Selecting observers for source localization via error exponents

Sabina Zejnilović*†, João Xavier†, 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

***Abstract*—In today's large social and technological networks, since it is unfeasible to observe all the nodes, the source of diffusion is determined based on the observations of a subset of nodes. The probability of source localization error depends on the particular choice of observer nodes. We propose a criterion for observer node selection based on the minimal pairwise Chernoff distance between distributions of different source candidates. The proposed approach is optimal for the fastest error decay with vanishing noise. Although suboptimal for non-negligible noise, through simulation, we demonstrate its applicability in achieving low error probability. We also analyze the effect of network topology on the resulting error by bounding the smallest Chernoff distance for some specific networks.**

***Keywords*—source localization, error exponent, multiple hypothesis testing, subset selection, graphs**

I. Introduction

In order to quickly curb infectious diseases, prevent spreading of rumors in social networks or identify individuals who spread computer viruses, it is important to localize the source of such diffusion. Due to the size of typical social or technological networks, limited resources, and privacy issues, it is not practically possible to monitor all the network nodes. Hence, source localization is often performed knowing only the times when a subset of nodes became informed or infected. Nodes whose observations are available are denoted as observer nodes, and the performance of the source estimator depends on the choice of the observed subset [1], [2]. Consequently, choosing the most informative subset of observer 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 [1]. Selection strategies based on different centrality measures are experimentally evaluated in [2]. In [3], a graph theoretic approach is used to determine the smallest subset of observer nodes that achieves correct source localization, under a simple deterministic propagation model.

In our diffusion model, we assume that a constant time is needed for infection to spread from a node to its neighbor, but each node shows infection symptoms only after some variable time. This models the existence of incubation period, which is the time that passes between the moment when an individual contracts a virus until the symptoms are exhibited. Similarly, a person might hear about a certain product and might continue spreading the news about it to friends, but will purchase that product only after some time. As the exponential noise is the worst possible non-negative additive noise [4], we model the duration of the incubation period as a random variable with exponential distribution, independent and identically distributed across observer nodes.

With the goal of selecting the subset of observers that would result in low error probability, we formulate source localization as a multiple hypothesis problem. Based on an analysis of the error exponent for vanishingly small noise, we propose a new criterion for observer node selection. Our criterion is based on the smallest pairwise Chernoff distance between distributions that characterize different source candidates and it is optimal for the asymptotically noiseless case. However, even in the presence of non-negligible noise, we illustrate through simulation that selecting the observer nodes based on the proposed criterion results in low error probability.

We also show that the proposed criterion is a true distance metric. Additionally, in order to capture the effect of the topology on the performance of the source estimator, we bound and evaluate the maximum possible values that the proposed metric can achieve for some specific networks.

II. Source localization model

We assume a widely studied Susceptible-Infected propagation model, where once a node is infected (informed), it remains infected (informed) [1], [5]. Initially, there is only a single infected/informed node in the network, the source node. At a known time, assumed 0, the source node initiates the network diffusion, modeling the scenarios when some known external event triggers propagation. We will assume the network to be an undirected graph, as infections and rumors spread through contact and ties which are typically bidirectional. Additionally, we will consider only a connected network, meaning that there is a path between any two nodes in a network, otherwise some parts of the network would be isolated and would not be relevant to the diffusion process. 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. Given such a model, the true time of infection of a node i corresponds to its distance to the source s, i.e. the number of edges in the shortest path between the node and the source, denoted as $d(i,s)$. Assuming that resources are limited, only a subset of nodes $O=\{o_1,\dots,o_r\}$ is monitored. The source is then identified based on the infection times of observer nodes. Due to scarcity of the resources, the number of observer nodes r is typically much smaller than the total number of nodes N.

As a result of the the existence of an incubation period, the exact time when a node becomes infected/informed is not known. Instead, the observations represent noisy versions of the observer node's distance to the unknown source. The random duration of the incubation period is modeled as additive noise with exponential distribution. Given a source node s, the *true* infection time of each observer o_l, $l = 1, \ldots, r$ we compactly denote $a_l^s = d(o_l, s)$. Denoting the *observed* infection time of observer o_l as x_l, we have

$$x_l = d(o_l, s) + n_l = a_l^s + n_l$$

where n_l is an exponential random variable with density $p(n) = \lambda e^{-\lambda n} u(n)$, where $u(n)$ is the discrete step function ($u(n) = 1$ for $n \geq 0$ and $u(n) = 0$ for $n < 0$). Stacking the observations from all observer nodes into a vector, we obtain

$$\boldsymbol{x} = \boldsymbol{d}(O, s) + \boldsymbol{n} = \boldsymbol{a}^s + \boldsymbol{n}.$$

