FINDING UNIQUE DENSE COMMUNITIES

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

Finding densely connected subgraphs, also called communities, in networks are of interest for many applications. In previous work, we showed an optimization method for efficiently finding subgraphs denser than the overall network [1]. This result is derived from our studies of network processes, dynamical processes that model interactions between individual agents in networks (i.e., spread of infection or cascading failures). In this paper, we prove that these subgraphs are also unique in the sense that there are no other subgraphs in the network isomorphic to these subgraphs.

Index Terms— network science, scaled SIS process, community detection, graph clustering, unique subgraph

1. INTRODUCTION

Network science studies networks, which describe relationships or interactions between multiple agents or components in a system [2, 3]. It is of particular interest to find subsets of nodes (i.e., agents) sharing some interesting properties. This is the problem of graph clustering [4, 5, 6]. Often the clusters of interests are communities where agents are highly connected. In this case, community detection is related to the problem of extracting dense subgraphs [7, 8]. For example, in epidemiology, densely connected communities are more vulnerable to the spread of infection through contagion. Alternatively, dense subgraphs are of interest for distributed computation in large, interconnected systems such as the power grid [9].

Previously in [1, 10], we studied a network process model called the scaled SIS (susceptible-infected-susceptible) process. The scaled SIS process models stochastic infection (i.e., failing) and healing (i.e., recovery) of agents whose interactions are described by an undirected, unweighted network, $G(V, E)$. The advantage of the scaled SIS process over other network process models is its equilibrium distribution can be described in closed-form without resorting to mean-field approximation [11] or assuming specific network topologies [12, 13].

We showed through analysis that for a range of dynamical parameters, the configurations with the highest equilibrium probability induce subgraphs that are denser than the overall network. These dense subgraphs can be found efficiently using Max-Flow/Min-Cut algorithm [14]. Solving for the most-probable configuration in a network with 4941 node takes 0.1 sec. on a standard desktop. In this paper, we show that these induced subgraphs are also provably *unique*. There are no other subgraphs in the network that are isomorphic to these induced subgraphs.

This work was partially supported by NSF grants CCF1011903 and CCF1513936

Section 2 briefly reviews the scaled SIS process and the equilibrium distribution. Section 3 describes the Most-Probable Configuration Problem, whose solutions correspond to subgraphs denser than the overall graph. Section 4 proves that these subgraphs are also unique. Section 5 concludes the paper.

2. SCALED SIS PROCESS

In [10], we developed and studied the scaled SIS (susceptible-infected-susceptible) process, $\{X(t), t \geq 0\}$, a binary-state, continuous-time Markov process on a finite-size, static, simple, unweighted, undirected, connected graph $G(V, E)$. The topology of the network captures the interactions amongst the N agents. The state of the ith agent, x_i, is either healthy ($x_i = 0$) or infected ($x_i = 1$). The state of the entire network at some time t,

$$X(t) = \mathbf{x} = [x_1, x_2, \ldots, x_N]^T,$$

is the state of all the agents at time t. We call $\mathbf{x}$ the network *configuration*. Due to the interactions, the evolution of an agent state, x_i, is no longer independent of the state of other agents. The scaled SIS process assumes that

1. $X(t)$ transitions to the configuration where 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. $X(t)$ transitions to the configuration where the ith agent ($i = 1, 2, \ldots, N$) is infected with transition rate

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

where $m_i = \sum_{j=1}^{N} \mathbb{1}(x_j = 1)\mathbf{A}_{ij}$ is the number of infected neighbors of node i. The symbol $\mathbb{1}(\cdot)$ is the indicator function, and $\mathbf{A} = [\mathbf{A}_{ij}]$ 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 when $m_i = 0$, the infection rate is 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 because of exogenous infection and healing. The state space of the scaled SIS process is $\mathcal{X} = \{\mathbf{x}\}$ and the size of the state space is 2^N.

