Collective influence maximization in threshold models of information cascading with first-order transitions

Sen Pei,[1] Xian Teng,[2] Jeffrey Shaman,[1] Flaviano Morone,[2] and Hernán A. Makse[2,*]

[1] *Department of Environmental Health Sciences, Mailman School of Public Health, Columbia University, New York, NY 10032, USA*

[2] *Levich Institute and Physics Department, City College of New York, New York, NY 10031, USA*

(Dated: June 1, 2016)

In spreading dynamics in social networks, there exists an optimal set of influencers whose activation can induce a global-scale cascade of information. To find the optimal, or minimal, set of spreaders, a method based on collective influence theory has been proposed for spreading dynamics with a continuous phase transition that can be mapped to optimal percolation. However, when it comes to diffusion processes exhibiting a first-order, or discontinuous transition, identifying the set of optimal spreaders with a linear algorithm for large-scale networks still remains a challenging task. Here we address this issue by exploring the collective influence in general threshold models of opinion cascading. Our analysis reveals that the importance of spreaders is fixed by the subcritical paths along which cascades propagate: the number of subcritical paths attached to each spreader determines its contribution to global cascades. The concept of subcritical path allows us to introduce a linearly scalable algorithm for massively large-scale networks. Results in both synthetic random graphs and real networks show that the set of spreaders predicted by our method is smaller than those identified by other linearly scalable heuristic approaches.

I. INTRODUCTION

In natural and social systems, spreading dynamics lie at the heart of a variety of complex phenomena including failure propagation in infrastructure [1], adoption of new behaviors [2], and diffusion of norms and innovations in social networks [3]. In these spreading processes, a small number of influencers arise as a consequence of the structural diversity of the underlying contact networks [4]. In different fields, it has been accepted that the initial activation of such "superspreaders", who usually hold prominent locations in networks, is capable of shaping the spreading dynamics of large populations [5–9]. From a practical point of view, identification of superspreaders could not only boost product promotion in a viral marketing campaign constrained by a limited budget, but also instruct the strategic protection of structurally pivotal units to maintain robust infrastructures at a lower cost. Given its great practical values in a range of important applications, the problem of locating superspreaders has attracted immediate attention in various disciplines [10–20].

In the simple case of finding single influential spreader, centrality-based heuristic measures such as degree [21], Betweenness [22], PageRank [23] and K-core [9, 10, 24, 25] are routinely adopted. Beyond this non-interacting problem of finding single spreaders, it becomes more complicated when trying to select a group of spreaders, due to the collective effects of multiple agents. In fact, searching for the optimal set of influencers in spreading dynamics is an NP-hard problem and remains to be a challenging conundrum in network science [4]. To address the many-body problem, an analytical framework of collective influence (CI) has been recently established for optimal percolation in random graphs [12]. Based on stability analysis of zero solution, CI theory applies to the dynamical processes that can be transformed to optimal percolation with a continuous phase transition. One of such spreading processes is Linear Threshold Model (LTM) with a fixed threshold $m_i = k_i - 1$, where k_i denotes node i's degree.

In a large variety of contexts, LTM has been frequently employed to describe a spreading process during which an individual make decisions under "peer pressure" [26–35]. That is, a node will adopt a piece of information only after a certain number of its neighbors have accepted it. The choice of threshold $m_i = k_i - 1$ in LTM guarantees a continuous phase transition thus CI method can be applied accordingly [12]. Nevertheless, for other choices of threshold, there also exist a wide class of LTMs that exhibit a first-order, or discontinuous phase transition. Under this condition, the stability analysis in CI method is not applicable any more. Therefore, how to determine the collective influence of the optimal seeds for first-order transitions in LTM should be studied separately. Although several approaches from different views have been developed, e.g., greedy search [4, 11] and message passing algorithm [13], the problem of developing an efficient linear algorithm for large-scale networks still needs to be explored.

Here, we examine the collective influence in general LTM, and develop a scalable algorithm to locate optimal spreaders. By analyzing the message passing equations of LTM, we find the form of interactions between spreaders and derive the analytical form of their contributions to information cascading. Each seed's contribution, defined as the *collective influence in threshold model* (CI-TM) with first-order transitions, is determined by the number

* hmakse@lev.ccny.cuny.edu

of subcritical paths emanating from it. Since the subcritical paths are such routes along which cascades can propagate, CI-TM can be considered as a reliable measure of seeds' structural importance in LTM. In an attempt to find the optimal superspreaders for first-order transitions in big-data analysis, we endeavor to maximize the number of active population by adaptively selecting nodes with the largest CI-TM values. Extensive comparisons with other plausible competing heuristics on both synthetic and realistic large-scale networks reveal that the proposed mechanism-based algorithm can indeed identify a smaller set of spreaders that could initiate global scale cascade.

II. COLLECTIVE INFLUENCE IN THRESHOLD MODELS: CI-TM

We present a theoretical framework to analyze the collective influence of individuals in general LTM. For a network with N nodes and M links, the topology is represented by the adjacency matrix $\{A_{ij}\}_{N\times N}$, where $A_{ij}=1$ if i and j are connected, and $A_{ij}=0$ otherwise. The vector $\mathbf{n}=(n_1,n_2,\cdots,n_N)$ records whether a node i is chosen as a seed ($n_i=1$) or not ($n_i=0$). The total fraction of seeds is therefore $q=\sum_i n_i/N$. During the spreading, the state of each node falls into the category of either active or inactive. The spreading starts from a q fraction of active seeds and evolves following a threshold rule: a node i becomes active when m_i neighbors get activated will a node i become active. This process terminates when there are no more newly activated nodes. We introduce ν_i as node i's probability in active ($\nu_i=1$) or inactive ($\nu_i=0$) state at the final stage, and denote $Q(q)$ as the size of the giant connected component of active population.

For a directed link $i\to j$, we introduce $\nu_{i\to j}$ as the probability of i being in an active state assuming node j is disconnected from the network [36]. If $n_i=1$, then $\nu_{i\to j}=1$. Otherwise, $\nu_{i\to j}=1$ only when there are at least m_i active neighbors excluding j. Since there exist many possible choices of these m_i neighbors, we define $P^{m_i}_{\partial i\setminus j}$ as the set of all combinations of m_i nodes selected from $\partial i\setminus j$, where $\partial i\setminus j$ is the set of nearest neighbors of i excluding j. Clearly, if i has k_i connections emanating from it, there are $\binom{k_i-1}{m_i}$ combinations, so the set $P^{m_i}_{\partial i\setminus j}$ contains $\binom{k_i-1}{m_i}$ elements, denoted by P_h, $h=1,\cdots,\binom{k_i-1}{m_i}$. Each element P_h has the form $P_h=\{p_{h_1},\cdots,p_{h_{m_i}}\}$ where $\{p_{h_1},\cdots,p_{h_{m_i}}\}$ are the m_i nodes in the hth combination. Figure 1(a) illustrates all three combinations P_1,P_2 and P_3 corresponding to $\nu_{i\to j}$ for node i with a threshold $m_i=2$. Should at least one combination is fully activated, we have $\nu_{i\to j}=1$.

Generally, for locally tree-like networks, we have the following message passing equation:

$$\nu_{i\to j}=n_i+(1-n_i)[1-\prod_{P_h\in P^{m_i}_{\partial i\setminus j}}(1-\prod_{p\in P_h}\nu_{p\to i})]. \quad (1)$$

The final state of i is given by

$$\nu_i=n_i+(1-n_i)[1-\prod_{P_h\in P^{m_i}_{\partial i}}(1-\prod_{p\in P_h}\nu_{p\to i})]. \quad (2)$$

