SUBGRAPH DENSITY AND EPIDEMICS OVER NETWORKS

June Zhang and José M.F. Moura

Carnegie Mellon University
Department of Electrical and Computer Engineering
Pittsburgh, PA 15213 USA
Email: {junez@andrew, moura@ece}.cmu.edu

ABSTRACT

We model a SIS (susceptible-infected-susceptible) epidemics over a static, finite-sized network as a continuous-time Markov process using the scaled SIS epidemics model. In our previous work, we derived the closed form description of the equilibrium distribution that explicitly accounts for the network topology and showed that the most probable equilibrium state demonstrates threshold behavior. In this paper, we will show how subgraph structures in the network topology impact the most probable state of the long run behavior of a SIS epidemics (i.e., stochastic diffusion process) over any static, finite-sized, network.

***Index Terms*—** Reversible Markov process, equilibrium distribution, SIS epidemics, graph density, densest subgraph, k-densest subgraph, topology dependent random interaction model, networks

1. INTRODUCTION

The study of networks is becoming increasingly important in a variety of disciplines. Incorporating explicitly the network topology leads to unexpected behaviors that cannot be predicted by standard models that discount topology [1].

In our work, we have been studying a diffusion process (e.g., epidemics) over a population whose interactions are captured by a fixed topology network. We are interested in understanding how network topology affects the epidemics process. Previous works by others have shown that the time in which an epidemics dies out or becomes endemic relates to the largest eigenvalue of the underlying network topology ([2], [3], [4]). However, the model assumed by these references does not consider the possibility that healthy agents may spontaneously become infected *without* viral transmission from infected neighbors; this type of infection is exogenous to the population. In [5], [6], we introduced the *scaled SIS (susceptible-infected-susceptible) epidemics* model, which accounts for both exogenous infection and exogenous healing, in addition to endogenous infection (i.e., viral transmission from infected neighbors). We derived the closed form expression of the equilibrium distribution of an epidemics over an arbitrary network topology. This paper continues our analysis of solving for the most probable network state (i.e., the network state with the maximum equilibrium probability). We introduce the model and the relevant question we address in section 2. In sections 3 and 4, we interpret network states as subgraphs of the network. In section 5, we determine how the subgraph density affects the most probable network state. Finally, section 6 concludes the paper.

This work was partially supported by AFOSR grant FA95501010291, and by NSF grants CCF1011903 and CCF1018509

2. MODEL

Consider a population of N agents whose interconnections are represented by a static, simple, unweighted, undirected, connected graph, $G(V, E)$, where $V(G)$ is the set of vertices and $E(G)$ is the set of edges. The topology of G is captured by the symmetric, $N \times N$ adjacency matrix, A. The state of the ith agent is denoted by x_i. Each agent can be in one of two possible states: healthy ($x_i = 0$) or infected ($x_i = 1$); we use the term *infectives* to refer to infected agents and *susceptibles* to refer to healthy agents. Let

$$\mathbf{x} = [x_1, x_2, \ldots, x_N]^T.$$

We will refer to x_i as the *agent state* and to $\mathbf{x}$ as either the *network state* or the *network configuration* in this paper. Let $\mathcal{X} = \{\mathbf{x}\}, |\mathcal{X}| = 2^N$, be the network state space.

The *scaled SIS (susceptible-infected-susceptible) epidemics* model captures the evolution of the network state over time. Let $X(t) = \mathbf{x}$ be the state of the network at time t, $t \geq 0$. The SIS epidemics assumes that infectives can heal and become reinfected; this means that the total number of agents in the population remains constant, unlike SIR epidemics (susceptible-infected-removed). We showed in [5] that we can model $X(t)$ as a continuous-time Markov process.

The time the Markov process spends in a particular state is random and exponentially distributed. The transition rate matrix Q is the object of interest with continuous-time Markov process. Adapting notation from [7], we define 2 operators on the network state, $\mathbf{x} = [x_1, x_2, \ldots x_i, \ldots x_j, \ldots, x_N]^T$:

$$H_i\mathbf{x} = [x_1, x_2, \ldots, x_i = 1, \ldots, x_N]^T$$
$$H_{j\bullet}\mathbf{x} = [x_1, x_2, \ldots, x_j = 0, \ldots, x_N]^T.$$

The operator H_i defines the operation that agent i becomes infected. If agent i is already infected, the operator does nothing. The operator $H_{j\bullet}$ defines the operation that agent j is healed. If agent j is already uninfected, the operator does nothing.