In [10], we proved that the equilibrium distribution of the scaled SIS process is

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

where Z is the partition function and is defined as

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

The term $1^T\mathbf{x}$ is the number of infected agents in configuration $\mathbf{x}$ and $\frac{\mathbf{x}^T\mathbf{A}\mathbf{x}}{2}$ is the number of infected edges (i.e., edges whose end nodes are infected).

2.1. Induced Subgraphs and Graph Density

Alternatively, as we showed in [1], we can consider each configuration, $\mathbf{x}$, from the perspective of induced subgraphs. The graph F is an induced subgraph of the network configuration $\mathbf{x}$ if the nodes in the subgraph F are the infected nodes in $\mathbf{x}$ and the edges of F are edges where both end nodes are infected.

$$\begin{aligned} V(F(\mathbf{x})) &= \{v_i \in V(G) \mid x_i = 1\} \\ E(F(\mathbf{x})) &= \{(i,j) \in E(G) \mid x_i = 1, x_j = 1\}. \end{aligned} \tag{5}$$

Definition The set of all possible induced subgraphs of G is $\mathcal{F} = \{F(\mathbf{x})\}, \forall\, \mathbf{x} \in \mathcal{X}$.

The set $\mathcal{F}$ includes the empty graph, $\emptyset$, which is induced by the configuration $\mathbf{x}^0 = [0, 0, \ldots, 0]^T$, and G, which is the subgraph induced by the configuration $\mathbf{x}^N = [1, 1, \ldots, 1]^T$. The cardinality of $|\mathcal{F}| = |\mathcal{X}|$.

From [7] The density of a graph G is

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

There is an alternative definition for graph density that is the number of edges divided by the total number of possible edges [15]. Unfortunately, these two definitions of density are not equivalent.

We will refer to the density of the entire network, $d(G) = d(F(\mathbf{x}^N))$, as the *network density*, and the density of an induced subgraph of G as the *subgraph density*. The density of the empty graph, $d(F(\mathbf{x}^0))$, is 0 by definition.

Alternatively, the equilibrium distribution of the scaled SIS process can be written as a function of the induced subgraphs

$$\pi(F(\mathbf{x})) = \frac{1}{Z}\left(\left(\frac{\lambda}{\mu}\right)\gamma^{d(F(\mathbf{x}))}\right)^{|V(F(\mathbf{x}))|}, \quad F \in \mathcal{F}, \tag{6}$$

where $d(F(\mathbf{x}))$ is the density of the induced subgraph F and Z is the partition function.

3. MOST-PROBABLE CONFIGURATION PROBLEM

The Most-Probable Configuration Problem solves for the configuration with the highest equilibrium probability. The *most-probable configuration*, $\mathbf{x}^*$, is the network configuration that we would most likely observe at equilibrium:

$$\mathbf{x}^* = \arg\max_{\mathbf{x}\in\mathcal{X}} \pi(\mathbf{x}) = \arg\max_{\mathbf{x}\in\mathcal{X}} \left(\frac{\lambda}{\mu}\right)^{1^T\mathbf{x}} \gamma^{\frac{\mathbf{x}^T A\mathbf{x}}{2}}. \tag{7}$$

The equilibrium distribution of the scaled SIS process (3) is a Gibbs distribution. In the context of Markov random field literatures, the Most-Probable Configuration Problem is the MAP (Maximum A Posteriori) problem [16]. The most-probable configuration is called the ground state problem in statistical mechanics. It is the state of minimum energy [17].

3.1. Regime II) Endogenous Infection Dominant

The solution of (7) depends on the adjacency matrix of the underlying network, $\mathbf{A}$, and the dynamics of the process through $\frac{\lambda}{\mu}$ and γ. The most-probable configuration is particularly interesting in Regime II) **Exogenous Infection Dominant**, where $\frac{\lambda}{\mu} > 1, 0 < \gamma \leq 1$. In this regime, the topology-dependent process (controlled by γ) opposes the effect of the topology-independent process (controlled by λ, μ). We showed in [1, 10] that many of the most-probable configurations are non-degenerate (i.e., $\mathbf{x} \neq \{\mathbf{x}^0, \mathbf{x}^N\}$); in these non-degenerate most-probable configuration, subsets of agents are infected while others are healthy.