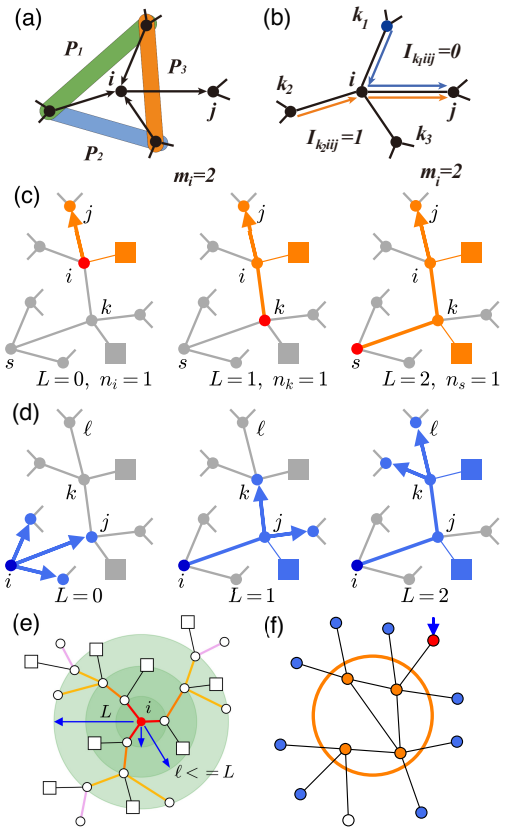

FIG. 1. Subcritical paths and collective influence of spreaders. (a), Three combinations of neighbors P_1, P_2 and P_3 corresponding to $\nu_{i\to j}$ in message passing equation. Node i has a threshold $m_i=2$. The full activation of at least one combination will lead to $\nu_{i\to j}=1$.(b), For link $i\to j$ with an active neighbor k_1 and inactive ones k_2 and k_3, $I_{k_1\to i,i\to j}=0$ since i has 0 ($<m_i-1$) active neighbor excluding k_1 and j, while $I_{k_2\to i,i\to j}=1$ because i has 1 ($=m_i-1$) active neighbor k_1 excluding k_2 and j. (c), Illustrations of subcritical paths ending with link $i\to j$ for $L=0$, 1, 2. Red dots stand for seeds, while squares represent $m-1$ active neighbors attached to subcritical nodes. Subcritical paths are highlighted by thick links. (d), The contribution of seed i to $\|\nu_{\to}\|$ exerted through subcritical paths of length $L=0$, 1, 2. (e), Calculation method of CI-TM$_L(i)$. Subcritical paths starting from i with length $\ell\le L$ are displayed by different colors. (f), An example of subcritical cluster. Assuming a uniform threshold $m=3$, nodes inside the circle are subcritical since they all have 2 active neighbors, represented by blue nodes. Activation of the red node will trigger a cascade covering all subcritical nodes.

The above equations Eq. (1-2) describe the general cases of LTM. For the special choice of threshold $m_i = k_i - 1$, there is only one combination in $P^{m_i}_{\partial i\backslash j}$, and the transition becomes continuous. In this case, Eq. (1) recovers the message passing equation of optimal percolation as treated in Ref. [12] (See Appendix A). For any other choice of m_i (except for $m_i = 1$, which is equivalent to the problem of optimal immunization [12]), the multiplicity of activated neighbors leads to a first-order transition and cascading as discussed here.

For all the $2M$ directed links $i \to j$, Eq. (1) is a nonlinear function of $\nu_\to = (\cdots, \nu_{i\to j}, \cdots)^T$:

$$\nu_\to = \mathbf{n}_\to + \mathbf{G}(\nu_\to). \tag{3}$$

In Eq. (3), $\mathbf{n}_\to = (\cdots, n_{i\to j}, \cdots)^T$ in which $n_{i\to j} = n_i$ for link $i \to j$, and $\mathbf{G} = (\cdots, G_{i\to j}, \cdots)^T$ where $G_{i\to j}$ is the nonlinear function of vector $\nu_\to$ for link $i \to j$. Given the initial configuration of seeds $\mathbf{n}$, the final state of $\nu_\to$ is fully determined by the self-consistent Eq. (3). Unfortunately, it cannot be solved directly due to the exponentially growing number of combinations in $P^{m_i}_{\partial i\backslash j}$. Therefore, for a small number of seeds, we adopt the iterative method to estimate the solution. In this point of view, Eq. (3) can be treated as a discrete dynamical system

$$\nu^{t+1}_\to = \mathbf{n}_\to + \mathbf{G}(\nu^t_\to) \tag{4}$$

with the initial condition $\nu^0_\to = \mathbf{n}_\to$.

To simplify the calculation, we approximate the nonlinear function $G_{i\to j}$ by linearization. Define $G'_{i\to j}(\nu_\to) = (\cdots, \frac{\partial G_{i\to j}}{\partial \nu_{k\to\ell}}, \cdots)$. By Eq. (1), we know that $\frac{\partial G_{i\to j}}{\partial \nu_{k\to\ell}} = 0$ for $\ell \neq i$. While in the case of $\ell = i$ and $k \neq j$, we have

$$\frac{\partial G_{i\to j}}{\partial \nu_{k\to i}} = (1-n_i) \prod_{P_h \in P^{m_i}_{\partial i\backslash j}, k\notin P_h} (1 - \prod_{p\in P_h} \nu_{p\to i}) \sum_{P_h \in P^{m_i}_{\partial i\backslash j}, k\in P_h} [(\prod_{p\in P_h\backslash k} \nu_{p\to i}) \prod_{P'_h \neq P_h, k\in P'_h} (1 - \prod_{p\in P'_h} \nu_{p\to i})]. \tag{5}$$

Although Eq. (5) has a complex form, in fact it is only determined by a simple quantity $a_{k\to i,i\to j} = \sum_{p\in\partial i\backslash(k,j)} \nu_{p\to i}$, which is interpreted as the number of i's active neighbors excluding k and j when i is absent from the network. On one hand, if $a_{k\to i,i\to j} \geq m_i$, at least one term of $\prod_{p\in P_h} \nu_{p\to i}$ equals one, since we are selecting m_i elements from a set containing at least m_i elements of value 1. Under such condition, $\frac{\partial G_{i\to j}}{\partial \nu_{k\to i}} = 0$. On the other hand, if $a_{k\to i,i\to j} \leq m_i - 2$, all the terms $\prod_{p\in P_h\backslash k} \nu_{p\to i}$ are zeros because we are selecting $m_i - 1$ elements from a set containing at most $m_i - 2$ nonzero elements, which also leads to $\frac{\partial G_{i\to j}}{\partial \nu_{k\to i}} = 0$. When $a_{k\to i,i\to j} = m_i - 1$, all the terms $\prod_{p\in P_\ell} \nu_{p\to i}$ are zeros, and only the exact combination of these $m_i - 1$ nonzero elements would lead to $\prod_{p\in P_h\backslash k} \nu_{p\to i} = 1$. Therefore, we have $\frac{\partial G_{i\to j}}{\partial \nu_{k\to i}} = 1 - n_i$. Based on the above reasoning, we define a quantity $I_{k\to\ell,i\to j}$ for links $k \to \ell$ and $i \to j$ as follows:

$$I_{k\to\ell,i\to j} = \begin{cases} 1 & \text{if } \ell = i, k \neq j, a_{k\to i,i\to j} = m_i - 1, \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

The definition of $I_{k\to\ell,i\to j}$ is reminiscent of the Hashimoto non-backtracking (NB) matrix $\mathcal{B}$ [12, 37–42]. In the case of $m_i = k_i - 1$, our quantity $I_{k\to\ell,i\to j}$ can be transformed to the corresponding element of NB matrix $\mathcal{B}_{k\to\ell,i\to j}$ in optimal percolation [12] (See Appendix A). In fact, $I_{k\to\ell,i\to j}$ is closely related to the concept of subcritical nodes. Recall that a node i is subcritical if it has $m_i - 1$ active neighbors [33–35]. This implies that one more activation of its neighbors will cause i activated. From Eq. (6) we know that $I_{k\to\ell,i\to j} = 1$ only if the links $k \to \ell$ and $i \to j$ are connected, non-backtracking, and additionally, node i is subcritical in the absence of node k and j. In Fig. 1(b), node i has an active neighbor k_1 and two inactive ones k_2 and k_3. By definition, for a threshold $m_i = 2$, we conclude $I_{k_1\to i,i\to j} = 0$ since i has no active neighbor excluding k_1 and j, while $I_{k_2\to i,i\to j} = 1$ because i has 1 $(= m_i - 1)$ active neighbor excluding k_2 and j.

For a small $\nu_\to$, a standard linearization around origin $\mathbf{0}$ gives $G_{i\to j}(\nu_\to) \approx G_{i\to j}(\mathbf{0}) + G'_{i\to j}(\mathbf{0})\nu_\to$. However this will cause degeneracy since Eq. (5) constantly gives $G'_{i\to j}(\mathbf{0}) = \mathbf{0}$. Therefore, we approximate $G_{i\to j}(\nu_\to)$ by $G_{i\to j}(\mathbf{0}) + G'_{i\to j}(\nu_\to)\nu_\to$ given $\nu_\to$ is close to $\mathbf{0}$. In Appendix B, we prove that this linearization has an approximation accuracy of $O(|\nu_\to|^2)$ ($|\cdot|$ is the vector norm), same as the linear Taylor expansion.

