

Spectral Bounds for Independent Cascade Model with Sensitive Edges

Eun Jee Lee*, Sudeep Kamath[†], Emmanuel Abbe^{*†}, Sanjeev R. Kulkarni[†]

^{*}Program in Applied and Computational Mathematics

[†]Department of Electrical Engineering
Princeton University

{ejlee, sukamath, eabbe, kulkarni}@princeton.edu

Abstract—This paper studies independent cascade models where influence propagates from seed-nodes along edges with independent probabilities. Upper-bounds for the expected number of influenced nodes were previously proposed using the spectral norm of a Hazard matrix. However, these bounds turn out loose in many cases, in particular with respect to sensitive edges such as bottlenecks, seed adjacent, and high probability edges. This paper proposes a similar bound that improves in such cases by handling sensitive edges more carefully.

I. INTRODUCTION

There is a broad range of applications for influence propagation which include determining a collective behavior on binary decisions [1], [2], [3], finding an effective strategy to vaccinate the disease [4], computing the transmission rate of a mobile ad-hoc network [5], studying the importance of individuals in spreading new ideas or products [6], characterizing the information diffusion in social network services [7], and studying the process for viral marketing [8].

Among various research questions related to the influence propagation, influence maximization problem—given a social network whose edges have the probabilities of influence between nodes, finding the top k seed nodes that maximizes the expected number of influenced nodes—has been widely studied. Kempe et al., [9] proposed a heuristic algorithm for this problem and many following studies also proposed various algorithms with improved accuracy and performance [10], [11], [12], [13].

One notable work taking a theoretical approach is the work [14] which finds the expected number of influenced nodes at the end of propagation (the influence) by proving an equivalence between bond percolation and a special case of Susceptible Infective Removed model [15] where all disease-transmission probabilities are equal. More recent theoretical works have also been focusing on estimation of the influence when a set of seed influencing nodes are given. Draief et al., [16] introduced an upper bound for the influence and tighter bounds were later suggested [17].

This paper shares the goal with mentioned works. We improve the upper bound for the influence in [17] by accounting the importance of bottleneck in the network. The proposed bound gives significant improvement, in particular,

for stochastic block models and networks with small degrees. In addition, the result is supported by empirical data.

II. STOCHASTIC DIFFUSION MODELS

We briefly introduce the models commonly used in influence propagation, namely, stochastic diffusion models. Consider a network as a directed graph $G = (V, E)$ whose edges are assigned transmission probabilities, i.e. the probability that one end of the edge influences the other end. The influence propagation starts at the set $S_0 \subseteq V$ of seeds, the initially informed nodes. Then, the influence spreads with a random process, which defines different types of stochastic diffusion models. We next define two standard stochastic diffusion models.

Definition 1 (Independent cascade model). *The model was defined in [18]. Let $G = (V, E)$ where $|V| = n$, \mathcal{P} be the $n \times n$ transmission probability matrix, and $S_0 \subseteq V$ be the set of initially influenced nodes. Then, the independent cascade model $IC(G, \mathcal{P}, S_0)$ generates the active set S_t for any discrete time $t \geq 1$ by the following process. First, set $S_t = S_{t-1}$ and for every edge $(i, j) \in \{(u, v) \in E | u \in S_{t-1} \setminus S_{t-2}, v \in V \setminus S_{t-1}\}$, i activates j with probability \mathcal{P}_{ij} . If j becomes activated, add j to S_t . The cascade stops if $S_T = S_{T+1}$ at the end of the activation process at time T .*

In [18], it is proven that the independent cascade model is equivalent to the following model, live-arc graph model with independent arc selection.

Definition 2 (Live-arc graph model with independent arc selection). *The model was defined in [18]. Let $G = (V, E)$ where $|V| = n$, \mathcal{P} be the $n \times n$ transmission probability matrix, and $S_0 \subseteq V$ be the set of initially influenced nodes. The live-arc graph model with independent arc selection $LI(G, \mathcal{P}, S_0)$ generates the set of live edges E_L as follows. For each $(i, j) \in E$, activate (i, j) with probability \mathcal{P}_{ij} . If (i, j) becomes activated, add (i, j) to E_L . Let $G_L = (V, E_L)$ be the subgraph of G . Then, the active set S is the set of all nodes in V that are reachable by S_0 in G_L .*

Notations. For any live-arc graph model with independent arc selection $LI(G, \mathcal{P}, S_0)$, we use random variable X_i as the

state of node $i \in V$ at the end of the propagation, i.e. $X_i = 1$ if node i is influenced, $X_i = 0$ otherwise, and E_{ij} as the state of edge (i, j) at the end of the propagation, i.e. $E_{ij} = 1$ (active) if the influence can be spread through the edge (i, j) , $E_{ij} = 0$ (inactive) otherwise. Notice that E_{ij} are independent Bernoulli random variables $\mathcal{B}(p_{ij})$. The number of nodes that are influenced at the end of propagation is defined as $N(S_0) = \sum_{i \in V \setminus S_0} \mathbf{1}(X_i = 1) + |S_0|$. The expected number of nodes that are influenced at the end of the propagation is defined as $\sigma_{\mathcal{N}}(S_0) = \mathbb{E}[N(S_0)]$ and called *the influence of the seeded network* $\mathcal{N} = LI(G, \mathcal{P}, S_0)$.