Corollary 5.8 in [1] proved that in Regime II) Endogenous Infection Dominant: $0 < \frac{\lambda}{\mu} \leq 1, \gamma > 1$, the subgraphs induced by the non-degenerate most-probable configurations have density larger than the overall graph

$$d(F(\mathbf{x}^*)) > d(G).$$

We also proved in [1] that in Regime II), the solution of the Most-Probable Configuration Problem corresponds to the minimum of a submodular function. We can use Max-Flow/Min-Cut algorithm to efficiently solve this combinatorial optimization problem [14].

Figure 1a shows the most-probable configuration when

$$\frac{\lambda}{\mu} = 0.06, \gamma = 3.53$$

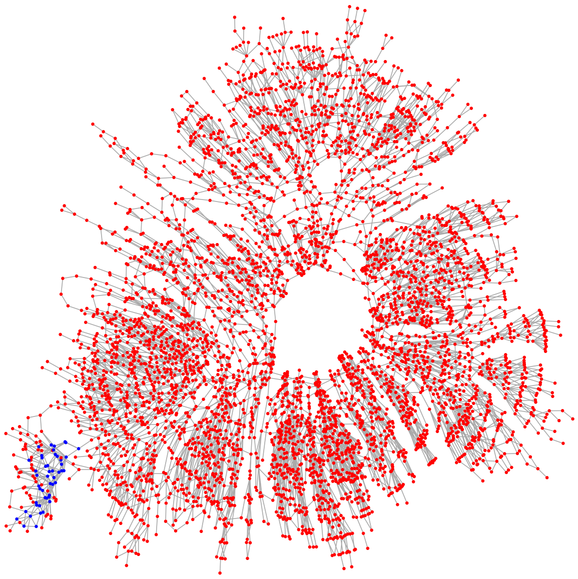
for the US Western Power Grid, a 4941-node network presentation of the western power grid [18]. The network can be obtained from [19]. The most-probable configuration is found in 0.026 sec. on a desktop with 3.7 GHz Quad Core Xeon processor and 16GB of RAM. Figure 1b shows the corresponding induced subgraph. The density of the induced subgraph is $d(F(\mathbf{x}^*)) = 3$. The density of the overall network is $d(G) = 1.335$.

4. UNIQUENESS OF THE MOST-PROBABLE CONFIGURATION

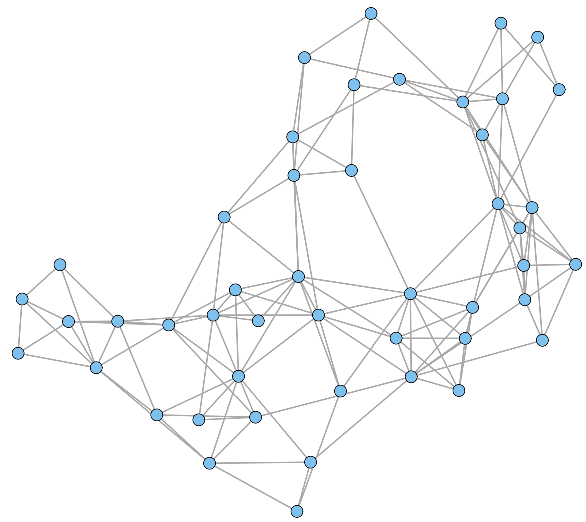
The graph $F(\mathbf{x}^*)$ is the subgraph induced by the solution of the Most-Probable Configuration Problem. In this section, we will prove that $F(\mathbf{x}^*)$ is *subgraph unique* in G in that there are no other subgraphs in G isomorphic to $F(\mathbf{x}^*)$. This does not guarantee that the solution of (7), a combinatorial optimization problem, is unique [20].

