resolving set of F.*

Proof. We omit the straightforward proof. □

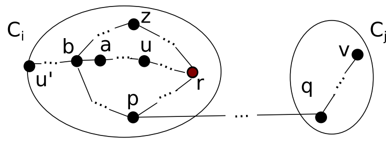


Fig. 3: Extended shortest path $r - u'$

In order to present a result on observability of a graph $F \in \mathcal{F}(k)$ with k arbitrary components, we review the concept of boundary of a graph. A node v is a boundary node of u if $d_G(w, u) \leq d_G(v, u)$, for all w that are neighbors of v [13]. A node v is a boundary node of G if it is a boundary node of some node of G. The boundary of graph G, $\partial(G)$, is the set of all the boundary nodes of G and can be easily determined by comparing distances among the neighbor nodes.

Theorem 5. *For any graph $F \in \mathcal{F}(k)$ to be observable, it is sufficient to observe the set of nodes $O = \cup_{i=1}^{k-1} \partial(C_i) \cup S_k$.*

Proof. Since the boundary is a resolving set [14], again we only analyze for $u \neq p$. The shortest path $p - u$ in C_i is extended to the shortest path $p - u'$ such that u' is a boundary node of p, and u can be u'. For a fixed shortest path $p - u'$ we have $d_{H_2}(u', v) = d_{H_2}(u', p) + d_{H_2}(p, v) = d_{H_1}(u', u) + d_{H_1}(u, p) + d_{H_2}(p, v) > d_{H_1}(u', u)$, completing the proof. □

The boundary of a tree is its leaves, the boundary of a grid are 4 corner nodes, while all the nodes are the boundary of a cycle [14]. As the boundary can be very large, we tighten the bound on the cardinality of a generalized resolving set.

Theorem 6. *Let $\partial(S_i)$ denote the boundary of the resolving set of C_i and let $O_i = S_i \cup \partial(S_i)$. For any graph $F \in \mathcal{F}(k)$ to be observable, it is sufficient to observe the set of nodes $O = \cup_{i=1}^{k-1} O_i \cup S_k$.*

Proof. For $r \in S_i$, following the notation of Theorem 5, the shortest path $r - u$ in C_i is extended to the shortest path $r - u'$, such that $u' \in \partial(r)$. Again, we analyze when $u \neq u'$.

Case I: There exists a shortest path from u' to p, and consequently to v, that passes through u. Then we have $d_{H_2}(u', v) = d_{H_1}(u', u) + d_{H_1}(u, p) + d_{H_2}(p, v) > d_{H_1}(u', u)$ and u and v have different distances to u'.

Case II: All shortest paths $u' - p$ do not pass through u. Let b be a node where path $p - u'$ connects to $r - u'$, as illustrated in Figure 3. Node b can be u', but not u, as that is Case I. We show that at least one shortest path $r - b$ passes through u, by assuming the opposite, i.e. there exists a node z with

$$d_{H_1}(r, u) + d_{H_1}(u, b) > d_{H_1}(r, z) + d_{H_1}(z, b). \quad (3)$$

Node z can also be p. Now let a be a node that immediately precedes b on the path $u - u'$. Such a node exists, as $b \neq u$ under our assumptions. Then $d_{H_1}(r, a) = \min \{d_{H_1}(r, z) + d_{H_1}(z, b) + 1, d_{H_1}(r, u) + d_{H_1}(u, b) - 1\}$. If the first value is smaller, we have $d_{H_1}(r, a) > d_{H_1}(r, b)$, which is not possible as the distance from node r does not decrease along the extended path $r - u'$. If the second value is smaller or values are equal, and yet we have that (3) holds, we get $d_{H_1}(r, a) = d_{H_1}(r, b)$. This implies that the shortest path $r - u$ could be extended only to $r - a$, and not to $r - b$, which is a contradiction and proves the claim.