Since we assume the noise in different nodes to be independent, the observation density is given by

$$p(\boldsymbol{x}; \boldsymbol{s}) = \lambda^r e^{-\lambda \sum_{l=1}^{r} x_l - a_l^s} \prod_{l=1}^{r} u(x_l - a_l^s). \quad (1)$$

Next, we frame source localization as a multiple hypothesis testing problem. Each node $s = 1, \ldots, N$ in a network is a potential source candidate and represents a hypothesis H_s with conditional density of observations $p_s(\boldsymbol{x})$ given by (1). Without any prior knowledge on the source position, we assume each hypothesis has the same prior probability $\pi = 1/N$. The maximum a posteriori probability decision rule minimizes the Bayesian probability of error, and in our case of equal priors, it corresponds to the maximum likelihood estimator [6]. Then the source s^* is selected according to

$$s^* = \arg\max_s p_s(\boldsymbol{x}),$$

where the ties are broken at random. The corresponding probability of error is typically difficult to calculate and usually not tractable. Consequently, determining the subset of observer nodes that minimizes the overall error is not analytically feasible. Instead, we resort to a popular approach of asymptotically characterizing the error decay for vanishing noise, and then selecting the subset of observer nodes that minimizes the dominant error exponent.

III. Observer node selection

A. Error exponent

In the binary hypothesis case, with hypotheses H_i and H_j, the overall probability of error of the maximum likelihood estimator P_e is bounded as [6]

$$P_e \leq \int \left(\frac{1}{\pi_i} p_i(\boldsymbol{x}) \right)^{\alpha} \left(\frac{1}{\pi_j} p_j(\boldsymbol{x}) \right)^{1-\alpha} d\boldsymbol{x}, \quad (2)$$

for $0 \leq \alpha \leq 1$. Extending to the multiple hypothesis case, the bound on the overall error probability is

$$P_e \leq \sum_{i=1}^{N} \sum_{j>i} P_e(i,j) = \sum_{i=1}^{N} \sum_{j>i} e^{\log P_e(i,j)}, \quad (3)$$

where $P_e(i,j)$ is the pairwise error for hypotheses H_i and H_j, bounded as shown in (2). For the source localization problem, we insert (1) and the uniform prior into (2) to obtain

$$\begin{aligned} P_e(i,j) &\leq \frac{1}{N} \int p(\boldsymbol{x}; i)^{\alpha} p(\boldsymbol{x}; j)^{1-\alpha} d\boldsymbol{x} \\ &= \frac{1}{N} \prod_{l=1}^{r} \int p(x_l; i)^{\alpha} p(x_l; j)^{1-\alpha} dx_l \\ &= \frac{1}{N} \prod_{l=1}^{r} \lambda e^{\lambda\left(\alpha a_l^i + (1-\alpha) a_l^j\right)} \int_{\max\left\{a_l^i, a_l^j\right\}}^{\infty} e^{-\lambda x_l} dx_l \\ &= \frac{1}{N} \prod_{l=1}^{r} e^{\lambda\left(\alpha a_l^i + (1-\alpha) a_l^j\right) - \lambda \max\left\{a_l^i, a_l^j\right\}}. \end{aligned} \quad (4)$$

The error probability bound (2) holds for $0 \leq \alpha \leq 1$, and the minimum can be taken to achieve a tighter bound [6]. Substituting α that minimizes the bound and taking the logarithm on both sides of (4), we further get

$$\log P_e(i,j) \leq \log \frac{1}{N} - \lambda \mathcal{C}(i,j), \quad (5)$$

where

$$\mathcal{C}(i,j) = \begin{cases} \sum_{l=1}^{r} \mathbb{I}_{\left\{a_l^i > a_l^j\right\}} \left(a_l^i - a_l^j\right), & \mathbf{1}^T \boldsymbol{a}^i \geq \mathbf{1}^T \boldsymbol{a}^j \\ \sum_{l=1}^{r} \mathbb{I}_{\left\{a_l^j > a_l^i\right\}} \left(a_l^j - a_l^i\right), & \mathbf{1}^T \boldsymbol{a}^j > \mathbf{1}^T \boldsymbol{a}^i \end{cases}, \quad (6)$$