Definition [21] Two graphs, G and F, are equivalent if they are isomorphic: there is a bijection between the vertex sets of G and F, $f : V(G) \to V(F)$, such that any two vertices u and v of G are adjacent in G if and only if $f(u)$ and $f(v)$ are adjacent in F. This means that edge (u, v) is in G if and only if edge $(f(u), f(v))$ is in F.

Lemma 4.1. *[Proof in Appendix A] Given an undirected graph G described by adjacency matrix $\mathbf{A}$ and two different configurations $\mathbf{x}_1$ and $\mathbf{x}_2$, which induce subgraphs $F(\mathbf{x}_1)$ and $F(\mathbf{x}_2)$, respectively. $F(\mathbf{x}_1)$ is isomorphic to $F(\mathbf{x}_2)$ if and only if $1^T\mathbf{x}_1 = 1^T\mathbf{x}_2$ and $\mathbf{x}_1^T\mathbf{A}\mathbf{x}_1 = \mathbf{x}_2^T\mathbf{A}\mathbf{x}_2$.*



(a) $\mathbf{x}^*$, Blue = Infected, Red = Healthy



(b) $F(\mathbf{x}^*)$

Fig. 1: Most-Probable Configuration, $\mathbf{x}^*$, for $\left(\frac{\lambda}{\mu} = 0.06, \gamma = 3.53\right)$. Computation time = 0.026 sec.

Theorem 4.1. *[Proof in Appendix B] Consider Regime II)* ***Endogenous Infection Dominant:*** $0 < \frac{\lambda}{\mu} \leq 1, \gamma > 1$ *and network G. The most-probable configuration,* $\mathbf{x}^*$*, induces the subgraph* $F(\mathbf{x}^*)$ *with density* $d(F(\mathbf{x}^*))$*. If*

$$\frac{\lambda}{\mu}\gamma^{d(F(\mathbf{x}^*))} > 1,$$

then $\mathbf{x}^*$ *is subgraph unique.*

Theorem 4.2. *If* $\mathbf{x}^*$ *is the most-probable configuration in Regime II)* ***Endogenous Infection Dominant:*** $0 < \frac{\lambda}{\mu} \leq 1, \gamma > 1$*, then it is subgraph unique.*

Proof. Theorem 4.1 states the condition for which a configuration, $\mathbf{x}^*$, is subgraph unique. Theorem 5.6 in [1] states the following necessary and sufficient conditions: The most-probable configuration $\mathbf{x}^* \neq \mathbf{x}^0$ if and only if there exists at least one induced subgraph $F \in \mathcal{F}$ with density $d(F)$ for which $\frac{\lambda}{\mu}\gamma^{d(F)} > 1$.

Since the induced subgraph of $\mathbf{x}^*$ satisfies the condition that $\frac{\lambda}{\mu}\gamma^{d(F(\mathbf{x}^*))} > 1$, then $\mathbf{x}^* \neq \mathbf{x}^0$. However, since $\mathbf{x}^0$ induces the empty graph, a solution where $\mathbf{x}^* = \mathbf{x}^0$ is also subgraph unique. As a result, any solution to the Most-Probable Configuration Problem in Regime II) is subgraph unique.

□

5. CONCLUSION

Theorem 4.2 states that all the possible solutions of the Most-Probable Configuration Problem in Regime II) are subgraph unique. This means that the corresponding induced subgraphs, $F(\mathbf{x}^*)$, are unique. This result is particularly interesting in the case when $\mathbf{x}^*$ is a non-degenerate configuration since it would be computationally infeasible to iterate through all the possible subgraphs in a network to find unique subgraphs. By studying network processes, we can gain insights into both the behavior of the dynamical process as well as properties of the underlying graph structure.

6. ACKNOWLEDGEMENT

We wish to thank Prof. Cristopher Moore of the Santa Fe Institute for discussions regarding Max-Flow/Min-Cut algorithms.

