Uncertainty Reduction for Stochastic Processes on Complex Networks

Filippo Radicchi and Claudio Castellano

Uncertainty reduction for stochastic processes on complex networks

Filippo Radicchi[1] and Claudio Castellano[2]

[1]*Center for Complex Networks and Systems Research, School of Informatics,*
Computing, and Engineering, Indiana University, Bloomington, Indiana 47408, USA[*]
[2]*Istituto dei Sistemi Complessi (ISC-CNR), Via dei Taurini 19, 00185 Roma, Italy*

Many real-world systems are characterized by stochastic dynamical rules where a complex network of interactions among individual elements probabilistically determines their state. Even with full knowledge of the network structure and of the stochastic rules, the ability to predict system configurations is generally characterized by large uncertainty. Selecting a fraction of the nodes and observing their state may help to reduce the uncertainty about the unobserved nodes. However, choosing these points of observation in an optimal way is a highly nontrivial task, depending on the nature of the stochastic process and on the structure of the underlying interaction pattern. In this paper, we introduce a computationally efficient algorithm to determine quasi-optimal solutions to the problem. The method leverages network sparsity to reduce computational complexity from exponential to almost quadratic, thus allowing the straightforward application of the method to mid-to-large-size systems. Although the method is exact only for equilibrium stochastic processes defined on trees, it turns out to be effective also for out-of-equilibrium processes on sparse loopy networks.

Stochastic phenomena are studied in any field of science, including biology [1], ecology [2], physics [3], neuroscience [4], and finance [5]. In a stochastic system composed of multiple elements, the states of the elements obey probabilistic rules that depend on the states of other elements. Often, a sparse network describes how elements interact one with the other [6]. Consider flu spreading for example. The epidemics starts from a few initial seeds. A person not immunized can contract the disease with a certain probability only if in contact with an infected individual. At the same time, infected people can spontaneously recover. The social network underlying the spreading process determines how the state of every individual depends on the others. At any given time, the system is characterized by some uncertainty, in the sense that different configurations have a non-vanishing probability to appear. Such an uncertainty is due to the stochasticity of the process, and it is present regardless of the knowledge possessed about the probabilistic dynamical rules and about the contact pattern.

To reduce uncertainty, one can observe the state of a sample of elements. In the example of flu spreading, this means obtaining full knowledge about the health state of some people. With such a knowledge, the prediction of the state of unobserved elements becomes less uncertain. In particular, the larger the sample, the lower the uncertainty, with the limiting case of null uncertainty when the entire system is observed. Resource constraints make complete observation usually impossible. Is there an efficient way of identifying the best elements to observe so that the uncertainty for the rest of system is minimized? The question is answered, from an information-theoretic point of view, by the principle of maximum entropy sampling (MES) [7]. Its rationale is intuitive: to reduce uncertainty about the system as much as possible, the elements for which joint uncertainty is maximal must be observed. MES is often used as a solution to problems of experimental design [8]. An example is the problem of where to place thermometers in a room to provide the most accurate picture of the temperature in the entire room [9]. In spe-

cial settings, MES can be efficiently approximated or achieved exactly with ad-hoc algorithms [10–12], These studies have, however, considered very small systems because the computational complexity of the proposed algorithms grows exponentially. Further, the problem has been studied only in regular topologies, such as lattices or fully connected networks. The present paper considers the MES problem when the interaction pattern is given by a large complex network. In this case, the sparsity of the topology can be leveraged to make the application of MES feasible in rather large systems.

To avoid any potential confusion, we stress that our goal is the selection of a fraction of observed nodes in order to minimize the uncertainty on the stochastic variables associated with unobserved nodes. This is distinct from the problem of optimally sampling a network to reduce uncertainty on its unknown properties (e.g., degree distribution, diameter, size) [13, 14]. Also, our problem is different from active learning in networks [15–17], where the goal is to infer a model able to predict the value of the unobserved variables in a specific configuration of the system. In our case, we do not infer parameter models. Further, we are not interested in making predictions about a specific configuration. Instead, for all possible configurations that the system may exhibit, we want to identify what nodes we need to observe in order to minimize our uncertainty about such configurations. In this respect, our problem is similar to the one studied in Ref. [18], with the difference that we deal with stochastic rather than deterministic systems.