$\mathbb{I}$ is an indicator function for the condition in $\{.\}$ and $\mathbf{1}$ is a vector where all entries are 1. Plugging back (5) into the overall bound for error probability (3), we finally have

$$P_e \leq \sum_{i=1}^{N} \sum_{j>i} e^{\log \frac{1}{N} - \lambda \mathcal{C}(i,j)}. \quad (7)$$

Note that for different subsets of observer nodes O_r, we obtain different distances $\boldsymbol{a}^i = \boldsymbol{d}(O, i)$ of each node $i = 1, \ldots, N$, to the selected set O_r. This in turn influences the values obtained for the exponents $\mathcal{C}(i,j)$. Now we are ready to state the proposed criterion for selection of the observer nodes.

Theorem 1: Maximum likelihood estimation of the source based on the observations of a subset of observer nodes O_r^* that has the highest $\min_{i \neq j} \mathcal{C}(i,j)$ for all pairs $i, j = 1, \ldots, N$, $i \neq j$, where $\mathcal{C}(i,j)$ is given by (6), achieves the fastest error decay for vanishingly small noise among all possible observer subsets of cardinality r.

Proof: Let (i,j) be the pair of nodes that determines the worst exponent of (7), i.e.,

$$(i,j) = \arg\min_{(k,l)} \mathcal{C}(k,l).$$

Let us assume $\mathbf{1}^T \boldsymbol{a}^i < \mathbf{1}^T \boldsymbol{a}^j$. Analyzing the upper bound on the total error (7) asymptotically, for vanishing noise, we obtain the following bound on the error exponent

$$\limsup_{\lambda \to \infty} \frac{\log P(e)}{\lambda} \leq -\mathcal{C}(i,j). \quad (8)$$

In order to prove that the error decays for the vanishing noise with the rate of the worst exponent, i.e.,

$$\lim_{\lambda \to \infty} \frac{\log P(e)}{\lambda} = -\mathcal{C}(i,j), \quad (9)$$

we first need to prove the following lower bound holds

$$\liminf_{\lambda\to\infty} \frac{\log P(e)}{\lambda} \geq -\mathcal{C}(i,j). \tag{10}$$

Let $P(e|H_s)$ denote the probability of an error given that hypothesis H_s is true. Then we have

$$\begin{aligned} P(e) &= \sum_{s=1}^{N} P(e|H_s)P(H_s) = \frac{1}{N}\sum_{s=1}^{N} P(e|H_s) \\ &\geq \frac{1}{N}P(e|H_i) = \frac{1}{N}\int_{\Omega_i} p(\boldsymbol{x};H_i)d\boldsymbol{x} \end{aligned} \tag{11}$$

where $\Omega_i := \{\boldsymbol{x} : p(\boldsymbol{x};H_i) < p(\boldsymbol{x};H_k) \text{ for some } k \neq i\}$. Now, we consider the set $\Omega_{ij} = \{\boldsymbol{x} : \boldsymbol{x} \geq \max\{\boldsymbol{a}^i,\boldsymbol{a}^j\}\}$. Note that $\Omega_{ij} \subset \Omega_i$; as indeed, if $\boldsymbol{x} \in \Omega_{ij}$ then

$$p(\boldsymbol{x};H_i) = \lambda^r e^{-\lambda \sum_{l=1}^{r} x_l - a_l^i} \text{ and } p(\boldsymbol{x};H_j) = \lambda^r e^{-\lambda \sum_{l=1}^{r} x_l - a_l^j}.$$

Due to the assumption $\mathbf{1}^T\boldsymbol{a}^i < \mathbf{1}^T\boldsymbol{a}^j$ we see that $p(\boldsymbol{x};H_i) < p(\boldsymbol{x};H_j)$. Thus, $\boldsymbol{x} \in \Omega_i$.

Now, we further evaluate (11) as

$$\begin{aligned} P(e) &\geq \frac{1}{N}\int_{\Omega_{ij}} p(\boldsymbol{x};H_i)d\boldsymbol{x} \\ &= \frac{1}{N}\prod_{l=1}^{r} \lambda e^{\lambda a_l^i} \int_{\max\left\{a_l^i,a_l^j\right\}}^{\infty} e^{-\lambda x_l} dx_l \\ &= \frac{1}{N}\prod_{l=1}^{r} e^{\lambda a_l^i - \lambda \max\left\{a_l^i,a_l^j\right\}} \\ &= \frac{1}{N} e^{-\lambda \sum_{l=1}^{r}\left(\max\left\{a_l^i,a_l^j\right\} - a_l^i\right)}. \end{aligned} \tag{12}$$