A. PROOF FOR LEMMA 4.1

Lemma. *Given an undirected graph G described by adjacency matrix* $\mathbf{A}$ *and two different configurations* $\mathbf{x}_1$ *and* $\mathbf{x}_2$*, which induce subgraphs* $F(\mathbf{x}_1)$ *and* $F(\mathbf{x}_2)$*, respectively.* $F(\mathbf{x}_1)$ *is isomorphic to* $F(\mathbf{x}_2)$ *if and only if* $1^T\mathbf{x}_1 = 1^T\mathbf{x}_2$ *and* $\mathbf{x}_1^T\mathbf{A}\mathbf{x}_1 = \mathbf{x}_2^T\mathbf{A}\mathbf{x}_2$*.*

Proof. Recall that $1^T\mathbf{x}_1$ and $1^T\mathbf{x}_2$ are the number of infected nodes in the two configurations. Therefore, they are the number of nodes in each induced subgraph, $F(\mathbf{x}_1)$ and $F(\mathbf{x}_2)$. By the definition of the induced subgraphs, $\mathbf{x}_1^T\mathbf{A}\mathbf{x}_1$ and $\mathbf{x}_2^T\mathbf{A}\mathbf{x}_2$ are equal to twice the number of edges in each induced subgraph.

Necessity

If $F(\mathbf{x}_1)$ is isomorphic to $F(\mathbf{x}_2)$, then $1^T\mathbf{x}_1 = 1^T\mathbf{x}_2$ and $\mathbf{x}_1^T\mathbf{A}\mathbf{x}_1 = \mathbf{x}_2^T\mathbf{A}\mathbf{x}_2$.

This follows from the definition of isomorphism.

Sufficiency

If $1^T\mathbf{x}_1 = 1^T\mathbf{x}_2$ and $\mathbf{x}_1^T\mathbf{A}\mathbf{x}_1 = \mathbf{x}_2^T\mathbf{x}_2$, then $F(\mathbf{x}_1)$ is isomorphic to $F(\mathbf{x}_2)$.

We prove this by contrapositive. We need to prove that if $F(\mathbf{x}_1)$ is not isomorphic to $F(\mathbf{x}_2)$, then $1^T\mathbf{x}_1 \neq 1^T\mathbf{x}_2$ or $\mathbf{x}_1^T\mathbf{A}\mathbf{x}_1 \neq \mathbf{x}_2^T\mathbf{A}\mathbf{x}_2$.

There are two ways $F(\mathbf{x}_1)$ is not isomorphic to $F(\mathbf{x}_2)$: 1) There is no bijective function f or 2) There is a bijective function f but two vertices adjacent in $F(\mathbf{x}_1)$ are not adjacent in $F(\mathbf{x}_2)$. The bijection function f does not exist if the induced subgraphs, $F(\mathbf{x}_1)$ and $F(\mathbf{x}_2)$, have different number of nodes; this mean that $1^T\mathbf{x}_1 \neq 1^T\mathbf{x}_2$. 2) There is a bijection function, f, but nodes u and v are adjacent in

$F(\mathbf{x}_1)$ but are not adjacent in $F(\mathbf{x}_2)$. This would mean that $\mathbf{x}_1^T \mathbf{A} \mathbf{x}_1 \neq \mathbf{x}_2^T \mathbf{A} \mathbf{x}_2$. This is impossible by the definition of induced subgraph.

□

B. PROOF FOR THEOREM 4.1

Theorem. *Consider Regime II)* ***Endogenous Infection Dominant:*** $0 < \frac{\lambda}{\mu} \leq 1, \gamma > 1$ *and network* G. *The most-probable configuration,* $\mathbf{x}^*$, *induces the subgraph* $F(\mathbf{x}^*)$ *with density* $d(F(\mathbf{x}^*))$. *If*

$$\frac{\lambda}{\mu}\gamma^{d(F(\mathbf{x}^*))} > 1,$$

then $\mathbf{x}^*$ *is subgraph unique.*

Proof. Due to space constraint, we present a summarized version of the proof. The detailed proof is found in [20].