Combining all direct links, Eq. (4) can be approximated by a linear equation

$$\nu^{t+1}_\to = \mathbf{n}_\to + \mathcal{F}^t \nu^t_\to, \tag{7}$$

where $\mathcal{F}^t = (\cdots, G'_{i\to j}(\nu^t_\to), \cdots)^T$ is a $2M \times 2M$ matrix defined on the directed links $k \to \ell$, $i \to j$ with elements

$$\mathcal{F}^t_{k\to\ell,i\to j} = \left.\frac{\partial G_{i\to j}}{\partial \nu_{k\to\ell}}\right|_{\nu^t_\to}. \tag{8}$$

With the notion of $I_{k\to\ell,i\to j}$, we can write $\mathcal{F}^t$ as:

$$\mathcal{F}^t_{k\to\ell,i\to j} = (1-n_i) I^t_{k\to\ell,i\to j}. \tag{9}$$

Now we update the state of $\nu^t_\to$ following Eq. (7). For the convenience of calculation, we put the matrix $\mathcal{F}^t$ in a higher-dimensional space [12]:

$$\mathcal{F}^t_{k\ell ij} = (1-n_i) A_{k\ell} A_{ij} \delta_{i\ell} (1-\delta_{jk}) I^t_{k\ell ij}, \tag{10}$$

where function $\delta_{i\ell}$ is 1 if $i = \ell$, and 0 otherwise. The indices k, ℓ, i, j runs from 1 to N. Starting from $\nu^0_\to = \mathbf{n}_\to$, $\nu^1_\to = \mathbf{n}_\to + \mathcal{F}^0 \mathbf{n}_\to$ gives

$$\nu^1_{i\to j} = n_i + (1-n_i) A_{ij} \sum_k A_{ki} (1-\delta_{jk}) I^0_{kiij} n_k. \tag{11}$$

The physical meaning of Eq. (11) can be interpreted as follows. If node i is a seed, $\nu^1_{i\to j}=1$. Otherwise, $\nu^1_{i\to j}$ is nonzero if i is subcritical ($I^0_{kiij}=1$) and at least one of its corresponding neighbors k is a seed ($n_k=1$). Supposing i is not a seed, the contribution of a neighboring seed k is conveyed by the directed link $k\to i\to j$ that satisfies $n_k=1, n_i=0$ and $I^0_{kiij}=1$, which is shown in the second panel of Fig. 1(c).

For $t=2$, we have $\nu^2_{\to}=\mathbf{n}_{\to}+\mathcal{F}^1\mathbf{n}_{\to}+\mathcal{F}^1\mathcal{F}^0\mathbf{n}_{\to}$. Therefore,

$$\nu^2_{i\to j}=n_i+(1-n_i)A_{ij}\sum_k A_{ki}(1-\delta_{jk})I^1_{kiij}n_k+(1-n_i)$$
$$A_{ij}\sum_k(1-n_k)A_{ki}(1-\delta_{jk})I^1_{kiij}\sum_s A_{sk}(1-\delta_{is})I^0_{skki}n_s. \tag{12}$$

The last term in Eq. (12) is actually the contribution of node i's 2-step neighbors s to $\nu^2_{i\to j}$. The contribution of a seed s is conducted through a directed path $s\to k\to i\to j$ that satisfies $n_s=1, n_k=0, n_i=0$ and $I^0_{skki}=1, I^1_{kiij}=1$ (See Fig. 1(c)).

Inspired by Eq. (12), we define the concept of *subcritical paths*. For a directed link $i\to j$, the path $i_1\to i_2\to\cdots\to i_L\to i\to j$ is a subcritical path of length L if $n_{i_1}=1, n_{i_2}=0,\cdots,n_i=0$, $I^0_{i_1i_2i_2i_3}=1,\cdots,I^{L-1}_{i_Liij}=1$, and any two consecutive links are non-backtracking. If $i_1=i$, we set the path's length $L=0$. The subcritical paths of length $L=0$, 1 and 2 are displayed in Fig. 1(c). Notice that, the calculation of L-length subcritical paths is in fact implemented by the multiplication of $\mathcal{F}^{L-1}\cdots\mathcal{F}^0$. Following this idea, we can generalize Eq. (12) to $\nu^T_{i\to j}$ at a given time T. The exact formula for $\nu^T_{i\to j}$ is quite lengthy, so we do not display it here. But we should keep in mind that the form of $\nu^T_{i\to j}$ is nothing but n_i plus the contribution of seeds connected to i through subcritical paths with length $L\le T$ when $n_i=0$. Finally, we note that when $m_i=k_i-1$, the subcritical path becomes the minimal, or shortest, path in the problem of optimal percolation [12].

III. CI-TM ALGORITHM

To quantify the active population in LTM, we define $\|\nu_{\to}\|=\sum_{ij}\nu_{i\to j}$. Starting from $\|\nu_{\to}\|=0$ when no seed is selected, $\|\nu_{\to}\|$ increases as more seeds are activated. Therefore, we expect that the first-order transition would appear early if $\|\nu_{\to}\|$ is maximized for a given number of seeds.

Based on the form of each element in $\nu_{\to}$, we learn that the contribution of a seed i to $\|\nu_{\to}\|$ is composed of all its collective contributions to every potential element, exerted through the subcritical paths attached to i. Therefore, we employ a seed's contribution to $\|\nu_{\to}\|$ to define its Collective Influence in Threshold Model (CI-TM) to find the best influencers. For the trivial case of subcritical paths with length $L=0$, we define CI-TM$_0(i)=k_i$, where k_i is the degree of node i. Thus, at the zero-order approximation we recover the high-degree heuristic. The first panel of Fig. 1(d) illustrates CI-TM$_0(i)=3$ for node i. For $L\ge 1$, subcritical paths are involved in the definition of CI-TM. For $L=1$,

$$\text{CI-TM}_1(i)=k_i+\sum_{j\in\partial i}(1-n_j)\sum_{k\in\partial j\setminus i}I^0_{ijjk}. \tag{13}$$

As shown in Fig. 1(d), the contribution of node i to $\|\nu_{\to}\|$ through subcritical paths of length $L=1$ is 2. Therefore, we have CI-TM$_1(i)=5$. For $L=2$,

$$\begin{aligned}\text{CI-TM}_2(i)=&k_i+\sum_{j\in\partial i}(1-n_j)\sum_{k\in\partial j\setminus i}I^0_{ijjk}\\&+\sum_{j\in\partial i}(1-n_j)\sum_{k\in\partial j\setminus i}(1-n_k)I^0_{ijjk}\sum_{\ell\in\partial k\setminus j}I^1_{jkk\ell}\end{aligned} \tag{14}$$

In Fig. 1(d), the additional 2-length subcritical paths also contribute to CI-TM$_2(i)$, leading to CI-TM$_2(i)=7$. Moreover, in Fig. 1(d), we can observe that for the tree structure, the activation of node j in the first-step update will not affect $I^1_{jk\ell}$ in the second-step update, which means $I^1_{jk\ell}=I^0_{jk\ell}$. More generally, $I_{jk\ell}$ is not affected by the activation of k's any precedent nodes on the subcritical path. Therefore, we will leave out the superscript t in the definition of CI-TM for locally tree-like networks. We can generalize the above CI-TM calculation to any given L. In summary, the definition of node i's influence CI-TM in an area of length L is:

$$\begin{aligned}\text{CI-TM}_L(i)=&\text{number of subcritical paths starting}\\&\text{from } i \text{ with length } 0\le\ell\le L.\end{aligned} \tag{15}$$

Figure 1(e) illustrates the calculation of node i's CI-TM for $L=2$, in which subcritical paths with length $\ell\le L$ are distinguished by colors.

It is important to note that in the case of $m_i=k_i-1$, the subcritical path becomes shortest path. Therefore the calculation area in Fig. 1(e) recovers the ball $\mathcal{B}(i,L)$ of radius L centered at i. This is exactly the same ball in the definition of influence $CI_L(i)=(k_i-1)\sum_{j\in\partial\mathcal{B}(i,L)}(k_j-1)$ in Ref. [12]. Thus the algorithm to find influencers for the LTM with first-order transitions is a simple generalization of CI algorithm [12] for second order transition where the ball of influence $\mathcal{B}(i,L)$ in CI [12] defined by the shortest paths is replaced in CI-TM by a ball of influence, in this case, defined by the subcritical paths. This elegant mapping from a second order to a first-order transition model implies that all phenomenology found in Ref. [12] can be translated to the present case. In particular, weak nodes, such as those with modest degree but surrounded by hierarchy of hubs located across the subcritical paths, can be strong influencers in the LTM as well as in optimal percolation [12].