Next, we have $\max\left\{a_l^i,a_l^j\right\} - a_l^i = 0$, for $a_l^i > a_l^j$ and $a_l^j - a_l^i$ otherwise. Finally, we get the lower bound

$$P(e) \geq \frac{1}{N} e^{-\lambda \sum_{l=1}^{r} \mathbb{I}_{\left\{a_l^j - a_l^i\right\}}\left(a_l^j - a_l^i\right)}, \text{ for } \mathbf{1}^T\boldsymbol{a}^j > \mathbf{1}^T\boldsymbol{a}^i.$$

Reversing i and j we get the other part of the expression of $\mathcal{C}(i,j)$ in (6), for the condition $\mathbf{1}^T\boldsymbol{a}^i \geq \mathbf{1}^T\boldsymbol{a}^j$. When $\boldsymbol{a}^i = \boldsymbol{a}^j$, then the two hypotheses cannot be discerned even for the noiseless case, and hence there is no exponential error decay. Expression (6) still holds, as the error exponent then evaluates to zero.

Since both the upper (8) and lower (10) bound hold, so does (9), and we conclude that the error exponent for vanishing noise equals the smallest $\mathcal{C}(i,j)$. Hence, the subset O_r that has the highest $\min_{i\neq j}\mathcal{C}(i,j)$ achieves the greatest rate at which the error probability decreases with vanishing noise. ■

$\lambda\mathcal{C}(i,j)$, derived above, actually represents the Chernoff distance between distributions $p(\boldsymbol{x};i)$ and $p(\boldsymbol{x};j)$, i.e.

$$\lambda\mathcal{C}(i,j) = -\min_{0\leq\alpha\leq 1} \log \int p(\boldsymbol{x};i)^{\alpha} p(\boldsymbol{x};j)^{1-\alpha} d\boldsymbol{x},$$

and it represents the highest achievable exponent for the decay of the error probability in the binary hypothesis case with asymptotically large number of observations. The result of Theorem 1 mirrors the result available for multiple hypothesis testing, where the best achievable exponent of the error decay for increasing number of observations is the minimum of $N(N-1)$ Chernoff distances between $p_i(\boldsymbol{x})$ and $p_j(\boldsymbol{x})$ [7].

The criterion from Theorem 1 is optimal for minimizing the error probability only for $\lambda \to \infty$. However, since exact minimization of the error probability is often computationally unfeasible, optimization of sub-optimum performance measures, such as the Chernoff distance, is a popular approach in practice [8].

B. Properties of the selection criterion

We have shown that the Chernoff distance between distributions characterizing source candidates i and j equals $\lambda\mathcal{C}(i,j)$, where $\mathcal{C}(i,j)$ has the form given by (6). Note that the term $\mathcal{C}(i,j)$ is completely independent of the noise level λ and captures the effect of choice of observer nodes and graph topology on the source localization error in a noiseless scenario. Next, we show that $\mathcal{C}(i,j)$ has some other interesting properties. First, we slightly change the notation, from $\mathcal{C}(i,j)$ to $\mathcal{C}(\boldsymbol{a}^i,\boldsymbol{a}^j)$, for easier presentation, as the latter form emphasizes that (6) is directly a function of vectors $\boldsymbol{a}^i$ and $\boldsymbol{a}^j$.

Theorem 2: $\mathcal{C}(\boldsymbol{a}^i,\boldsymbol{a}^j)$, given by (6), is a metric.

Proof: To show that $\mathcal{C}(\boldsymbol{a}^i,\boldsymbol{a}^j)$ satisfies all the conditions for a metric, we first reformulate (6). Let $\boldsymbol{v}^+$ denote a vector where negative entries of $\boldsymbol{v}$ are replaces with 0, i.e., $v_i^+ = \max\{0,v_i\}$. Similarly, let $\boldsymbol{v}^-$ denote a vector where positive entries of $\boldsymbol{v}$ are set to 0, but instead of negative values, their absolute values are taken. Then for any vector $\boldsymbol{v}$, we have $\boldsymbol{v} = \boldsymbol{v}^+ - \boldsymbol{v}^-$ and

$$\begin{aligned} \|\boldsymbol{v}\|_1 &= \mathbf{1}^T\boldsymbol{v}^+ + \mathbf{1}^T\boldsymbol{v}^- \\ \mathbf{1}^T\boldsymbol{v} &= \mathbf{1}^T\boldsymbol{v}^+ - \mathbf{1}^T\boldsymbol{v}^-. \end{aligned} \tag{13}$$