The two types of state transitions $X(t)$ captures corresponding to healing and infection events respectively:

1) $X(t)$ jumps to the network state where, a single agent, the jth agent ($j = 1, \ldots, N$), is healed with transition rate:

$$q(\mathbf{x}, H_{j\bullet}\mathbf{x}) = \mu, \quad \mathbf{x} \neq H_{j\bullet}\mathbf{x}. \tag{1}$$

2) $X(t)$ jumps to the network state where a single agent, the ith agent ($i = 1, 2, \ldots, N$), is infected with transition rate

$$q(\mathbf{x}, H_i\mathbf{x}) = \lambda\gamma^{d_i} \tag{2}$$

where $d_i = \sum_{j=1}^{N} \mathbb{1}(x_j = 1)A_{ij}$ is the number of infected neighbors of node i. The symbol $\mathbb{1}(\cdot)$ is the indicator function, and $A = [A_{ij}]$ is the adjacency matrix of G.

We call the parameter μ the healing rate. The parameter λ is the exogenous (i.e., spontaneous) infection rate since if $d_i = 0$, the infection rate is still λ. The parameter γ is the endogenous infection rate since it is dependent on the number of infected neighbors; consequently, we will also refer to γ as the topology dependent parameter and to λ and μ as topology independent parameters. The scaled SIS epidemics model does not have an absorbing state (i.e., network states where all agents are infectives or all agents are susceptibles) because it accounts for both exogenous infection and the exogenous healing [2], [3], [4].

We can derive the closed form equilibrium distribution of the scaled SIS epidemics model. In contrast, there is no known closed form equilibrium distribution for the more commonly used topology dependent SIS epidemics model used in research [2, 3, 8]. We call these type of models the *additive SIS epidemics model* since their infection transition rate $q(\mathbf{x}, H_i\mathbf{x}) = \lambda + d_i\gamma$.

2.1. Equilibrium Distribution and Most Probable Configuration

In [5], [6], we proved that the equilibrium distribution for $X(t)$ is

$$\pi(\mathbf{x}) = \frac{1}{Z}\left(\frac{\lambda}{\mu}\right)^{1^T\mathbf{x}} \gamma^{\frac{\mathbf{x}^T A\mathbf{x}}{2}}, \quad \mathbf{x} \in \mathcal{X} \tag{3}$$

where Z, known as the partition function, is the normalization constant and is defined as

$$Z = \sum_{\mathbf{x}\in\mathcal{X}} \left(\frac{\lambda}{\mu}\right)^{1^T\mathbf{x}} \gamma^{\frac{\mathbf{x}^T A\mathbf{x}}{2}}. \tag{4}$$

The equilibrium distribution is the true limiting distribution of the process. In [6], we solved for configuration(s) with the highest probability since these are configurations that we would most likely observe at equilibrium. Using (3), the most probable network state, $\mathbf{x}^*$, is

$$\mathbf{x}^* = \arg\max_{\mathbf{x}\in\mathcal{X}} \pi(\mathbf{x}) = \arg\max_{\mathbf{x}\in\mathcal{X}} \left(\frac{\lambda}{\mu}\right)^{1^T\mathbf{x}} \gamma^{\frac{\mathbf{x}^T A\mathbf{x}}{2}} \tag{5}$$

Our work in [6] focused on solving (5) for different network topologies without having to solve a binary integer programming problem. We divided the parameter space into 4 different regimes:

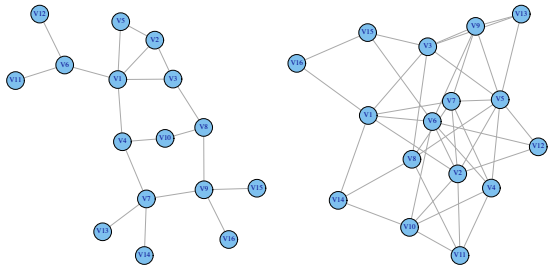
I) **Healing Dominant:** $0 < \frac{\lambda}{\mu} \leq 1, 0 < \gamma \leq 1$;

II) **Endogenous Infection Dominant:** $0 < \frac{\lambda}{\mu} \leq 1, \gamma > 1$;

III) **Exogenous Infection Dominant:** $\frac{\lambda}{\mu} > 1, 0 < \gamma \leq 1$;

IV) **Infection Dominant:** $\frac{\lambda}{\mu} > 1, \gamma > 1$.

