

# Ph.D. Defense

## Localizing a diffusion source on graphs: analysis and design of node selection strategies

**Sabina Zejnilovic**

April 26, 2016

| 12:00PM EST

| PH B34



### Abstract:

Many different phenomena, such as spreading of infectious diseases in networks and dissemination of information through social networks are modeled as a diffusion over a network of nodes. Localizing the source of diffusion becomes crucial for quarantine efforts, as well as for identifying influential individuals who start certain trends. In many real world networks, due to their size, it is unfeasible to observe the infection times of all nodes. In this case, identification of the source node is carried out based only on a subset of nodes, called observer nodes. The choice of observer nodes heavily impacts the accuracy of source localization.

In this thesis we specifically focus on the analysis and design of observer selection strategies with the goal of understanding which are the best nodes to observe that contribute the most to the disambiguation of the source on graphs for different diffusion scenarios. We introduce and study the concept of network observability which characterizes the ability to identify the source node based only on partial network knowledge, provided by the observer nodes. There are four main contributions of our work:

- We study the effect of the choice of observers on source localization when there is no uncertainty present - in a purely deterministic scenario. We model the dynamics of network diffusion as a linear time-varying system, and present a necessary and sufficient condition for network observability in the context of the proposed model, as well as in the context of graph theory. Relating the observer selection problem to a known problem of finding a resolving set of a graph, we design observer selection strategies with performance guarantees.
- We extend the concept of observability for the scenario when the network topology is not fully known and the edges between different communities are hidden. We present necessary and sufficient conditions for network observability when the components are all trees, cycles, grids and complete graphs. We also give sufficient conditions when the components are of arbitrary structure.
- We formulate a metric, based on error exponents, that can be used to compare different observer subsets from the perspective of source localization error. We evaluate the error exponent for vanishing noise for three different diffusion scenarios; in each, the infection times are modeled as random variables with different distributions: Gaussian, Laplacian and exponential.
- We develop different sequential selection strategies, optimal and efficient sub-optimal, for dynamic observer selection in both deterministic and stochastic setting, and demonstrate the applicability of one of the proposed strategies on a real-world dataset of a cholera outbreak.

### Committee

Prof. Bruno Sinopoli,  
advisor

Prof. Joao Gomes,  
University of Lisbon,  
advisor

Prof. Pulkit Grover

Prof. Massimo  
Franceschetti,  
University of California,  
San Diego

Prof. Joao Xavier,  
University of Lisbon

Prof. Joao Sobrinho,  
University of Lisbon