For a given fraction q of seeds, our goal is to maximize $\|\nu_{\to}\|$. As we have explained, the CI-TM value of a seed depends on the choice of other seeds. Therefore, it is

hard to obtain the optimal configuration $\{\mathbf{n}|\sum_i n_i/N = q\}$ without turning to extremely time-consuming algorithms. To compromise and obtain a linear algorithm, we propose an adaptive CI-TM algorithm following a greedy approach. Define $C(i, L)$ as the set of node i plus subcritical vertices belonging to all subcritical paths originating from i with length $\ell \leq L$. Beginning with an empty seed set S, we remove the top CI-TM nodes as follow. The calculation proceeds following the CI-TM algorithm.

Algorithm 1 CI-TM algorithm

1: Initialize $S = \emptyset$
2: Calculate CI-TM$_L$ for all nodes
3: **for** $l = 1$ to qN **do**
4: Select i with the largest CI-TM$_L$
5: $S = S \bigcup \{i\}$
6: Remove $C(i, L)$, and decrease the degree and threshold of $C(i, L)$'s existing neighbors by 1
7: Update CI-TM$_L$ for nodes within $L + 1$ steps from $C(i, L)$
8: **end for**
9: Output S

In the above algorithm, we remove $C(i, L)$ once i is added to the seed set. The reason lies in that it is unnecessary to select nodes in $C(i, L)$ as seeds in later calculation, because the activation of i will definitely active $C(i, L)$ (See Fig. 1(f)). Besides, $C(i, L)$ can be identified during the computation of CI-TM$_L$ without additional cost. In traditional centrality-based methods, seeds may have significant overlap in their influenced population. It has been reported that the performance of these methods, such as K-core, suffers a lot from this phenomenon [10]. On the contrary, in our algorithm, this problem is alleviated by the removal of subcritical nodes in $C(i, L)$, which successfully reduces the overlap and improves the efficacy of each seed. Although such greedy strategy is not guaranteed to give the exact optimal spreaders, we expect a good performance in comparison with other heuristic methods in large-scale networks, as already found in Ref. [12].

More importantly, the CI-TM algorithm is linearly scalable for large networks with a computational complexity $O(N \log N)$ as $N \to \infty$. On one hand, computing CI-TM$_L$ is equivalent to iteratively visiting subcritical neighbors of each node layer by layer within L radius. Because of the finite search radius, computing CI-TM$_L$ for each node takes $O(1)$ time. Initially, we have to calculate CI-TM$_L$ for all nodes. However, during later adaptive calculation, there is no need to update CI-TM$_L$ for all nodes. We only have to recalculate for nodes within $L + 1$ steps from the removed vertices, which scales as $O(1)$ compared to the network size as $N \to \infty$ as shown in Ref. [20]. On the other hand, selecting the node with maximal CI-TM can be realized by making use of the data structure of heap that takes $O(\log N)$ time [20]. Therefore, the overall complexity of ranking N nodes is $O(N \log N)$ even when we remove the top CI-TM nodes one by one. In addition, considering the relative small number of subcritical neighbors, the cost of searching for subcritical paths is far less than that when scanning all neighbors. This permits the efficient computation of CI-TM for considerably large L. In our later experiments on finite-size networks, we do not put a limit on L so as to calculate CI-TM thoroughly. But remember that we can always truncate L to speed up CI-TM algorithm for extremely large-scale networks.

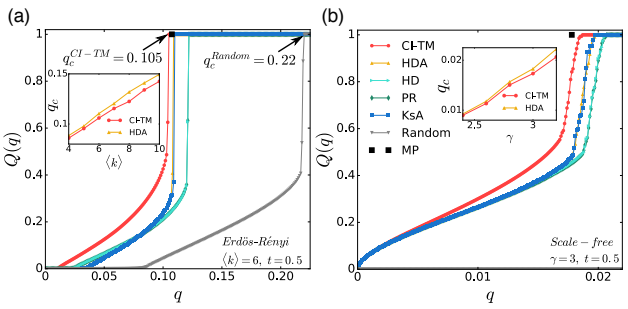

FIG. 2. Performance of CI-TM algorithm on random networks. (a), Size of active giant component $Q(q)$ versus the fraction of seeds q for ER networks with size $N = 2 \times 10^5$ and mean degree $\langle k \rangle = 6$. Different methods are distinguished by distinct markers and colors. Threshold is set as fractional $t = 0.5$. The CI-TM algorithm is run without limitation on L. MP is implemented by using $T = 40$ and a reinforcement parameter $r = 1 \times 10^{-4}$. For CI-TM, the identified critical value is $q_c^{\text{CI-TM}} = 0.105(1)$ while for Random selection $q_c^{\text{Random}} = 0.220(1)$. Inset presents the critical values q_c identified by HDA and CI-TM for different mean degree $\langle k \rangle$. (b), Comparison for scale-free networks with size $N = 2 \times 10^5$, power-law exponent $\gamma = 3$, minimal degree $k_{min} = 2$ and maximal degree $k_{max} = 1000$. The fractional threshold of LTM is also set as $t = 0.5$. Inset shows the critical values q_c for different exponents γ. All the results are averaged over 50 realizations.

IV. TEST OF CI-TM ALGORITHM

We first simulate LTM dynamics on synthetic random networks, including Erdös-Rényi (ER) and scale-free (SF) networks. In the models, we adopt a fractional threshold rule $m_i = \lceil t k_i \rceil$, which means that a node will be activated once t fraction of its neighbors are active ($\lceil \cdot \rceil$ is the ceiling function). In order to verify the efficacy of CI-TM algorithm, we compare its performance against several widely-used ranking methods, including high degree (HD) [21], high degree adaptive (HDA), PageRank (PR) [23] and K-core adaptive (KsA) [24], and message passing algorithm (MP) based on statistical physics [13]. Details about these strategies are explained in Appendix C. As a reference, we also display the result of random selection of seeds in the same figures.

Figure 2(a) presents $Q(q)$ versus q on ER networks ($t = 0.5$, $N = 2 \times 10^5$, $\langle k \rangle = 6$). Similar to bootstrap percolation on homogeneous networks, $Q(q)$ first under-

goes a continuous transition from $Q(q) = 0$ to nonzero, and then a first-order transition at a higher value of q_c [32]. Remarkably, CI-TM algorithm achieves the best performance for first-order transition by identifying the smallest number of spreaders ($q_c^{\text{CI-TM}} = 0.105$) to trigger global cascade. Besides, it also brings about the earliest continuous transition. Among all the strategies, random selection works worst, with a critical value $q_c^{\text{Random}} = 0.220$. Although the original K-core ranking has an unsatisfactory performance for multi-source spreading [10], the adaptive version KsA gives a better result similar to HDA. MP predicts a rather small set of spreaders as well, similar to CI-TM $q_c^{\text{MP}} = 0.107$, but it has the issue with convergence. Near the threshold, MP and other algorithms based on belief propagation (BP) do not converge to a stable solution. This is a general property of BP algorithm as applied to NP-hard problems. In order to drive the system to converge by brute force, a reinforcement process with a parameter r is applied, which produces a convergence iteration time of order $O(r^{-1})$ [13]. In implementation, the parameter r should be set small so that the true optimal result is not altered too much. As a result, the convergence time of MP increases accordingly. We find that, for a large-scale ER network with $N = 10^7$, CI-TM algorithm ($L = 3$) can be completed in less than 2.4 hours, while MP ($T = 40$, $r = 10^{-4}$) is estimated to require 132 days by extrapolation (See Fig. A1(d) in Appendix C). However, even when MP has prohibitive running time for large datasets, the performance is also expected to degrade as the system size increases. This is because the convergence time is expected to increase for $N \to \infty$ if performance needs to be the same. Alternatively, by increasing N and keeping the parameters T and r the same, a clear degrading in the performance is obtained where q_c^{MP} increases with N. This is clearly seen in the inset of Fig. A1(d) with an increase of q_c^{MP} as $N \to \infty$, away from the more optimal value $q_c^{\text{CI-TM}}$ as $N \to \infty$. We also provide the first-order critical value q_c for CI-TM and HDA on ER networks with different average degree $\langle k \rangle$ in the inset of Fig. 2(a). With the growth of $\langle k \rangle$, q_c increases and so does the difference between CI-TM and HDA. In some cases, q_c can be further improved by a simple modification on CI-TM (See Appendix D).