We will prove by contradiction. We can show that when the solutions to the Most-Probable Configuration Problem are two equally probable configurations, $\mathbf{x}_1^*$ and $\mathbf{x}_2^*$, whose induced subgraphs, $F_1 = F(\mathbf{x}_1^*)$ and $F_2 = F(\mathbf{x}_2^*)$, are isomorphic, we can always create another subgraph by combining F_1 and F_2 such that the configuration that induces this third subgraph will have a higher probability than $\mathbf{x}_1^*$ and $\mathbf{x}_2^*$. Let $F_1 \cap F_2$ denote $V(F_1) \cap V(F_2)$ and $F_1 \cup F_2$ denote $V(F_1) \cup V(F_2)$.

Suppose that the solutions to the Most-Probable Configuration Problem are two equally probable configurations, $\mathbf{x}_1^*$ and $\mathbf{x}_2^*$, whose induced subgraphs, $F_1 = F(\mathbf{x}_1^*)$ and $F_2 = F(\mathbf{x}_2^*)$, are isomorphic. From Lemma 4.1, this means that $|V(F_1)| = |V(F_2)| = N_1$, $|E(F_1)| = |E(F_2)| = E_1$, $d(F_1) = d(F_2)$. Additionally, we know that $\frac{\lambda}{\mu}\gamma^{d(F_1)} = \frac{\lambda}{\mu}\gamma^{d(F_2)} > 1$.

We now consider two cases: 1) $F_1 \cap F_2 = \emptyset$ and 2) $F_1 \cap F_2 \neq \emptyset$.

$F_1 \cap F_2 = \emptyset$

The induced subgraphs F_1 and F_2 are disconnected.

Define a new subgraph

$$\widetilde{F} = F_1 \cup F_2.$$

We know that $|V(\widetilde{F})| = 2N_1$ and $|E(\widetilde{F})| = 2E_1$, while $d(\widetilde{F}) = d(F_1) = d(F_2)$. The subgraph $\widetilde{F}$ has the same density as F_1, which means

$$\frac{\lambda}{\mu}\gamma^{d(\widetilde{F})} = \frac{\lambda}{\mu}\gamma^{d(F_1)} > 1.$$

Additionally $\widetilde{N} > N_1$. As we are in Regime II), by (6), the configuration that induces $\widetilde{F}$ has a larger equilibrium probability than $\mathbf{x}_1^*, \mathbf{x}_2^*$, thereby contradicting the premise that $\mathbf{x}_1^*, \mathbf{x}_2^*$ are the most-probable configurations.

$F_1 \cap F_2 \neq \emptyset$

The induced subgraphs F_1 and F_2 are not disconnected.

Define a new subgraph

$$\hat{F} = F_1 \cap F_2. \tag{8}$$

We know then that $|V(\hat{F})| = \hat{N} < N_1$, $|E(\hat{F})| = \hat{E} < E_1$, with density $d(\hat{F}) = \frac{\hat{E}}{\hat{N}}$. We have 3 cases to consider: 1) $d(\hat{F}) = d(F_1)$, 2) $d(\hat{F}) < d(F_1)$, 3) $d(\hat{F}) > d(F_1)$.

In case 1) $d(\hat{F}) = d(F_1)$, we can show that the new subgraph

$$\widetilde{F} = F_1 \cup F_2$$

has the same density as F_1. With $\widetilde{N} > N_1$, the configuration that induces $\widetilde{F}$ is more probable than $\mathbf{x}_1^*$, leading to a contradiction.

In case 2) $d(\hat{F}) < d(F_1)$, we can show that the new subgraph

$$\widetilde{F} = F_1 \cup F_2$$

is denser than F_1. With $\widetilde{N} > N_1$, the configuration that induces $\widetilde{F}$ is more probable than $\mathbf{x}_1^*$, leading to a contradiction.