We summarize the main results of [6]. In regimes I and IV, for any underlying network topology, $\mathbf{x}^* = \mathbf{x}^0 = [0, 0, \ldots, 0]$ and $\mathbf{x}^* = \mathbf{x}^N = [1, 1, \ldots, 1]$. Network topology dependence exists only in regime II and regime III, where the effect of the topology dependent process (controlled by γ) opposes the effect of the topology independent process (control by λ, μ). More interestingly for regimes II and III, $\mathbf{x}^*$ exhibits threshold behavior depending on both the model parameters and on the network topology. In the rest of this paper, we focus only on regime II.



(a) Graph A (b) Graph B

Fig. 1: Example Network Topologies

3. SPURIOUS MOST PROBABLE NETWORK STATE IN REGIME II

We showed in [6], that, in regime II) **Endogenous Infection Dominant:** $0 < \frac{\lambda}{\mu} \leq 1, \gamma > 1$, for three network topologies: k-regular, complete multipartite, and complete multipartite with k-regular islands, the solution to (5) can only be either $\mathbf{x}^0 = [0, 0, \ldots, 0]$, the configuration where all agents are susceptibles, or $\mathbf{x}^N = [1, 1, \ldots, 1]$, the configuration where all agents are infectives. Furthermore, we derived the exact threshold for when $\mathbf{x}^*$ transitions between these two configurations.

We observed that, for other network topologies, the solution to (5) may be configurations other than $\mathbf{x}^0$ or $\mathbf{x}^N$ for certain range of the parameters. We call these solutions *spurious* configurations. Consider graph A shown in Figure 1a and graph B shown in Figure 1b. Graph A is a 16 node Erdős-Rényi random graph, while graph B is a 16 node Watts-Strogatz graph [1]. For these network topologies, if we hold the λ and μ parameters constant, the solution to (5) is a function of γ. However, this function is highly discontinuous as shown in Figure 2 and Figure 3. We set $\lambda = 1, \mu = 2$ while varying γ from 1 to 4 with incremental step size of 0.1. On the Y-axis, we plot the total number of infectives in $\mathbf{x}^*$.

Since $\frac{\lambda}{\mu} = 0.5$, the exogenous healing rate μ is larger than the exogenous infection rate λ. When γ is low, the effect of the topology independent process (λ, μ) dominates the topology dependent process (γ), so the most probable network state we expect is $\mathbf{x}^0 = [0, 0, \ldots, 0]$. But at high γ values, the effect of the topology independent process (λ, μ) is dominated by the effect of the topology dependent process (γ) so the solution to (5) is $\mathbf{x}^N = [1, 1, \ldots, 1]$.

Note that for a narrow range of γ values, the number of infectives in $\mathbf{x}^*$ is neither 0 nor 16. In the Erdős-Rényi graph A, for γ approximately between 1.6 and 2, it is 9. In the Watts-Strogatz graph B, for γ between 1.3 and 1.5, it can be 14 or 15, but not 16. These correspond to the spurious most probable network states. Figure 4a shows the actual spurious $\mathbf{x}^*$ for graph A where 9 agents are infected but 7 agents remain healthy. Figure 4b and Figure 4c show these spurious configurations for graph B.

In [6], we only made observations regarding these spurious configurations. Now we are are ready to answer the following questions regarding these spurious solutions to (5):

1. What kind of network topologies exhibit these spurious configurations?
2. What model parameters (λ, γ, μ) will result in these spurious configurations?

We will show that the existence of these spurious configurations is related to subgraphs within the network and to the concept of

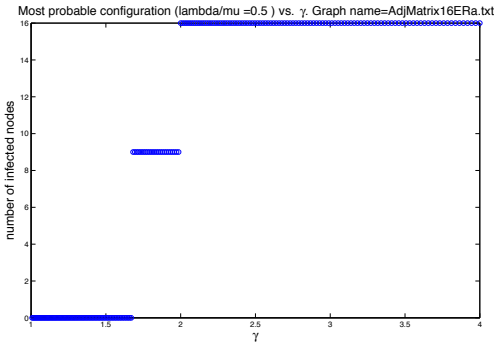


Fig. 2: Most Probable Configuration $\mathbf{x}^*$ for Graph A

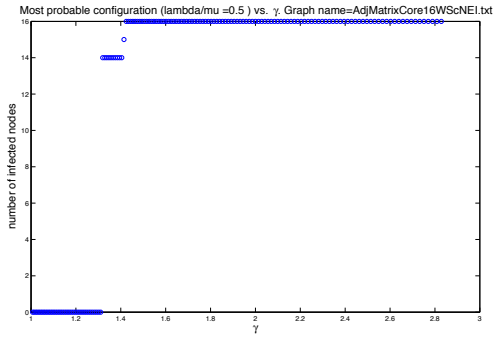


Fig. 3: Most Probable Configuration $\mathbf{x}^*$ for Graph B

graph density.