We consider a dynamical stochastic process defined on a graph $\mathbb{G}$, composed of N nodes. Every node $i \in \mathbb{G}$ is characterized by a state variable x_i that can assume K distinct values; $\mathbf{x} = (x_1, x_2, \ldots, x_N)$ corresponds to a specific microscopic configuration of the system. We assume that the process is Markovian and that the change of the state of a single node is determined only by local interactions with the nodes directly connected to it. Hence, the graph $\mathbb{G}$ fully determines how microscopic configurations are related one to the other. Let us indicate with $p(\mathbf{x})$ the stationary probability distribution

associated to each of the K^N possible microscopic configurations that the system can assume. Despite full knowledge of the graph structure and of the stochastic process, we are still left with potentially large uncertainty quantified by the information-theoretic joint entropy

$$\mathcal{H}(\mathbb{G}) = -\sum_{\mathbf{x}} p(\mathbf{x}) \log_2[p(\mathbf{x})] . \tag{1}$$

Suppose we can observe a subset of nodes $\mathbb{O} \subseteq \mathbb{G}$. Observing these nodes removes any uncertainty on their state, and thus conditions the joint probability distribution of the unobserved part of the graph, $\mathbb{G} \setminus \mathbb{O}$, to the state of the observed nodes $\mathbb{O}$, namely $p(x_{u_1}, \dots, x_{u_{N-O}} | x_{o_1}, \dots, x_{o_O}) = p(\mathbf{x}_{\mathbb{G} \setminus \mathbb{O}} | \mathbf{x}_{\mathbb{O}})$, where $u_1, \dots, u_{N-O} \in \mathbb{G} \setminus \mathbb{O}$, $o_1, \dots, o_O \in \mathbb{O}$, and we defined $\mathbf{x}_{\mathbb{G} \setminus \mathbb{O}} = (x_{u_1}, \dots, x_{u_{N-O}})$ and $\mathbf{x}_{\mathbb{O}} = (x_{o_1}, \dots, x_{o_O})$. For a particular choice of the set $\mathbb{O}$, the uncertainty about the rest of the system is quantified by the conditional entropy

$$\mathcal{H}(\mathbb{G} \setminus \mathbb{O} | \mathbb{O}) = -\sum_{\mathbf{x}_{\mathbb{O}}} p(\mathbf{x}_{\mathbb{O}}) \sum_{\mathbf{x}_{\mathbb{G} \setminus \mathbb{O}}} p(\mathbf{x}_{\mathbb{G} \setminus \mathbb{O}} | \mathbf{x}_{\mathbb{O}}) \log_2[p(\mathbf{x}_{\mathbb{G} \setminus \mathbb{O}} | \mathbf{x}_{\mathbb{O}})] . \tag{2}$$

For $\mathbb{O} = \emptyset$, Eq. (2) is identical to Eq. (1). For $\mathbb{O} = \mathbb{G}$, we have instead $\mathcal{H}(\mathbb{G} \setminus \mathbb{O} | \mathbb{O}) = \mathcal{H}(\emptyset) = 0$.

We look for the optimal selection of a number O of nodes such that their observation minimizes the conditional entropy of Eq. (2). In particular, since $\mathcal{H}(\mathbb{G} \setminus \mathbb{O} | \mathbb{O}) = \mathcal{H}(\mathbb{G}) - \mathcal{H}(\mathbb{O})$, the minimization of Eq. (2) is equivalent to finding the group of nodes $\mathbb{O}^*$ having maximum joint entropy, i.e.,

$$\mathbb{O}^* = \arg \max_{\mathbb{O}} \mathcal{H}(\mathbb{O}) , \tag{3}$$

where $\mathcal{H}(\mathbb{O}) = -\sum_{\mathbf{x}_{\mathbb{O}}} p(\mathbf{x}_{\mathbb{O}}) \log_2[p(\mathbf{x}_{\mathbb{O}})]$. The maximization is performed over all sets $\mathbb{O}$ of fixed size O. This principle is known as MES, and the associated problem is *NP*-hard [7]. The exact solution of this optimization requires the consideration of all possible choices of the set $\mathbb{O}$, and for each of them the computation of the associated joint entropy. The computational complexity of both operations scales exponentially with O.