We then examine CI-TM's performance on SF networks with power-law degree distributions $P(k) \sim k^{-\gamma}$ in Fig. 2(b). We generate SF networks of size $N = 2 \times 10^5$ and power-law exponent $\gamma = 3$ following the configuration model [43]. It can be seen that the critical value of first-order transition becomes rather small for SF networks, due to the existence of highly connected hubs. Still, CI-TM algorithm outperforms other heuristic approaches and shows a comparable performance as MP. In addition, CI-TM algorithm produces the largest active component $Q(q)$ for any give q before the first-order transition. For different values of power-law exponent γ, CI-TM consistently has a lower critical value q_c than HDA, as shown in the inset of Fig. 2(b).

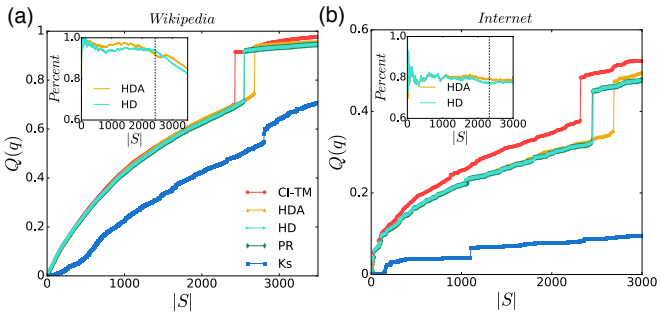

FIG. 3. Performance of CI-TM algorithm on large-scale real-world networks. (a), The relationship between the size of active giant component $Q(q)$ and the number of seeds $|S|$ for Wikipedia communication network, calculated by different methods. The LTM fractional threshold is set as $t = 0.65$. Inset displays the percentage of "real influencers" predicted by CI-TM that HDA and HD have successfully identified. The vertical dash line indicates the critical point of CI-TM. (b), Same analysis for Internet network with fractional threshold $t = 0.45$.

In applications, we are frequently faced with large-scale networks which exhibit more complicated topological characteristics than random graphs. Thus, it is more necessary and challenging to find a feasible strategy to efficiently approximate the optimal spreaders for those networks. Next, we explore CI-TM algorithm's performance for real networks. We examine two representative datasets: Wikipedia communication network ($N = 622,999, M = 2,890,729$) [44] and Internet autonomous system network ($N = 1,464,020, M = 10,863,640$) [45]. Wikipedia is an open access internet encyclopedia, whose contents are provided and edited by numerous contributors. Wikipedia network, as a typical social contact network used in information spreading, is constructed by the communication instances among users in English Wikipedia. An edge from user i to j indicates that i has written at least a message on the talk page of j. The Internet network records the autonomous system level topological structure of Internet, which provides an example of infrastructure network on which malicious attack and failure propagation may occur. Both networks are treated as undirected in the analysis.

Figure 3(a) displays $Q(q)$ for different number of seeds $|S|$ for Wikipedia network. CI-TM is able to trigger the global cascade with the smallest group of seeds, whose size is quite small compared to the entire network due to the heavy-tailed degree distribution. We also discover that, in the setting of first-order transitions, some "weak nodes" as defined in Ref. [12] are proven to be more collectively influential than those highly-connected hubs. As shown in the inset figure, we present the percentage of "real influencers" (predicted by CI-TM) that HDA and HD have identified, with the vertical dash line indicating CI-TM's q_c. At q_c, HDA and HD locate up to 90% overlapping seeds with CI-TM algorithm, most of which are tagged as hubs. However, due to the collective nature

of LTM, seeding the set of privileged nodes in the non-interacting view does not guarantee the maximization of collective influence. The other proportion of spreaders with lower degree, although may be inefficient as single spreaders, are responsible for bridging the collective influence of hubs. With the help of both hubs and bridging weak-nodes, CI-TM achieves the best performance against other heuristic measures. The Internet network also exhibits similar phenomenon in Fig. 3(b). In this case, HDA and HD can only find 80% optimal influencers at the first-order transition of CI-TM algorithm, missing a substantial amount of nodes with lower degree but indispensable in integrating the collective influence of high-degree seeds. Such kind of weak-node effect has also been discovered for second-order transition in optimal percolation [12].

Although the performance of K-core can be improved by adaptive calculation in Fig. 2, for large-scale real networks, we do not display the result of KsA due to its $O(N^2)$ computational complexity and only show the curve of K-core. One cause for the unsatisfactory result of K-core is that it is not designed as a multiple spreaders finder since high K-core nodes tend to form densely connected clusters in the same shell, which prevents the expanding of information cascade. However, when finding individual spreaders, K-core is more optimal than other heuristics [9, 10].

In Appendix C, we further compare CI-TM algorithm with other methods, including Betweenness Centrality (BC), Closeness Centrality (CC) and Greedy Algorithm (GA). Results from regular, ER and SF networks prove that CI-TM algorithm outperforms computationally expensive BC, CC and GA. Furthermore, with a much higher efficiency, CI-TM achieves a comparable performance as the skillfully designed MP algorithm which is designed based on constraint satisfaction optimization. We should note that, MP can only find the specific configuration at the transition point, whereas CI-TM algorithm is capable of optimizing the influence for any given fraction of seeds q.

V. DISCUSSION

Identification of superspreaders in LTM has great practical implications in a wide range of dynamical processes. However, the complicated interactions among multiple spreaders prevent us from accurately locating the optimal influencers in LTM. Despite the existence of many heuristic strategies seeking to find optimal spreaders, they generally fail to take the collective effect of seeds into account, thus can only obtain suboptimal configurations. In this work, we propose a theoretical framework to analyze the collective influence of individuals in general LTM. By iteratively solving the linearized message passing equations, we decompose $\|\nu_{\rightarrow}\|$ into separate components, each of which corresponds to the contribution made by a single seed. Particularly, we find that the contribution of a seed is largely determined by its interplay with other nodes through subcritical paths. The subcritical path maximizing influence spreading in first-order transition is a generalization of the minimal path that defines the ball of influence in CI optimal percolation in second-order phase transitions. Thus we see that the concept of collective influence is quite universal, encompassing continuous and discontinuous transitions, including generalization to network of networks in the brain [46] and socio-economics [47]. In order to maximize the active population, we develop a scalable CI-TM algorithm. Our solution, rooted in the collective influence theory of general LTM, optimizes the active population by a fast adaptive scheme. The CI-TM algorithm, which is feasible for large-scale networks, outperforms other frequently used ranking strategies in synthetic random graphs and real-world networks. This provides a practical algorithm that can be employed in relevant applications such as viral marketing and information spreading in big-data analysis.

Appendix A: Relation to optimal percolation

In terms of the special choice of threshold $m_i = k_i - 1$ for each node with degree k_i, the set $P^{m_i}_{\partial i\backslash j}$ in Eq. (1) only has one combination. Therefore, Eq. (1) simplifies to

$$\nu_{i\to j} = n_i + (1 - n_i) \prod_{p\in\partial i\backslash j} \nu_{p\to i}. \tag{A1}$$

It has been proved that LTM with threshold $m_i = k_i - 1$ can be mapped to optimal percolation [12]. In the notation of optimal percolation, $\bar{\nu}_{i\to j}$ is defined as node i's probability NOT in active state when j is disconnected from the network, while $\bar{n}_i = 0$ if i is a seed and $\bar{n}_i = 1$ otherwise. The above definition of notation is exactly dual to what we define in Eq. (A1). Substituting $\nu_{i\to j} = 1 - \bar{\nu}_{i\to j}$ and $n_i = 1 - \bar{n}_i$ in Eq. (A1), through a simple rearrangement we have

$$\bar{\nu}_{i\to j} = \bar{n}_i[1 - \prod_{p\in\partial i\backslash j} (1 - \bar{\nu}_{p\to i})], \tag{A2}$$

which is exactly the message passing equation for optimal percolation in Ref. [12]. Therefore, our formula designed for LTM also contains the situation of second-order transition as a limiting case.