4. NETWORK CONFIGURATION INDUCED SUBGRAPH

The graph H is an induced subgraph of G if: 1) two vertices in H are connected if and only if they are connected in G [9]; and 2) the vertex set and edge set of H are subsets of the vertex set and edge set of G.

$$V(H) \subseteq V(G), E(H) \subseteq V(G)$$

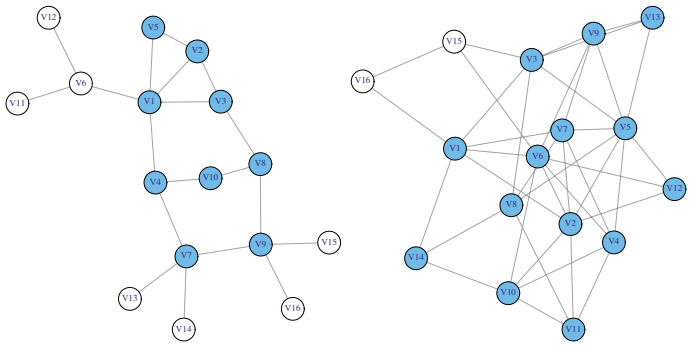
In this paper, we say that H is an induced subgraph of the network state $\mathbf{x}$ if the nodes in the subgraph H are the infected nodes in $\mathbf{x}$ and the edges of H are edges where both end nodes are infected.

$$\begin{aligned} V(H(\mathbf{x})) &= \{v_i \in V(G) \mid x_i = 1\} \\ E(H(\mathbf{x})) &= \{(i,j) \in E(G) \mid x_i = 1, x_j = 1\}. \end{aligned} \tag{6}$$

For example, The entire network G is its own subgraph; it is the subgraph induced by the configuration $\mathbf{x}^N$. The empty graph is the subgraph induced by the configuration $\mathbf{x}^0$. Figures 5 and 6 show two network configurations and their corresponding induced subgraphs.

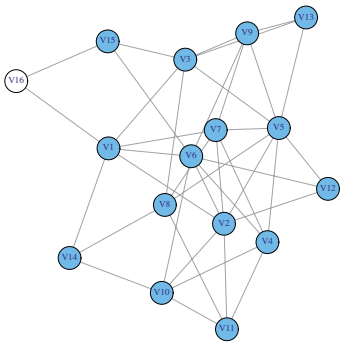
Theorem 4.1. *If the induced subgraphs of two network configurations, $\mathbf{x}_1$ and $\mathbf{x}_2$, are isomorphic, then $\pi(\mathbf{x}_1) = \pi(\mathbf{x}_2)$.*

Proof. By definition, $|V(H(\mathbf{x}))| = 1^T\mathbf{x}$ and $|E(H(\mathbf{x}))| = \frac{\mathbf{x}^T A\mathbf{x}}{2}$. Consider that $\mathbf{x}_1$ and $\mathbf{x}_2$ induce two graphs $H(\mathbf{x}_1)$ and $H(\mathbf{x}_2)$. Since $H(\mathbf{x}_1)$ is isomorphic to $H(\mathbf{x}_2)$, the number of nodes and the number of edges are the same for $H(\mathbf{x}_1)$ and $H(\mathbf{x}_2)$ [9]. By construction, we know that the number of nodes in $H(\mathbf{x}_1) = 1^T\mathbf{x}_1 = 1^T\mathbf{x}_2$, and the number of edges in $H(\mathbf{x}_1) = \frac{\mathbf{x}_1^T A\mathbf{x}_1}{2} = \frac{\mathbf{x}_2^T A\mathbf{x}_2}{2}$. From (3), we can conclude that $\pi(\mathbf{x}_1) = \pi(\mathbf{x}_2)$. □