A quasi-optimal solution can be obtained at a reduced computational cost, exploiting the sub-modularity of the entropy function [19]. Such a property allows us to implement a greedy strategy, where the set of observed nodes is built sequentially, leading to a solution provably close to the optimum [20]. The greedy strategy provides a solution corresponding to a value of the function to be optimized that is at least $(1 - 1/e) = 0.63\dots$ times the value of the global maximum [20]. In the present context, the greedy algorithm consists in sequentially adding, to the set of observed nodes, the node with maximal entropy conditioned to the set of variables already observed. More specifically, the algorithm starts at stage $t = 0$ with an empty set, $\mathbb{O}_{t=0} = \emptyset$. The t-th point of observation, namely o_t, is chosen, among the nodes not yet part of the observed set $\mathbb{O}_{t-1} = \{o_1, o_2, \dots, o_{t-1}\}$, according to the rule

$$o_t = \arg \max_{i \notin \mathbb{O}_{t-1}} \mathcal{H}(i | o_1, \dots, o_{t-1}) . \tag{4}$$

The algorithm can be run up to arbitrary values $1 \le t \le N$. This procedure addresses the issue of the extensive search over all possible groups of nodes. However, at every stage t, the computation of each of the $N - (t - 1)$ conditional entropies in Eq. (4) still requires a number of operations scaling as K^t. This makes the algorithm usable only for constructing very small sets of observed nodes.

To make the greedy algorithm applicable to large sets, one must introduce approximations to reduce the computational complexity of the calculation of $\mathcal{H}$ in Eq. (4). The simplest and most popular ansatz in the study of processes on networks is the so-called individual-based mean-field (IBMF) approximation [21], according to which the joint distribution $p(\mathbf{x})$ is seen as the product of the marginal probabilities of the individual nodes, i.e., $p(x_1, x_2, \dots, x_N) = p(x_1)p(x_2)\cdots p(x_N)$. The approximation allows us to write the entropy of any set $\mathbb{O}$ of variables as

$$\mathcal{H}_{ind}(\mathbb{O}) = \sum_{o \in \mathbb{O}} \mathcal{H}(o) , \tag{5}$$

where $\mathcal{H}(o)$ is the unconditional entropy of the node o. Under this approximation, the optimal set $\mathbb{O}_{ind}$ corresponds to the O nodes with maximal unconditional entropy.

In this paper, we propose a refined approximation, much less drastic than the IBMF approach, based on two assumptions. First, we assume that the graph $\mathbb{G}$ fully determines dependencies among variables. If the pair of nodes i and j is connected by an edge, then the variables x_i and x_j are directly dependent one on the other. Otherwise, the variables still depend one on the other but only through at least another variable in the system. This seems a reasonable way of improving the IBMF approximation, as we expect that the most important dependencies are present among pairs of variables with a direct interaction. This assumption is exact for equilibrium configurations of processes with rates depending on the states of direct neighbors and satisfying detailed balance [22]. Second, we assume that the graph $\mathbb{G}$ is a tree. Both assumptions are used in our proposed algorithm, that allows us to efficiently compute the entropy, namely $\mathcal{H}_{pair}(\mathbb{O})$, for an arbitrary subset of variables $\mathbb{O}$ [and, as a consequence, the conditional entropies in Eq. (4)]. If the set $\mathbb{O}$ coincides with the entire graph, then the algorithm is equivalent to the one used to compute the Bethe free-entropy on trees [23]. The algorithm works sequentially, in the sense that the function $\mathcal{H}_{pair}(\mathbb{O})$ is computed by iteratively adding single nodes to the set $\mathbb{O}$. This allows us to use the algorithm directly in the greedy maximization of Eq. (4).