Summing the two expressions of (13) gives

$$\mathbf{1}^T\boldsymbol{v}^+ = \frac{1}{2}\left(\|\boldsymbol{v}\|_1 + \mathbf{1}^T\boldsymbol{v}\right). \tag{14}$$

Now we can rewrite (6) as

$$\begin{aligned} &\mathcal{C}(\boldsymbol{a}^i,\boldsymbol{a}^j) = \mathbb{I}_{\{\mathbf{1}^T\boldsymbol{a}^i > \mathbf{1}^T\boldsymbol{a}^j\}}\mathbf{1}^T\left(\boldsymbol{a}^i - \boldsymbol{a}^j\right)^+ + \\ &\mathbb{I}_{\{\mathbf{1}^T\boldsymbol{a}^i = \mathbf{1}^T\boldsymbol{a}^j\}}\mathbf{1}^T\left(\boldsymbol{a}^i - \boldsymbol{a}^j\right)^+ + \mathbb{I}_{\{\mathbf{1}^T\boldsymbol{a}^i < \mathbf{1}^T\boldsymbol{a}^j\}}\mathbf{1}^T\left(\boldsymbol{a}^j - \boldsymbol{a}^i\right)^+. \end{aligned} \tag{15}$$

When $\mathbf{1}^T\boldsymbol{a}^i = \mathbf{1}^T\boldsymbol{a}^j$ holds, then $\mathbf{1}^T\left(\boldsymbol{a}^i - \boldsymbol{a}^j\right) = 0$ and substituting $\boldsymbol{v} = \boldsymbol{a}^i - \boldsymbol{a}^j$ in (14), we obtain $\mathbf{1}^T\left(\boldsymbol{a}^i - \boldsymbol{a}^j\right)^+ = \frac{1}{2}\left\|\boldsymbol{a}^i - \boldsymbol{a}^j\right\|_1$. Again, using (14), we can rewrite (15) as

$$\begin{aligned} \mathcal{C}(\boldsymbol{a}^i,\boldsymbol{a}^j) =& \mathbb{I}_{\{\mathbf{1}^T\boldsymbol{a}^i > \mathbf{1}^T\boldsymbol{a}^j\}}\frac{1}{2}\left(\left\|\boldsymbol{a}^i - \boldsymbol{a}^j\right\|_1 + \mathbf{1}^T\left(\boldsymbol{a}^i - \boldsymbol{a}^j\right)\right) \\ &+ \mathbb{I}_{\{\mathbf{1}^T\boldsymbol{a}^i = \mathbf{1}^T\boldsymbol{a}^j\}}\frac{1}{2}\left\|\boldsymbol{a}^i - \boldsymbol{a}^j\right\|_1 \\ &+ \mathbb{I}_{\{\mathbf{1}^T\boldsymbol{a}^i < \mathbf{1}^T\boldsymbol{a}^j\}}\frac{1}{2}\left(\left\|\boldsymbol{a}^j - \boldsymbol{a}^i\right\|_1 + \mathbf{1}^T\left(\boldsymbol{a}^j - \boldsymbol{a}^i\right)\right) \end{aligned}$$

Finally, we have

$$\mathcal{C}(\boldsymbol{a}^i,\boldsymbol{a}^j) = \frac{1}{2}\left\|\boldsymbol{a}^i - \boldsymbol{a}^j\right\|_1 + \frac{1}{2}\left|\mathbf{1}^T\boldsymbol{a}^i - \mathbf{1}^T\boldsymbol{a}^j\right|. \tag{16}$$

From (16), we directly see that positive-definiteness and symmetry hold. Substituting $\boldsymbol{x} = \boldsymbol{a}^i - \boldsymbol{a}^j$ and $\boldsymbol{y} = \boldsymbol{a}^j - \boldsymbol{a}^k$ into $\|\boldsymbol{x}+\boldsymbol{y}\|_1 \leq \|\boldsymbol{x}\|_1 + \|\boldsymbol{y}\|_1$, and $\boldsymbol{x} = \mathbf{1}^T\boldsymbol{a}^i - \mathbf{1}^T\boldsymbol{a}^j$ and

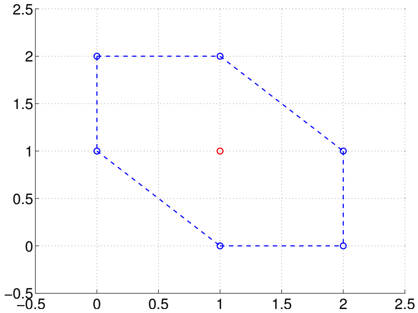

Fig. 1. Unit ball in two-dimensional space for $\mathcal{C}(\boldsymbol{x}, \boldsymbol{y})$, centered at point $(1, 1)$ denoted with a red circle.