(a) Spurious $\mathbf{x}^*$ for Graph A with 9 infectives

(b) Spurious $\mathbf{x}^*$ for Graph B with 14 infectives



(c) Spurious $\mathbf{x}^*$ for Graph B with 15 infectives

Fig. 4: Spurious $\mathbf{x}^*$ (Grey = Infectives, White = Susceptibles)

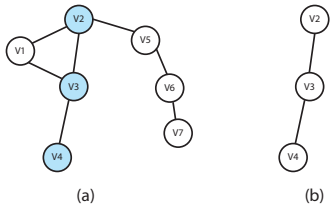


Fig. 5: (a) configuration $\mathbf{x}_1 = [0,1,1,1,0,0,0]^T$; (b) induced subgraph $H(\mathbf{x}_1)$

It's important to note that Theorem 4.1 is only applicable to the scaled SIS model. For the more commonly used additive SIS model, we proved in [8] that two network configurations, $\mathbf{x}_1$ and $\mathbf{x}_2$ are equally probable at equilibrium if their corresponding induced colored graphs are isomorphic; this is a much stricter condition to satisfy. Consider the network configurations $\mathbf{x}_1$ and $\mathbf{x}_2$ as shown in Figures 5 and 6. The overall colored graphs are not isomorphic whereas the induced subgraphs $H(\mathbf{x}_1)$ and $H(\mathbf{x}_2)$ are; so these two configurations are equally probable at equilibrium only in the scaled SIS model and not in the more common additive SIS model.

In the next section, we discuss how the existence of spurious configurations is related to the density of these induced subgraphs.

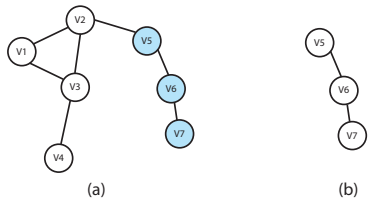


Fig. 6: (a) configuration $\mathbf{x}_2 = [0,0,0,0,1,1,1]^T$; (b) induced subgraph $H(\mathbf{x}_2)$

5. GRAPH DENSITY

The density of graph G is defined as [10]

$$d(G) = \frac{|E(G)|}{|V(G)|}.$$

We will refer to the density of the entire graph G as the *network density*, and the density of the subgraphs of G as the *subgraph density*. By definition, the density of the empty graph, $H(\mathbf{x}^0)$, is 0.

In regime II) **Endogenous Infection Dominant:** where $0 < \frac{\lambda}{\mu} \leq 1, \gamma > 1$, the existence of spurious configuration(s) as solution(s) to (5) is related to the existence of subgraphs whose density is *larger* than the overall network density $d(G)$. We omit the proof here due to space constraints and simply state the following result:

Theorem 5.1. *When $0 < \frac{\lambda}{\mu} \leq 1, \gamma > 1$, if there exists at least one subgraph H in G with density $d(H)$ for which*

$$0 < \frac{\log\left(\frac{\lambda}{\mu}\gamma^{d(G)}\right)}{\log\left(\frac{\lambda}{\mu}\gamma^{d(H)}\right)} < \frac{N'}{N}, \tag{7}$$

then $\mathbf{x}^* = \mathbf{x}' \in \mathcal{X}, \mathbf{x}^* \neq \mathbf{x}^0$, and $\mathbf{x}^* \neq \mathbf{x}^N$.

The implication of Theorem 5.1 is that in, regime II, subgraphs which are denser than the overall graph may also be solutions to (5). We know from Theorem 4.1 that it is possible that these solutions are not unique since multiple dense subgraphs may be isomorphic to each other.

We will illustrate Theorem 5.1 using Graph A, shown in Figure 1a. We know from prior numerical calculation what the spurious configuration is (see Figure 4a), so we will use Figure 2 to show the validity of theorem 5.1.

For Graph A, $N = 16$, $E = 19$, so the network density is $\frac{19}{16}$. There is at least one denser induced subgraph in Graph A (i.e., the subgraph containing only infectives) as shown in Figure 4a. This subgraph has 9 nodes and 12 edges, so its density is $\frac{12}{9}$. We know that $\frac{\lambda}{\mu} = 0.5$ and the density of each subgraph; solving

$$\frac{\log\left(0.5\gamma^{\frac{19}{16}}\right)}{\log\left(0.5\gamma^{\frac{12}{9}}\right)} > 0, \tag{8}$$

we see that the range of γ values for which this is true is $(2^{\frac{19}{16}}, \infty)$. Since this is a lower bound, we can see from figure 2 that when γ is slightly above $2^{\frac{19}{16}} = 1.79$, the most probable configuration is the spurious configuration.