Since the shortest path from $r - b$ goes through u, we have

$$d_{H_1}(r, u) + d_{H_1}(u, b) \leq d_{H_1}(r, p) + d_{H_1}(p, b). \quad (4)$$

As r did not distinguish u and v, Condition 2 holds. Using this in (4), it follows $d_{H_1}(u, b) < d_{H_1}(p, b)$. Now, $d_{H_2}(v, u') = d_{H_2}(v, p) + d_{H_2}(p, b) + d_{H_2}(b, u') > d_{H_2}(v, p) + d_{H_2}(u, b) + d_{H_2}(b, u') > d_{H_1}(u, u')$, which completes the proof. □

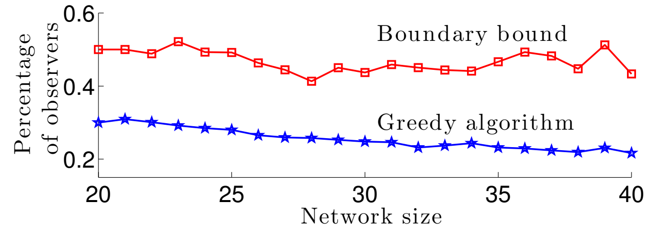


Fig. 4: Performance of the boundary bound

IV. Simulation results

We compared the sufficient number of observers chosen by Theorem 6 and by a greedy algorithm adapted from [12], as the calculation of the true minimum is computationally intensive even for modest network sizes. The adapted greedy algorithm selects one by one a node that distinguishes the most node pairs across all possible topologies that can be constructed by adding an edge between two components of the graph. We have generated 200 graphs, where each graph comprises two Erdős-Rényi components, random size n_i $10 - 20$ nodes, $n_i p = 4$, where p is the edge probability. For each graph, we calculated the boundary nodes of a (not necessarily the smallest) resolving set of each component, using the original greedy algorithm and we have averaged the results for same size networks. Figure 4 shows that the average percentage of observers selected by the bound is higher than by the greedy algorithm. However, unlike the latter, Theorem 6 does not require comparison of distance vectors through an exponential number of topologies, but can be applied in polynomial time.

V. Conclusions

We analyzed the problem of finding the minimum number of nodes that need to be observed when the connections between network components are not known, in order to localize the source correctly. We formulated this as a problem of finding a modified version of the smallest resolving set which is known to be an NP-hard problem for arbitrary graphs. We calculated the minimum number of nodes that is necessary and sufficient when the network components are either all trees, grids, cycles or complete graphs. When the network components are of arbitrary structure, we presented an upper bound on the minimum number of nodes in terms of the boundary nodes of components. Through simulation, we have illustrated the performance of the bound, which, although generally not very tight, can be determined in polynomial time.

REFERENCES

[1] D. Shah and T. Zaman, “Rumors in a network: who’s the culprit?” *IEEE Trans. on Information Theory*, 2011.

[2] P. Pinto, P. Thiran, and M. Vetterli, “Locating the source of diffusion in large-scale networks,” *Physical Review Letters*, August 2012.

[3] W. Luo and W. P. Tay, “Estimating infection sources in a network with incomplete observations,” *IEEE GlobalSIP*, pp. 301 – 304, 2013.

[4] 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. 2850–2864, June 2013.

[5] C. H. Comin and L. da Fontoura Costa, “Identifying the starting point of a spreading process in complex networks,” *Physical Review*, vol. 84, 2011.

[6] E. Seo, P. Mohapatra, and T. F. Abdelzaher, “Identifying rumors and their sources in social networks,” *SPIE Defense, Security, and Sensing*, April 2012.

[7] S. Zejnilovic, J. Gomes, and B. Sinopoli, “Network observability and localization of the source of diffusion based on a subset of nodes,” *Allerton*, pp. 847– 852, 2013.

[8] ——, “Network observability and localization of sources of diffusion in tree networks with missing edges,” *EUSIPCO*, 2014.

[9] M. Gomez-Rodriguez, J. Leskovec, and A. Krause, “Inferring networks of diffusion and influence,” *ACM Transactions on Knowledge Discovery from Data*, vol. 5, no. 4, pp. 21–37, February 2012.

[10] M. Kim and J. Leskovec, “The network completion problem: Inferring missing nodes and edges in networks,” *SDM*, pp. 47–58, 2011.

[11] M.O.Jackson, *Social and Economic Networks*. Princeton University Press, 2008.

[12] S. Khuller, B. Raghavachari, and A. Rosenfeld, “Landmarks in graphs,” *Discrete Applied Mathematics*, vol. 70, no. 3, pp. 217–229, 1996.

[13] G. Chartrand, D. Erwin, G. L. Johns, and P. Zhang, “Boundary vertices in graphs,” *Discrete Mathematics*, vol. 263, pp. 25–34, 2003.

[14] C. Hernando, M. Mora, I. M. Pelayo, and C. Seara, “Some structural, metric and convex properties on the boundary of a graph,” *Electronic notes in Discrete Mathematics*, vol. 24, pp. 203–209, 2006.