$\boldsymbol{y} = \mathbf{1}^T \boldsymbol{a}^j - \mathbf{1}^T \boldsymbol{a}^k$ into $|\boldsymbol{x} + \boldsymbol{y}| \leq |\boldsymbol{x}| + |\boldsymbol{y}|$, we see that $\mathcal{C}(\boldsymbol{a}^i, \boldsymbol{a}^k) \leq \mathcal{C}(\boldsymbol{a}^i, \boldsymbol{a}^j) + \mathcal{C}(\boldsymbol{a}^j, \boldsymbol{a}^k)$ also holds. Thus $\mathcal{C}(\boldsymbol{a}^i, \boldsymbol{a}^j)$ is a metric. ∎

In the proof of Theorem 2, we have made no assumption on the arguments $\boldsymbol{a}^i$ and $\boldsymbol{a}^j$, and we have shown that $\mathcal{C}(\boldsymbol{x}, \boldsymbol{y})$ is a metric for any two real vectors $\boldsymbol{x}$ and $\boldsymbol{y}$. A unit ball defined with this metric is shown in Figure 1. The blue line in Figure 1 contains all the points which are at distance 1 from the red point, where distance is defined with $\mathcal{C}(x, y)$. Since in our specific case spatial coordinates represent graph distances to observer nodes, which are integers, only the vertices of the unit ball are possible values for $\boldsymbol{a}^i$ at distance 1 (given that the ball is centered at a point with integer coordinates).

Although each pairwise distance can be calculated in $\mathcal{O}(r)$ time, still $\binom{N}{2}$ pairwise Chernoff distances need to be calculated for each of $\binom{N}{r}$ different subsets that are examined in order to select an optimal subset. Next, we present some basic bounds for the smallest pairwise distance. Let $C(O_r)$ denote the smallest $\mathcal{C}(i, j)$ for all i, j pairs for a fixed subset O_r, while $C(r)$ denotes the smallest $\mathcal{C}(i, j)$ for all i, j pairs for any observer subset of cardinality r.

Theorem 3: The following bounds hold

$$C(O_r) > 0 \iff O_r \text{ is a resolving set.} \tag{17}$$

$$C(O_r) \leq C(O_r \cup o_{r+1}) \tag{18}$$

$$C(r) \leq r. \tag{19}$$

Proof: From (6) it follows that $\mathcal{C}(i, j) = 0$ if and only if $\boldsymbol{a}^i = \boldsymbol{a}^j$, i.e., if nodes i and j are equidistant to all the observer nodes. Then the set O_r cannot be a resolving set, which is by definition a set of nodes O such that each pair of nodes has a different distance to at least one node from O [9].

The following inequality (18) is intuitive, as it states that the distances between distributions that characterize source candidates will not decrease if a new node is observed, and consequently, the error will not decrease any slower if an additional observation is included. Including a new observer node adds a new entry to $\boldsymbol{a}^i$ and $\boldsymbol{a}^j$, which cannot decrease either $(\boldsymbol{a}^i - \boldsymbol{a}^j)^+$ or $(\boldsymbol{a}^j - \boldsymbol{a}^i)^+$. Then each pairwise distance as seen from (15) does not decrease. This also holds for the smallest distance. Note that here we only claim that for a fixed subset, a new observer cannot decrease the distance, and hence we also have $C(r) \leq C(r+1)$, but it is not generally true that $C(O_r) \leq C(O_{r+1})$, if $O_r \not\subset O_{r+1}$.

Inequality (19) comes from analyzing $\mathcal{C}(i, j)$, when i and j are neighbor nodes. For any observer o_l, we have that $d(i, o_l) \in \{d(j, o_l) - 1, d(j, o_l), d(j, o_l) + 1\}$ when i and j are connected by an edge. Hence $\left|a_l^i - a_l^j\right| = |d(i, o_l) - d(j, o_l)| \leq 1$, and for r observers we have $\left\|\boldsymbol{a}^i - \boldsymbol{a}^j\right\|_1 \leq r$ and $\left|\mathbf{1}^T \boldsymbol{a}^i - \mathbf{1}^T \boldsymbol{a}^j\right| \leq r$. Using these in (16), we have that $\mathcal{C}(i, j) \leq r$, when i and j are neighbors, and the same then holds for the minimal $\mathcal{C}(i, j)$. ∎

In order to understand how the source localization error differs across different topologies and what is the best possible exponent that can be reached with any subset selection for a given topology, we analyze the idealized case when all nodes are monitored.