Solving for the RHS of (7),

$$\frac{\log\left(0.5\gamma^{\frac{19}{16}}\right)}{\log\left(0.5\gamma^{\frac{12}{9}}\right)} > \frac{9}{16}, \tag{9}$$

the range of range of γ values for which this is true is $(0.5^{(-1)}, \infty)$. Since this is an upper bound, we can see in Figure 2 that when γ is slightly below $0.5^{(-1)} = 2$, the most probable configuration is a spurious configuration.

Intuitively, this shows that if the topology dependent parameter γ is particularly large or small with respect to a constant $\frac{\lambda}{\mu}$, then the subgraph structures in the network topology may not matter. Otherwise, the solutions to the (5) will be dependent on the subgraph structures in the network topology.

Finding these thresholds require that we find the dense subgraphs in the network topology. In graph theory, there are two major problems related to dense subgraphs: 1) densest subgraph problem, 2) k-densest subgraph problem. The *densest subgraph problem* is the problem of finding the subgraph with the maximum density with no constraint on the number of nodes in the subgraph. It is known that this problem can be solved in polynomial time exactly and in linear time in approximation for undirected graphs. The *k-densest subgraph problem* is the problem of finding the subgraph with maximum density containing exactly k nodes. The k-densest subgraph problem (DkS) is known to be NP-hard [10]. Knowing the solutions to the k-densest subgraph problem will give us all the spurious configurations in a given topology. We will show with rigor in future work that just knowing the densest subgraph will give us an idea of if we should worry about subgraph structures or not. For instance, if we treat the epidemics as a design problem, then we can bias the model parameters in such a way that we can ensure that the solution to (5) can only be $\mathbf{x}^0$ or $\mathbf{x}^N$ for any network topology.

6. CONCLUSION

In this paper, we have shown that, for the scaled SIS epidemics $X(t)$, we can interpret the network states of the epidemics as induced subgraphs; therefore, induced subgraphs that are isomorphic are equally probable in equilibrium. This is a less strict condition than the one we derived for the additive SIS epidemics in [8]. Second, we showed that in regime II) **Endogenous Infection Dominant:** where $0 < \frac{\lambda}{\mu} \leq 1, \gamma > 1$, the spurious configurations (i.e., solutions to (5) other than the non infected $\mathbf{x}^0$ and the whole population infected $\mathbf{x}^N$) correspond to subgraph structures in the network topology that are *denser* than the overall network topology. In future work, we will prove that if we know the density of these subgraphs then we can find the thresholds where $\mathbf{x}^*$ transitions from $\mathbf{x}^0$ to the spurious configuration(s) and from the spurious configuration(s) to $\mathbf{x}^N$.

7. REFERENCES

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

[2] Y. Wang, D. Chakrabarti, C. Wang, and C. Faloutsos, "Epidemic spreading in real networks: an eigenvalue viewpoint," in *2003. 22nd International Symposium on Reliable Distributed Systems*, Florence, Italy, 2003, pp. 25–34.

[3] A. Ganesh, L. Massoulie, and D. Towsley, "The effect of network topology on the spread of epidemics," in *INFOCOM*

2005. 24th Annual Joint Conference of the IEEE Computer and Communications Societies, Miami, USA, Mar. 2005, pp. 1455–1466 vol. 2.

[4] M. Draief, A. Ganesh, and L. Massoulié, "Thresholds for virus spread on networks," in *Proceedings of the 1st International Conference on Performance Evaluation Methodolgies and Tools*. New York, NY, USA: ACM, 2006.

[5] J. Zhang and J. M. F. Moura, "Threshold behavior of epidemics in regular networks," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Vancouver, Canada, May 2013.

[6] ——, "Diffusion in social networks as SIS epidemics: beyond full Mixing and complete graphs," submitted for publication, 2013.

[7] F. P. Kelly, *Reversibility and Stochastic Networks*. Cambridge University Press, 2011.

[8] J. Zhang and J. M. F. Moura, "Accounting for topology in spreading contagion in non-complete networks," in *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Kyoto, Japan, Mar. 2012, pp. 2681–2684.

[9] G. R. C. Godsil, *Algebraic Graph Theory*. Springer-Verlag, 2001.

[10] S. Khuller and B. Saha, "On finding dense subgraphs," in *Automata, Languages and Programming*. Springer, 2009, pp. 597–608.