In case 3) $d(\hat{F}) > d(F_1)$, we need to consider both

$$\widetilde{F} = F_1 \cup F_2$$

and

$$\hat{F} = F_1 \cap F_2.$$

We can show that the equilibrium probability of $\mathbf{x}_1^*, \mathbf{x}_2^*$ can not simultaneously be larger than the equilibrium probability of the configuration, $\hat{\mathbf{x}}$, which induces the subgraph $\hat{F}$ *and* the equilibrium probability of the configuration, $\widetilde{\mathbf{x}}$, which induces the subgraph $\widetilde{F}$. This contradicts the premise that $\mathbf{x}_1^*$ and $\mathbf{x}_2^*$ are the two most-probable configurations.

□

C. REFERENCES

[1] J. Zhang and J. M. F. Moura, "Role of subgraphs in epidemics over finite-size networks under the scaled SIS process," *Journal of Complex Networks*, 2015.

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

[3] M. Newman, *Networks: an Introduction*. Oxford University Press, 2010.

[4] S. E. Schaeffer, "Graph clustering," *Computer Science Review*, vol. 1, no. 1, pp. 27–64, 2007.

[5] S. Fortunato, "Community detection in graphs," *Physics Reports*, vol. 486, no. 3, pp. 75–174, 2010.

[6] J. Leskovec, K. J. Lang, and M. Mahoney, "Empirical comparison of algorithms for network community detection," in *Proceedings of the 19th international conference on World wide web*. ACM, 2010, pp. 631–640.

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

[8] V. E. Lee, N. Ruan, R. Jin, and C. Aggarwal, "A survey of algorithms for dense subgraph discovery," in *Managing and Mining Graph Data*. Springer, 2010, pp. 303–336.

[9] J. Mohammadi, J. Zhang, S. Kar, G. Hug, and J. M. F. Moura, "Multilevel distributed approach for DC optimal power flow," in *Proceedings of IEEE Global Conference on Signal and Information Processing (GlobalSIP)*. IEEE, to be published 2015.

[10] J. Zhang and J. M. F. Moura, "Diffusion in social networks as SIS epidemics: Beyond full mixing and complete graphs," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 4, pp. 537–551, Aug 2014.

[11] A. Barrat, M. Barthelemy, and A. Vespignani, *Dynamical Processes on Complex Networks*. Cambridge University Press, 2008, vol. 1.

[12] P. Van Mieghem and E. Cator, "Epidemics in networks with nodal self-infection and the epidemic threshold," *Phys. Rev. E*, vol. 86, p. 016116, Jul 2012. [Online]. Available: http://link.aps.org/doi/10.1103/PhysRevE.86.016116

[13] R. Durrett, "Some features of the spread of epidemics and information on a random graph," *Proceedings of the National Academy of Sciences*, vol. 107, no. 10, pp. 4491–4498, 2010. [Online]. Available: http://www.pnas.org/content/107/10/4491.abstract

[14] V. Kolmogorov and R. Zabin, "What energy functions can be minimized via graph cuts?" *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 2, pp. 147–159, 2004.

[15] S. Wasserman, *Social Network Analysis: Methods and Applications*. Cambridge University Press, 1994, vol. 8.

[16] D. Koller and N. Friedman, *Probabilistic Graphical Models: Principles and Techniques*. MIT press, 2009.

[17] A. K. Hartmann and M. Weigt, *Phase Transitions in Combinatorial Optimization Problems: Basics, Algorithms and Statistical Mechanics*. John Wiley & Sons, 2006.

[18] D. J. Watts and S. H. Strogatz, "Collective dynamics of small-world networks," *Nature*, vol. 393, no. 6684, pp. 440–442, 1998.

[19] J. Leskovec and A. Krevl, "SNAP Datasets: Stanford large network dataset collection," http://snap.stanford.edu/data, Jun. 2014.

[20] J. Zhang, "Network process: How topology impacts the dynamics of epidemics and cascading failures," Ph.D. dissertation, Carnegie Mellon University, 2015.

[21] D. B. West *et al.*, *Introduction to Graph Theory*. Prentice hall Upper Saddle River, 2001, vol. 2.