In the stability analysis of optimal percolation, a linearization around zero solution $\bar{\nu}_{\rightarrow} = \mathbf{0}$ is performed [12]. Mapping to our notation, the zero solution in optimal percolation corresponds to $\nu_{\rightarrow} = \mathbf{1}$. At this point, the condition $a_{k\to i,i\to j} = m_i - 1$, or equivalently $k_i - 2$, is naturally fulfilled for any two consecutive non-backtracking links $k \to i$ and $i \to j$, since all the $k_i - 2$ neighbors p excluding k and j have $\nu_{p\to i} = 1$. Therefore, $I_{k\to\ell,i\to j}$ recovers to the corresponding element of NB matrix $\mathcal{B}_{k\to\ell,i\to j}$ as defined in Ref. [12]. It should be

noticed that the stability analysis of the NB matrix via its largest eigenvalue developed in Ref. [12] cannot be applied to the general LTM since LTM presents a first-order phase transition where linear stability analysis is not valid.

Appendix B: Linearization of $G_{i\to j}$

The conventional method to linearize the nonlinear function $G_{i\to j}(\nu_\to)$ is Taylor expansion around the fixed point $\mathbf{0}$: $G_{i\to j}(\nu_\to) \approx G_{i\to j}(\mathbf{0}) + G'_{i\to j}(\mathbf{0})\nu_\to$. However, for our specific function $G_{i\to j}$, the gradient $G'_{i\to j}(\mathbf{0})$ is constantly $\mathbf{0}$ according to Eq. (5). To avoid the degeneracy, other linear approximation method should be applied.

For a differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ and $\mathbf{x}$, $\mathbf{y} \in \mathbb{R}^n$, the mean value theorem guarantees that there exists a real number $c \in (0, 1)$ such that $f(\mathbf{y}) - f(\mathbf{x}) = \nabla f((1-c)\mathbf{x}+c\mathbf{y})\cdot(\mathbf{y}-\mathbf{x})$. Here ∇ is the gradient and $\cdot$ denotes the dot product. Set $f = G_{i\to j}$, $\mathbf{x} = \mathbf{0}$, and $\mathbf{y} = \nu_\to$, we have $G_{i\to j}(\nu_\to) = G_{i\to j}(\mathbf{0}) + G'_{i\to j}(c\nu_\to)\nu_\to$. Notice that, if we set $c = 0$, the above equation becomes the classical linear Taylor expansion: $G_{i\to j}(\nu_\to) \approx G_{i\to j}(\mathbf{0}) + G'_{i\to j}(\mathbf{0})\nu_\to$, where the approximation accuracy is $O(|\nu_\to|^2)$ ($|\cdot|$ is the norm of vectors).

To deal with the degeneracy of $G'_{i\to j}(\mathbf{0})$, we approximate $G_{i\to j}(\nu_\to)$ by setting $c = 1$ for small $\nu_\to$: $G_{i\to j}(\nu_\to) \approx G_{i\to j}(\mathbf{0}) + G'_{i\to j}(\nu_\to)\nu_\to$. The approximation error can be calculated by $e = |G_{i\to j}(\nu_\to) - G_{i\to j}(\mathbf{0}) - G'_{i\to j}(\nu_\to)\nu_\to| \le |G'_{i\to j}(c\nu_\to) - G'_{i\to j}(\nu_\to)||\nu_\to|$. Recall that $G'_{i\to j} = (\cdots, \frac{\partial G_{i\to j}}{\partial \nu_{k\to \ell}}, \cdots)$. In a finite-size network, for a small $\nu_\to$ with elements $|\nu_{i\to j}| \le 1$, the gradient of each element $\frac{\partial G_{i\to j}}{\partial \nu_{k\to \ell}}$ is bounded according to Eq. (5). For all $2M$ elements, there exists a uniform upper bound for all the gradients $\nabla\frac{\partial G_{i\to j}}{\partial \nu_{k\to \ell}}$. Applying the mean value theorem to the differentiable function $\frac{\partial G_{i\to j}}{\partial \nu_{k\to \ell}}$, there should be a constant L such that $|\frac{\partial G_{i\to j}}{\partial \nu_{k\to \ell}}(c\nu_\to) - \frac{\partial G_{i\to j}}{\partial \nu_{k\to \ell}}(\nu_\to)| \le L|\nu_\to|$ for all the elements of $G'_{i\to j}$. Summing up all the elements in $G'_{i\to j}$, we conclude that $|G'_{i\to j}(c\nu_\to) - G'_{i\to j}(\nu_\to)| \le |(\cdots, L, \cdots)||\nu_\to|$. Therefore, the approximation error $e \le |(\cdots, L, \cdots)||\nu_\to|^2$. This proves that the accuracy of the linear approximation is $O(|\nu_\to|^2)$, which is same as the linear Taylor expansion.

Appendix C: More comparisons with competing methods

A growing number of methods aimed to rank nodes' influence in networks have been proposed in previous studies. Here we introduce some of the most widely used competing methods and perform a thorough comparison with CI-TM algorithm.

High degree (HD) Degree, defined as the number of connections attached to a node, is the most widely-used measure of influence [21]. In HD method, we rank nodes according to their degrees in a descending order, and sequentially select them as information sources. For HD method, the selected hubs intend to link with each other due to the assortative mixing property, making their influence areas overlap significantly. In this case, the selected seeds could rarely be optimal. High degree adaptive **(HDA)** is the adaptive version of HD method. After each removal, the degree of each node is recalculated. Such adaptive procedure can usually mitigate the overlapping and improve the performance of HD.

K-core (Ks) Through a k-shell decomposition process, K-core method assigns each node a k_S value to distinguish whether it locates in the core region or peripheral area [24]. In k-shell decomposition, nodes are iteratively removed from the network according to their current degrees. During the removal, all the nodes are classified into different k-shells. The K-core method selects nodes within high k-shells as the spreaders. In practice, single influential spreaders can be identified effectively by K-core ranking, which has been confirmed by both simulations and real-world data [9, 10, 25]. However, K-core ranking has the disadvantage of severe overlap of seeds' influence areas, and therefore performs poorly for multiple node selection. This limitation can be alleviated with an adaptive scheme where we recompute the K-core after each removal. Since there exists many nodes in the same k-shell, we select the node with the largest degree to further distinguish nodes within the highest k-shell. Such K-core adaptive **(KsA)** method can effectively enhance the performance of K-core.

PageRank (PR) PageRank is a popular ranking algorithm of webpages which was developed and used by search engine Google [23]. Over the years, PageRank has been adopted in many practical ranking problems. Generally speaking, PageRank measures a webpage's stationary visiting probability by a random walker following the hyperlinks in the network. As a special case of eigenvector centralities, PageRank evaluates the score of a node by taking into account its neighbors' scores. Even though such score-propagating mechanism works well for some purposes such as webpage ranking, an unfavorable consequence may be a heavy accumulation of scores near the high-degree nodes, specially for scale-free networks [38].

Greedy algorithm (GA) In GA, starting from an empty set of seeds, nodes with the maximal marginal gain are sequentially added to the seed set. Kempe *et al.* have proven that for a class of LTM with the attribute of submodularity, GA has a performance guarantee of $1-1/e \approx 63\%$, which means it could achieve at least 63% of real maximal influence [4]. Although guaranteed, this performance is still well sub-optimal. This result relies on the submodular property defined by a diminishing returns effect: the marginal gain from adding a node to the seed set S decreases with the size of S. It has been proven that several classes of LTM have submodular property, such as a random choice of thresholds. However, for LTM with a fixed threshold, it is not generally submodular. As

a consequence, GA is not guaranteed to provide a such approximation of the optimal spreaders for general LTM which is anyway quite far (63%) from the optimal q_c. Furthermore, the greedy search of GA requires massive simulations, which makes GA unscalable and thus limits its application in large-scale social networks.

Betweenness centrality (BC) BC quantifies the importance of node i in terms of the number of shortest paths cross through it [22]. Therefore, nodes with large BC usually occupy the pivotal positions in the shortest pathways connecting large numbers of nodes. In BC method, we select nodes with high BC scores as seeds. Although BC has been widely applied in social network analysis, its relatively high computational complexity makes BC prohibitive for large-scale networks. So far, the most efficient algorithm takes $O(MN)$ to calculate BC for a network with N nodes and M links [48], which is still not applicable to modern social networks with millions of users.