Theorem 4: For a complete network $C(N) = 1$, for a star $C(N) = 2$, for a path $C(N) = \left\lceil \frac{N}{2} \right\rceil$ and for a tree $C(N) \leq \max\{2(\tau_a + 1), 2(\tau_b + 1)\}$, where a and b are nodes at distance 1 to a common ancestor, and τ_s is the number of descendants of node s.

Proof: Observing all nodes in a complete graph, for any two nodes i and j, we have that $\boldsymbol{d}(i, O_N)$ and $\boldsymbol{d}(j, O_N)$ differ only in two entries, as $0 = d(i, i) \neq d(i, j) = 1$ and vice versa, hence $C(N) = \mathcal{C}(i, j) = 1$.

In a star network, $\boldsymbol{d}(i, O_N)$ and $\boldsymbol{d}(j, O_N)$ for any two leaf nodes differ only in the entries i and j. Since $d(i, j) = 2$, this pair determines the minimal $\mathcal{C}(i, j)$, as $\boldsymbol{d}(i, O_N)$ and $\boldsymbol{d}(c, O_N)$, where c represents the central node, differ in all the entries.

In a path network, we label the nodes sequentially, and denote with $\beta = \left\lfloor \frac{d(i,j)}{2} \right\rfloor^2$ if $d(i, j)$ is an odd number, and $\beta = \frac{d(i,j)}{2} \left(\frac{d(i,j)}{2} - 1 \right)$ otherwise. Then it can be shown, and we omit the proof for brevity, that $\mathcal{C}(i, j) = \beta + d(i, j) + d(i, j) \max\{i - 1, N - i - d(i, j)\}$. The minimum is reached for $d(i, j) = 1$ and $i = \left\lceil \frac{N}{2} \right\rceil$, and equals $\left\lceil \frac{N}{2} \right\rceil$.

The bound for $C(N)$ in trees is obtained by analyzing $\mathcal{C}(i, j)$, when a and b are both at a distance 1 to a common ancestor. Then, nodes a and b are equidistant to all the nodes except themselves and their descendants. For any node o that is a descendent of a, we have $d(o, a) - d(o, b) = d(a, b) = 2$, from which the bound follows. Although not straightforward, this bound is useful in cases such as when a tree has two leaves connected to the same node (like in a star network). Then regardless of the remaining structure of the tree, or selected observer subset, the error exponent can be at most 2λ. ∎

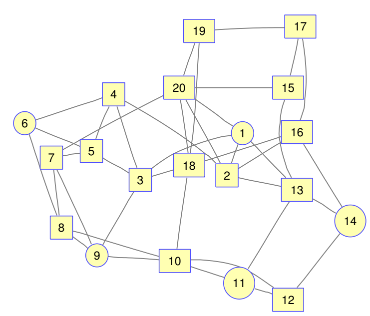

Fig. 2. A small world network example used in simulations

IV. Simulation results

The proposed criterion for observer node selection is only optimal for vanishing noise. Here, we illustrate its usefulness even in a more practical scenario, on a randomly generated small world network with 20 nodes, shown in Figure 2. Small world networks represent a popular model for social networks and the Internet, as they have a small diameter and a large clustering coefficient. Although networks of interest are much larger, we present a small example, as we wish to calculate the exact probability of error for different observer node subsets. This is computationally very intensive and we are able to calculate the exact error probability only for smaller networks and observer subsets. However, computing the Chernoff distances for a fixed subset can be performed much faster, even if it is in the order of $\mathcal{O}(N^2)$. We analyze the performance of different subsets of 5 randomly selected nodes. There are in total $\binom{20}{5} = 15504$ possible subsets, of which around half have a minimum pairwise Chernoff distance of 0, the other half have the minimum distance of 1λ, while only a single subset has the distance of 2λ. The nodes of this subset with the highest minimal Chernoff distance are labeled with circles in Figure 2.