Properties of the entropy function alone allow us to write the inequality

$$\mathcal{H}(\mathbb{O}) \le \mathcal{H}_{pair}(\mathbb{O}) \le \mathcal{H}_{ind}(\mathbb{O}) \tag{6}$$

for any $\mathbb{O} \subseteq \mathbb{G}$. Essentially, our approximation always leads to an upper-bound of the true entropy function that is tighter than the one predicted using the standard IBMF approximation. The approximation is exact, i.e., $\mathcal{H}(\mathbb{O}) \equiv \mathcal{H}_{pair}(\mathbb{O})$, only

in the case of equilibrium distributions of systems satisfying detailed balance on a tree.

Suppose we are at stage t of the algorithm. We need to compute the conditional entropy $\mathcal{H}_{pair}(o_t|o_1,\ldots,o_{t-1})$ for the next node o_t that we are adding to the set. Thanks to Bayes rule, we can write

$$\mathcal{H}_{pair}(o_t|o_1,\ldots,o_{t-1}) = \mathcal{H}(o_t) + \mathcal{H}_{pair}(o_1,\ldots,o_{t-1}|o_t) \\ -\mathcal{H}_{pair}(o_1,\ldots,o_{t-1}) \quad . \tag{7}$$

Under our two main assumptions, the second term can be written [24], as the sum of pairwise conditional entropies, one per observed node

$$\mathcal{H}_{pair}(o_1,\ldots,o_{t-1}|o_t) = \sum_{j=1}^{t-1} \mathcal{H}(o_j|s_{o_j}). \tag{8}$$

Every observed node corresponds to a term in the sum, given by the entropy associated with that observed node, o_j, conditioned to another node $s_{o_j} \in (\mathbb{O}_t \setminus \{o_j\})$. Such a node s_{o_j} is either the first observed node encountered along the unique path connecting o_j to o_t, or node o_t itself. Finally, thanks to the chain rule, the rightmost term in Eq. (7) can be expressed in terms of quantities computed at previous stages

$$\mathcal{H}_{pair}(o_1,\ldots,o_{t-1}) = \mathcal{H}_{pair}(o_{t-1}|o_1,\ldots,o_{t-2})+ \\ \mathcal{H}_{pair}(o_1,\ldots,o_{t-2}) \quad . \tag{9}$$

If the graph is not a tree, the algorithm is still applicable as long as the structure is sufficiently treelike. Many real-world networks satisfy this condition [6], and, very often, tree-like approximations are effective even if the graphs are not treelike [34]. In our proposal, if the graph contains loops, $\mathcal{H}_{pair}(o_1,\ldots,o_{t-1}|o_t)$ is computed under the tree assumption by generating a spanning tree rooted in o_t, and using again Eq. (8). This provides us with an upper-bound of the true entropy, since neglecting dependencies necessarily leads to an entropy larger than its true value. The rooted tree can be generated arbitrarily. However, to keep the upper-bound as tight as possible, we use a Djikstra-like algorithm suitably modified for this context [24]. Results presented here are based on this choice.

The algorithm requires prior knowledge of the unconditional entropy of individual nodes, and of the pairwise entropy among pairs of nodes. From a computational point of view, the running time scales as N^3 in the worst-case scenario [24]. However, some computational tricks allow for a great reduction of the complexity of the MES algorithm [19, 35], which effectively scales as $N^2 \log(N)$ [24]. This makes the algorithm easily applicable even to relatively large systems.

To validate the algorithm, we consider four different processes: i) the Ising model [36]; ii) the Modified version of the Susceptible-Infected-Susceptible (MSIS) model as proposed in Ref. [37]; iii) the standard version of the Susceptible-Infected-Susceptible (SIS) model [21]; iv) the Independent Cascade (IC) model [38]. The first two models satisfy detailed balance. The standard versions of the SIS and IC models are