Closeness centrality (CC) Closeness centrality quantifies how close a node to other nodes in the network [49]. Formally, CC is defined as the reciprocal of the average shortest distance of a node to others in a network. Thus, nodes with high CC values tend to locate at the center of network clusters or communities. In CC method, we pick the seeds according to nodes' CC scores. Same as BC, CC also requires the heavy task of calculating all possible shortest paths. Thus the high computational cost of CC makes it infeasible for large-scale networks.

Message passing (MP) Recently, message passing (MP) algorithm has attracted much attention due to its successful application to network-related optimization problems, such as containing epidemic outbreaks [14] and optimizing spread dynamics [13]. In particular, F. Altarelli *et al.* proposed a message passing algorithm aimed to identify initial conditions to maximize the final number of active nodes in threshold models. Precisely, the trajectory of nodes' states is parametrized by a configuration $\mathbf{t} = \{t_i, 1 \leq i \leq N\}$ where $t_i \in \mathcal{T} = \{0, 1, \cdots, T, \infty\}$ is the activation time of node i ($t_i = \infty$ if inactive). By mapping the optimization onto a constraint satisfaction problem, an energy-minimizing algorithm based on the cavity method of statistical physics is proposed. In the algorithm, a convolution process is employed to compute the Max-Sum updates. The technical details of the derivation and implementation of the MP algorithm can be found in Ref. [13]. In most cases MP algorithm does not converge, then a reinforcement strategy is implemented [13]. By imposing an external field slowly increasing over time with a growth rate r, the system is forced to converge to a higher energy, which increases with r. In addition, it requires $O(r^{-1})$ iterations to reach convergence. In practice, in order to acquire a good approximation, r should be set small, usually of order $O(10^{-4})$ or less. This causes an increase of computational burden of MP algorithm. For a node of degree k and threshold m_i, each update takes $O(Tk(k-1)m_i^2)$ operations [13]. Pre-computing the convolution can further save a factor of $k-1$. Considering the updates of all N nodes for $O(r^{-1})$ iterations, the overall complexity of MP is $O(T\sum_i k_i m_i^2/r)$. Therefore, the time complexity of MP depends on both the degree distribution of networks and the choice of threshold.

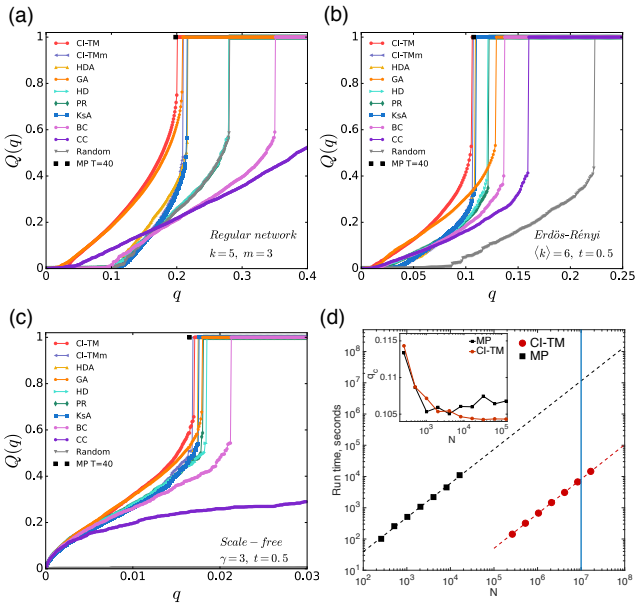

FIG. A1. Comparison of competing methods. Performance of different methods for (a), regular network ($N = 10^4$, $k = 5$, $m = 3$) (b), ER network ($N = 10^4$, $\langle k\rangle = 6$, $t = 0.5$) (c), and scale-free network ($N = 10^4$, $\gamma = 3$, $t = 0.5$). In addition to the methods we have examined in the main text, we also display the results of Greedy Algorithm (GA), Betweenness Centrality (BC), Closeness Centrality (CC) and Message Passing (MP). We set the parameter $T = 40$ in MP and implement reinforcement with parameter $r = 10^{-4}$. (d), Comparison of run time for MP ($T = 40$, $r = 10^{-4}$) and CI-TM ($L = 3$) on ER networks ($\langle k\rangle = 6$, $t = 0.5$). The dashed lines are power-law fitting, and the vertical line indicates network size $N = 10^7$. The inset presents the critical value q_c for MP and CI-TM with increasing network size N.

In Fig. A1(a)-(c), we provide the thorough comparisons of different methods on regular, ER and SF networks ($N = 10^4$), including MP algorithm and computationally expensive methods GA, BC and CC. We set $T = 40$ and a reinforcement parameter $r = 1 \times 10^{-4}$ in MP algorithm. For all the cases, our method outperforms other heuristic ranking strategies and has a comparable performance of MP algorithm. We display the comparison of run time for MP and CI-TM for ER networks with $\langle k\rangle = 6$ and threshold $t = 0.5$ as a function of N in Fig. A1(d). CI-TM algorithm with $L = 3$ can complete the calculation in about 2.4 hours for $N = 10^7$, while MP with $T = 40$ and $r = 10^{-4}$ will approximately require 132.7 days. Moreover, as N increases, the critical value q_c of MP shows a growing trend for the same parameters $T = 40$ and $r = 10^{-4}$, as shown in the inset of Fig. A1(d). This implies that in order to converge to a smaller seed

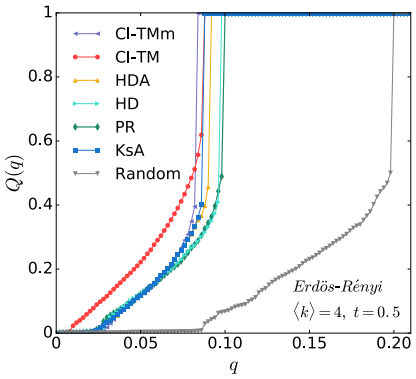

FIG. A2. Performance of modified CI-TM on ER networks. For ER networks ($N = 2 \times 10^5$, $\langle k \rangle = 4$, $t = 0.5$), the performance of CI-TMm surpasses CI-TM by excluding the fragmented vulnerable clusters in the adaptive calculation. Although $Q(q)$ for CI-TMm is lower at first, it exceeds other methods near the critical point and achieves the earliest first-order transition.

set for large N, MP requires a larger T and a smaller reinforcement parameter r, which in turn increases its computational complexity to prohibited limits.

Appendix D: Modified CI-TM algorithm

The CI-TM algorithm is essentially a greedy approach based on CI-TM values. The success of CI-TM algorithm depends on whether the currently selected seed has potentials to create more subcritical nodes that are helpful for the early formation of giant subcritical cluster. In LTM, there exists a special case of subcritical nodes with threshold $m = 1$, which is defined as vulnerable vertices in previous literature [30, 31]. Different from general subcritical nodes, vulnerable vertices are naturally subcritical since they have threshold $m = 1$ and do not rely on the states of others. During the calculation, a node of degree k becomes vulnerable when its $m - 1$ neighbors are removed, leaving $k-m+1$ links in the network. For ER networks with a low average degree, the limited number of remaining links of vulnerable vertices could only form fragmented clusters. In this case, CI-TM would bias to nodes connected to large numbers of small vulnerable clusters, such as a peripheral hub linked to numerous leaf nodes. Unfortunately, the activation of such small clusters provides little help to the formation of giant subcritical cluster. Because the scattered vulnerable clusters have very few links connected to existing non-subcritical nodes, their activations are not effective in producing subsequent subcritical nodes. Moreover, once global cascade appears, these fragmented vulnerable clusters will be activated subsequently, without additional seeds. In this case, we heuristically propose a modified CI-TM (CI-TMm) algorithm by excluding vulnerable nodes in the calculation of CI-TM value. The performance of CI-TMm algorithm is displayed in Fig.A2 for ER networks with an average degree $\langle k \rangle = 4$. The critical value q_c for CI-TMm is advanced compared to CI-TM algorithm. However, before the first-order transition, CI-TMm cannot optimize the spreading and has substantially lower $Q(q)$ than CI-TM. We should note that, CI-TMm presents a lower q_c only in the situation of fragmented vulnerable clusters. For ER networks with higher average degrees (e.g., $\langle k \rangle = 6$) where relatively large vulnerable clusters emerge, CI-TM still predicts earlier first-order transition.