We have randomly generated 200 subsets, and for each subset we have calculated the exact error probability for noise level $\lambda = 1$. Since the true infection times in a network correspond to graph distances, which take integer values (in our example network, they range from 0 to 4), additive noise with mean 1 is significant. Figure 3 shows the distribution of error probabilities vs. the minimum Chernoff distance for these subsets. Each scatter point represents a random subset of 5 nodes. From Figure 3 we observe that even for this level of noise, the subset with the highest distance performs better than the other subsets. We also note that subsets with the same Chernoff distance display a range of different error probabilities, but the average error decreases with increasing distance.

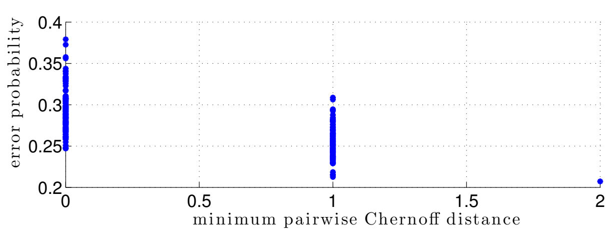

Fig. 3. Probability of error of different observer subsets of 5 nodes for $\lambda = 1$

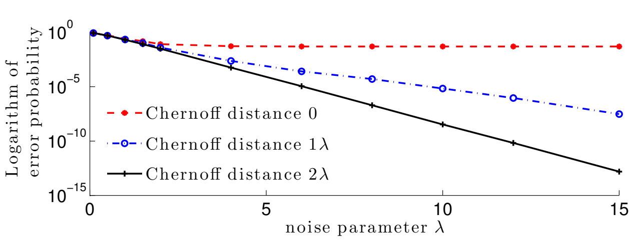

Fig. 4. Decay of error probability with vanishing noise for subsets of different minimum Chernoff distance

Next, for each value of a minimal Chernoff distance, we have selected a subset that achieved the lowest error, out of 200 randomly generates subsets shown in the previous Figure. For these three subsets, we have evaluated the probability of error for different noise levels. As λ increases, the noise level decreases and the obtained curves are shown on a logarithmic scale in Figure 4. As expected, the error probability for the subset with the highest Chernoff distance decays with the largest slope, illustrating the result stated by Theorem 1.

V. Conclusion

We have formulated localization of a source of network diffusion based on infection times of a subset of nodes as a multiple hypothesis testing problem. To accommodate for the existence of a variable incubation time of each node, we have assumed that the observations are corrupted by additive exponential noise. In order to bound the error probability we have derived the expression for the Chernoff distance between distributions that characterize different suspect nodes. We have shown that the subset of observer nodes with the highest minimal pairwise Chernoff distance achieves the fastest error decay for vanishing noise. We have proved that the proposed criterion represents a true metric, and we have provided bounds for it for some specific topologies. We have illustrated the usefulness of the selection criterion in achieving low probability of error even in the presence of non-negligible noise through a simulation example. As the error exponent does not depend on the noise level, for future work we leave the investigation of its applicability in scenarios with noise of different distributions.

Acknowledgment

The work of S. Zejnilović and J. Gomes was partially supported by Fundação para a Ciência e a Tecnologia (project FCT [UID/EEA/50009/2013] and a PhD grant from the Carnegie Mellon-Portugal program) and EU FP7 project MORPH (grant agreement no. 288704).

The work of J. Xavier was partially supported by Fundação para a Ciência e a Tecnologia (FCT), PTDC/EMS-CRO/2042/2012.

References

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

[2] E. Seo, P. Mohapatra, and T. F. Abdelzaher, "Identifying rumors and their sources in social networks," *SPIE DSS*, 2012.

[3] S. Zejnilović, J. Gomes, and B. Sinopoli, "Network observability and localization of the source of diffusion based on a subset of nodes," *Allerton*, 2013.

[4] A. Martinez, "Communication by energy modulation: The additive exponential noise channel," *IEEE Transactions on Information Theory*, vol. 57, no. 6, Jun 2011.

[5] D. Shah and T. Zaman, "Rumors in a network: who's the culprit?" *IEEE Transactions on Information Theory*, 2011.

[6] T. Cover and J. Thomas, *Elements of Information Theory 2nd Edition*. Wiley Series in Telecommunications and Signal Processing, 2006.

[7] N. Salikhov, "On one generalization of Chernov's distance," *Theory of Probability and Its Applications*, vol. 43, no. 2, pp. 239–255, 1999.

[8] T. Kailath, "The divergence and Bhattacharyya distance measures in signal selection," *IEEE Transactions on Communication technology*, vol. 15, no. 1, 1967.

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