instead prototypical examples of out-of-equilibrium processes that don't satify detailed balance. We analyze the behavior of all models for different parameter values and on different network substrates, including synthetic graphs and real-world networks. Results and details of our systematic analysis are reported in [24].
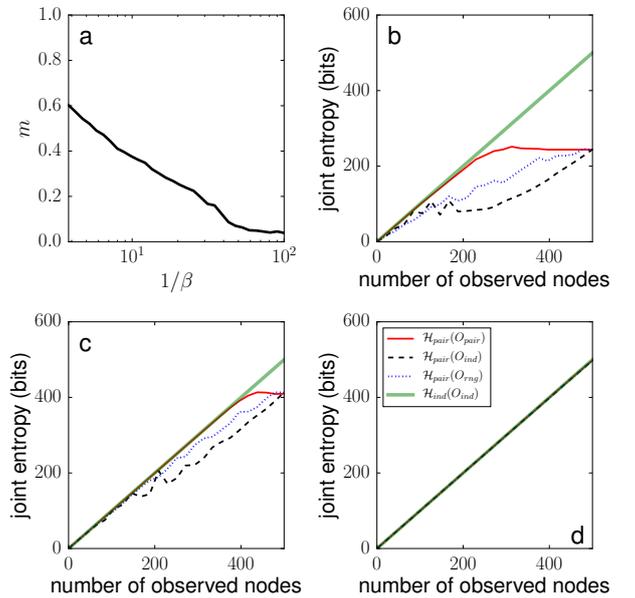

Figure 1. Ising model on the US air transportation network. a) Magnetization m as a function of the temperature $1/\beta$. b) Joint entropy of the observed set as a function of the set size. Here $1/\beta = 1$. Sampling techniques considered are: i) MES (red full line); ii) MES under the IBMF approximation (black dashed line); iii) random sampling, i.e., nodes are added to the observed set in random order (blue dotted line). For all sampling techniques, joint entropy is measured using the novel approximation. We plot also the joint entropy according to the IBMF approximation for the set $\mathbb{O}_{ind}$ (thick green line). c) Same as in panel b, but for $1/\beta = 15$. d) Same as in panels b, but for $1/\beta = 50$.

In Fig. 1 we show results for the Ising model applied to the US air transportation network (size $N = 500$) originally considered in [39]. The network contains loops, so that our approximation is not exact. We sample configurations reached by the system after a sufficiently long number of iterations of the Metropolis algorithm with fixed value of the temperature $1/\beta$ and external magnetic field $h = 1/N$. Every realization is obtained after $1,000\,N$ total spin flips. The phase diagram of the system is presented in Fig. 1a, showing the typical transition from ordered to disordered configurations as the temperature is increased. We first analyze statistical properties of microscopic configurations obtained at $1/\beta = 1$ in Fig. 1b. To estimate the unconditional entropy $\mathcal{H}(i)$ of a generic node i, and the pairwise conditional entropy $\mathcal{H}(j|i)$ of a generic pair (i, j), we rely on $T = 1,000$ sampled configurations. In addition to $\mathbb{O}_{pair}$ and $\mathbb{O}_{ind}$, we consider also the set of observed nodes $\mathbb{O}_{rng}$, built by adding nodes in random order. As the figure clearly

shows, our approximation generates noticeable improvements with respect to the the IBMF approximation in the computation of the entropy of subsets of the system. This is apparent from the large value of the difference $\mathcal{H}_{ind}(\mathbb{O}_{ind}) - \mathcal{H}_{pair}(\mathbb{O}_{ind})$. Concerning different sampling strategies, we also see a significant benefit from using our proposed technique over the naive version of MES. $\mathcal{H}_{pair}(\mathbb{O}_{pair})$ grows much quicker than $\mathcal{H}_{pair}(\mathbb{O}_{ind})$, and saturates at the maximum value after about 200 nodes are observed. This is an indication that the entire uncertainty of the system can be explained by looking at a fraction of the nodes in the network only. On the contrary, $\mathcal{H}_{pair}(\mathbb{O}_{ind})$ behaves very similarly to, if not worse than, $\mathcal{H}_{pair}(\mathbb{O}_{rng})$ testifying that the naive MES strategy is not effective in this specific system. As the temperature increases (Fig. 1c), the advantage of using our new approximation in place of the IBMF approximation becomes less apparent. At the same time, the benefit of using our greedy MES strategy compared to the naive version becomes less evident. For very large temperatures, all curves become identical (Fig. 1d).
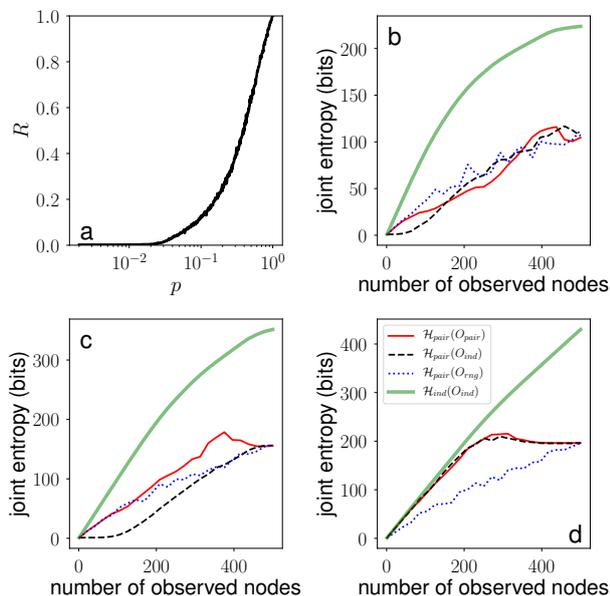