ACKNOWLEDGMENTS

This work was supported by NIH-NIGMS 1R21GM107641, NSF-PoLS PHY-1305476 and ARL Cooperative Agreement Number W911NF-09-2-0053, the ARL Network Science CTA (to H.A.M.), as well as US NIH grant GM110748 and the Defense Threat Reduction Agency contract HDTRA1-15-C-0018 (to J.S.).

[1] S.V. Buldyrev, R. Parshani, G. Paul, H.E. Stanley, and S. Havlin, *Catastrophic cascade of failures in interdependent networks,* Nature (London) **464,** 1025 (2010).
[2] D.J. Watts and J. Peretti, *Viral marketing for the real world,* Harvard Business Review 104 (May 2007).
[3] E.M. Rogers, *Diffusion of Innovation* (Free Press, New York, 1995).
[4] D. Kempe, J. Kleinberg, and É . Tardos, *Maximizing the spread of influence in a social network,* Proc. 9th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining 137 (2003).
[5] P. Domingos and M. Richardson, *Mining the network value of customers,* Proc. 7th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining 57 (2001).
[6] T.W. Valente and R.L. Davis, *Accelerating the diffusion of innovations using opinion leaders,* Ann. Am. Acad. Polit. Soc. Sci. **556,** 55 (1999).
[7] A. Galeotti and S. Goyal, *Influencing the influencers: a theory of strategic diffusion,* The RAND J. Econ. **40,** 509 (2009).
[8] D.J. Watts and P.S. Dodds *Influentials, networks, and public opinion formation,* J. Consum. Res. **34,** 441 (2007).
[9] S. Pei, L. Muchnik, J.S. Andrade Jr, Z. Zheng, and H.A. Makse, *Searching for superspreaders of information in real-world social media,* Sci. Rep. **4** 5547 (2014).
[10] M. Kitsak, L.K. Gallos, S. Havlin, F. Liljeros, L. Muchnik, H.E. Stanley, and H.A. Makse, *Identification of influential spreaders in complex networks,* Nature Phys. **6,**

888 (2010).
[11] J. Leskovec, A. Krause, C. Guestrin, C. Faloutsos, J. VanBriesen, and N. Glance, *Cost-effective outbreak detection in networks,* Proc. 13th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining 420 (2007).
[12] F. Morone and H.A. Makse, *Influence maximization in complex networks through optimal percolation,* Nature (London) **524,** 65 (2015).
[13] F. Altarelli, A. Braunstein, L. DallAsta, and R. Zecchina, *Optimizing spread dynamics on graphs by message passing,* J. Stat. Mech. **9,** P09011 (2013).
[14] F. Altarelli, A. Braunstein, L. DallAsta, J.R. Wakeling, and R. Zecchina, *Containing epidemic outbreaks by message-passing techniques,* Phys. Rev. X **4,** 021024 (2014).
[15] W. Chen, Y. Yuan, and L. Zhang, *Scalable influence maximization in social networks under the linear threshold model,* Proc. IEEE 10th Int. Conf. on Data Mining (ICDM) 88 (2010).
[16] W. Chen, C. Wang, and Y. Wang, *Scalable influence maximization for prevalent viral marketing in large-scale social networks,* Proc. 16th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining 1029 (2010).
[17] W. Chen, Y. Wang, and S. Yang, *Efficient influence maximization in social networks,* Proc. 15th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining 199 (2009).
[18] S. Mugisha and H.J. Zhou, *Identifying optimal targets of network attack by belief propagation,* arXiv preprint arXiv:1603.05781 (2016).
[19] A. Braunstein, L. Dall'Asta, G. Semerjian, and L. Zdeborová, *Network dismantling,* arXiv preprint arXiv:1603.08883 (2016).
[20] F. Morone, B. Min, L. Bo, R. Mari, and H.A. Makse, *Collective Influence Algorithm to find influencers via optimal percolation in massively large social media,* arXiv preprint arXiv:1603.08273 (2016).
[21] R. Albert, H. Jeong, and A.-L. Barabási, *Error and attack tolerance of complex networks,* Nature (London) **406,** 378 (2000).
[22] L.C. Freeman, *Centrality in social networks: Conceptual clarification,* Soc. Netw. **1,** 215 (1979).
[23] S. Brin and L. Page, *The anatomy of a large-scale hypertextual web search engine,* Computer Networks and ISDN Systems **30,** 107 (1998).
[24] S.B. Seidman, *Network structure and minimum degree,* Soc. Netw. **5,** 269 (1983).
[25] S. Pei and H.A. Makse, *Spreading dynamics in complex networks,* J. Stat. Mech. **12,** P12002 (2013).
[26] M. Granovetter, *Threshold models of collective behavior,* Am. J. Sociol. **83,** 1420 (1978).
[27] T.C. Schelling, *Micromotives and macrobehavior* (Norton, New York, 1978).
[28] T.W. Valente, *Network Models of the Diffusion of Innovations* (Hampton Press, Cresskill, NJ, 1995).
[29] J. Kleinberg, *Cascading behavior in networks: Algorithmic and economic issues,* Algorithmic Game Theory **24,** 613 (2007).
[30] D.J. Watts, *A simple model of global cascades on random networks,* Proc. Natl. Acad. Sci. USA **99,** 5766 (2002).
[31] M. Ramos, J. Shao, S.D. Reis, C. Anteneodo, J.S. Andrade Jr, S. Havlin, and H.A. Makse, *How does public opinion become extreme?* Sci. Rep. **5** 10032 (2015).
[32] G.J. Baxter, S.N. Dorogovtsev, A.V. Goltsev, and J.F. Mendes, *Bootstrap percolation on complex networks,* Phys. Rev. E **82**, 011103 (2010).
[33] A.V. Goltsev, S.N. Dorogovtsev, and J.F.F. Mendes, *k-core (bootstrap) percolation on complex networks: Critical phenomena and nonlocal effects,* Phys. Rev. E **73,** 056101 (2006).
[34] S.N. Dorogovtsev, A.V. Goltsev, and J.F.F. Mendes, *k-core architecture and k-core percolation on complex networks,* Physica D **224,** 7 (2006).
[35] J.M. Schwarz, A.J. Liu, and L.Q. Chayes, *The onset of jamming as the sudden emergence of an infinite k-core cluster,* Europhys. Lett. **73,** 560 (2006).
[36] M. Mézard and G. Parisi, *The cavity method at zero temperature,* J. Stat. Phys. **111**, 1 (2003).
[37] K. Hashimoto, *Zeta functions of finite graphs and representations of p-adic groups,* Adv. Stud. Pure Math. **15,** 211 (1989).
[38] T. Martin, X. Zhang, X, and M.E.J. Newman, *Localization and centrality in networks,* Phys. Rev. E **90,** 052808 (2014).
[39] F. Krzakala, C. Moore, E. Mossel, J. Neeman, A. Sly, L. Zdeborová, and P. Zhang, *Spectral redemption in clustering sparse networks,* Proc. Natl. Acad. Sci. USA **110,** 20935 (2013).
[40] B. Karrer, M.E.J. Newman, and L. Zdeborová, *Percolation on sparse networks,* Phys. Rev. Lett. **113,** 208702 (2014).
[41] M.E.J. Newman, *Spectral methods for community detection and graph partitioning,* Phys. Rev. E **88,** 042822 (2013).
[42] F. Radicchi, *Predicting percolation thresholds in networks,* Phys. Rev. E **91,** 010801(R) (2015).
[43] M. Molloy and B. Reed, *A critical point for random graphs with a given degree sequence,* Random Structures & Algorithms **6,** 161 (1995).
[44] J. Leskovec, D.P. Huttenlocher, and J.M. Kleinberg, *Governance in social media: A case study of the wikipedia promotion process,* Proc. 4th AAAI Int. Conf. on Weblogs and Social Media 98 (2010).
[45] J. Leskovec, J. Kleinberg, and C. Faloutsos, *Graphs over time: densification laws, shrinking diameters and possible explanations,* Proc. 11th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining 177 (2005).
[46] F. Morone, K. Roth, B Min, and H.A. Makse, *Robust network of networks inspired by a model of brain activation,* arXiv preprint arXiv:1602.06238 (2016).
[47] S. Luo, F. Morone, C. Sarraute, and H.A. Makse, *Inferring personal financial status from social network location,* arXiv preprint (2016).
[48] U. Brandes, *A faster algorithm for betweenness centrality,* J. Math. Sociol. **25,** 163 (2001).
[49] A. Bavelas, *Communication patterns in tasks oriented groups,* J. Acoust. Soc. Am. **22,** 271 (1950).