Figure 2. Independent cascade model on the US air transportation network. a) Relative size of the outbreak R as a function of the infection probability p. Panels b, c, and refer respectively to $p = 0.1$, $p = 0.2$, and $p = 0.5$. The description of the curves in these panels is identical to one of the curves appearing in Fig. 1.

In Fig. 2, we study the IC model applied to the same real-world network. We focus on microscopic configurations corresponding to the final stage of the dynamics, where nodes are either in the susceptible or recovered state. The initial condition of the dynamics is given by all nodes in the susceptible state, except for a single randomly chosen seed in the infected state. Infections propagate along each active edge with probability p. For every value of p, we consider $T = 100,000$ sampled configurations. In the IC model on a loopy graph, both assumptions at the basis of our approach are violated.

Nonetheless, the results reveal that our approximation represents a significant improvement over the basic IBMF approximation. First, we are able to provide estimates of the entropy of the system that are radically smaller, showing that pairwise correlations among variables are particularly significant in the system. Second, we are able to construct sets of observed nodes that are more representative for system uncertainty than those obtained by using the other sampling strategies.

To further strengthen our message, in [24] we include a comparison of the performance between our greedy algorithm for MES and other selection strategies: (i) degree centrality sampling, where nodes are added in decreasing (increasing) order based on their degree; (ii) closeness centrality sampling, the same as (i) but with node ranking based on closeness centrality. These strategies are chosen to test the performance of centrality-based metrics that rely on topological properties only. The most significant difference between them is that degree is a local metric, whereas closeness is global. We find that topological heuristics are not always reliable sampling strategies, and that their effectiveness is seriously affected by the underlying network structure and/or the parameter values of the stochastic models. Further in [24], we study analytically the behavior of the IC model in star networks and show that the choice of the best nodes to observe is highly sensitive not only to the parameter of the model, but also to the initial configuration of the stochastic dynamical process.

In summary, we introduce an algorithm to approximate the conditional entropy of a sample of nodes in a complex network. The algorithm relies on the sparsity of the graph to simplify computations otherwise unfeasible. Although the algorithm allows us to compute the conditional entropy of arbitrary node sets, it finds a particularly interesting application in the greedy approximation of the so-called MES principle. This principle corresponds to the optimal reduction of uncertainty of a stochastic process taking place on a network. Combining our algorithm with machine learning methods to create active supervised learning approaches is a potentially interesting direction for future investigation. Other extensions worth of consideration are also the generalization of our algorithm to devise computationally feasible selection strategies based on other information-theoretic principles, as for example the maximization of the mutual information rather than entropy.

* filiradi@indiana.edu

[1] P. C. Bressloff, *Stochastic processes in cell biology* (Springer, Heidelberg, 2014).

[2] R. Lande, S. Engen, and B.-E. Saether, *Stochastic population dynamics in ecology and conservation* (Oxford University Press, New York, 2003).

[3] N. G. van Kampen, *Stochastic processes in physics and chem-*

istry (Elsevier, Amsterdam, 1995).

[4] C. Laing and G. J. Lord, *Stochastic methods in neuroscience* (Oxford University Press, New York, 2010).

[5] P. Wolfgang and B. Jorg, *Stochastic Processes: From Physics to Finance* (Springer, Heidelberg, 2013).

[6] M. Newman, *Networks: an introduction* (Oxford University Press, New York, 2010).

[7] M. C. Shewry and H. P. Wynn, J. Appl. Stat. **14**, 165 (1987).

[8] K. Chaloner and I. Verdinelli, Statist. Sci. **10**, 273 (1995).

[9] A. Krause, A. Singh, and C. Guestrin, J. Mach. Learn. Res. **9**, 235 (2008).

[10] C.-W. Ko, J. Lee, and M. Queyranne, Oper. Res. **43**, 684 (1995).

[11] J. Lee (John Wiley & Sons, Chichester, 2002) pp. 1229 – 1234.

[12] C. Guestrin, A. Krause, and A. P. Singh, in *Proceedings of the 22nd international conference on Machine learning* (ACM, 2005) pp. 265–272.

[13] J. Leskovec and C. Faloutsos, in *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '06 (ACM, New York, NY, USA, 2006) pp. 631–636.

[14] S. K. Thompson, arXiv:1306.0817 (2013).

[15] M. Bilgic, L. Mihalkova, and L. Getoor, in *Proceedings of the 27th international conference on machine learning (ICML-10)* (2010) pp. 79–86.

[16] C. Moore, X. Yan, Y. Zhu, J.-B. Rouquier, and T. Lane, in *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining* (ACM, 2011) pp. 841–849.

[17] L. Peel, J. Compl. Netw. **3**, 431 (2014).

[18] A. Haber, F. Molnar, and A. E. Motter, IEEE Trans. Control Netw. Syst. **PP**, 1 (2017).

[19] A. Krause and D. Golovin, Tractability: Practical Approaches to Hard Problems **3**, 8 (2012).

[20] G. L. Nemhauser, L. A. Wolsey, and M. L. Fisher, Math. Program. **14**, 265 (1978).

[21] R. Pastor-Satorras, C. Castellano, P. Van Mieghem, and

[22] J. Moussouris, J. Stat. Phys. **10**, 11 (1974).

[23] M. Mezard and A. Montanari, *Information, physics, and computation* (Oxford University Press, New York, 2009) .

[24] See Supplemental Material [url] for details on the algorithm and additional empirical results, which includes Refs. [25-33].

[25] W.K. Hastings, Biometrika **57**, 97 (1970).

[26] J. Goldenberg, B. Libai, and E. Muller, Mark. Lett. **12**, 211 (2001).

[27] D.T. Gillespie, J. Chem. Phys. **81**, 2340 (1977).

[28] D. de Solla Price, J. Assoc. Inf. Sci. Technol. **27**, 292 (1976).

[29] M. Catanzaro, M. Boguná, and R. Pastor-Satorras, Phys. Rev. E **71**, 027103 (2005).

[30] R. Guimerá, L. Danon, A. Díaz-Guilera, F. Giralt, and A. Arenas, Phys. Rev. E **68**, 065103 (2003).

[31] D.J. Watts and S.H. Strogatz, Nature **393**, 440 (1998).

[32] L.A. Adamic and N. Glance, in *Proceedings of the 3rd international workshop on Link discovery* (ACM, 2005) pp. 3643.

[33] J. Leskovec, J. Kleinberg, and C. Faloutsos, in *Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining* (ACM, 2005) pp. 177187.

[34] S. Melnik, A. Hackett, M. A. Porter, P. J. Mucha, and J. P. Gleeson, Phys. Rev. E **83**, 036112 (2011).

[35] J. Leskovec, A. Krause, C. Guestrin, C. Faloutsos, J. Van-Briesen, and N. Glance, in *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining* (ACM, 2007) pp. 420–429.

[36] S. N. Dorogovtsev, A. V. Goltsev, and J. F. F. Mendes, Rev. Mod. Phys. **80**, 1275 (2008).

[37] J. P. Gleeson, Phys. Rev. X **3**, 021004 (2013).

[38] D. Kempe, J. Kleinberg, and É. Tardos, in *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining* (ACM, 2003) pp. 137–146.

[39] V. Colizza, R. Pastor-Satorras, and A. Vespignani, Nat. Phys. **3**, 276 (2007).