III. EXISTING BOUNDS

Since a node is influenced if at least one of its neighbors and the edge between them are both active, we have, for all $i \in V \setminus S_0$,

$$X_i = 1 - \prod_j (1 - X_j E_{ij}). \quad (1)$$

Recently, Lemonnier et al. proposed an upper bound for the influence $\sigma_{\mathcal{N}}(S_0)$.

Definition 3. [17] For any seeded network $\mathcal{N} = LI(G, \mathcal{P}, S_0)$, the Hazard matrix \mathcal{H} and $\mathcal{H}(S_0)$ are defined as

$$\mathcal{H}_{ij} = \begin{cases} -\ln(1 - p_{ij}) & \text{if } (i, j) \in E \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

and

$$\mathcal{H}_{ij}(S_0) = \mathbf{1}_{j \notin S_0} \mathcal{H}_{ij}. \quad (3)$$

Theorem 1. [17] For any seeded network $\mathcal{N} = LI(G, \mathcal{P}, S_0)$ with $|S_0| = n_0 < n$, let $\rho(S_0)$ be the spectral radius of $\frac{\mathcal{H}(S_0) + \mathcal{H}(S_0)^T}{2}$. Then, the expected number of nodes influenced by the seeds satisfies

$$\sigma_{\mathcal{N}}(S_0) \leq n_0 + \gamma(n - n_0), \quad (4)$$

where γ is the smallest solution in $[0, 1]$ of

$$\gamma - 1 + \exp\left(-\rho(S_0)\gamma - \frac{\rho(S_0)n_0}{\gamma(n - n_0)}\right) = 0. \quad (5)$$

IV. IMPROVED BOUNDS

Note that the bound in Theorem 1 accounts for all solutions of the equation (1). However, some solutions are implausible. Consider a graph with two cliques C_1 and C_2 , with node sets V_1 and V_2 respectively, where the cliques are connected to each other by a single edge E_c and the seeds are on the clique C_1 . Let all transmission probabilities are 1 except for E_c where $p_c \neq \{0, 1\}$. When $E_c = 0$, $X_i = 1, \forall i \in V$ is a solution to equation (1) as well as the correct solution $X_i = 1, \forall i \in V_1$ and $X_j = 0, \forall j \in V_2$. Hence, applying the upper bound on graphs with small cuts may result in substantial difference between the bound and the expectation.

Also, the bound does not consider for the fact that the edges near the seeds have more significant impact on the propagation. For example, in case of edges that are adjacent to the seeds being inactive, the only correct solution is

$X_i = 0, \forall i \in V \setminus S_0$, but the equations in (1) also have other solutions which lead to a looser bound.

Finally, using the Hazard matrix in Definition 3 when at least one edge has transmission probability arbitrarily close to 1 gives a trivial bound, n , since $\rho(S_0)$ becomes infinite. In this section, we propose a tighter upper bound to improve these limitations.

A. Conditioning on a set of sensitive edges

We select a subset $S \subseteq E$ of edges and define the set of random variables $E[S] = \{E_{ij} | (i, j) \in S\}$. Note that

$$\mathbb{E}[X_i] = \sum_{s \in \{0, 1\}^{|S|}} \mathbb{E}[X_i | E[S] = s] \mathbb{P}(E[S] = s), \quad (6)$$

and

$$\sigma_{\mathcal{N}}(S_0) = \sum_{s \in \{0, 1\}^{|S|}} \sigma_s(S_0) \mathbb{P}(E[S] = s), \quad (7)$$

where $\sigma_s(S_0) = \sum_i \mathbb{E}[X_i | E[S] = s]$. Hence, we may compute upper bounds for $\sum_i \mathbb{E}[X_i | E[S] = s]$ for all $s \in \{0, 1\}^{|S|}$ and average such bounds to find an upper bound for $\sigma_{\mathcal{N}}(S_0)$.

Notice that if $S = E$, $\mathbb{E}[X_i | E[S] = s]$ is a constant given the realization of the graph, but conditioning on E requires $2^{|E|}$ computations. Thus, we select *sensitive* edges that potentially create a large difference between the bound and the expectation, such as edges in a bottleneck (small cut), adjacent to the seeds, or having high transmission probabilities.

B. Obtaining hazard matrix of conditioned graph

Condition $E[S] = s$ results in fixing the edge variables in $E[S]$ to either 0 or 1 according to s . Hence, the condition is equivalent to replacing the transmission probability matrix with \mathcal{P}' where $\mathcal{P}'_e = s_e, \forall e \in S$ and $\mathcal{P}'_{e'} = \mathcal{P}_{e'}, \forall e' \in E \setminus S$. However, as explained above, transmission probability 1 results in a trivial bound, $\sigma_{\mathcal{N}}(S_0) \leq n$. Therefore, we modify the graph according to the conditioning, i.e. removing edges when they are set to inactive in the conditioning and contracting edges when they are set to active in the conditioning.

Definition 4. For any seeded network $\mathcal{N} = LI(G, \mathcal{P}, S_0)$, a subset $S \subseteq E$, and a realization s of $E[S]$, the conditioned seeded network $\mathcal{N}_s = LI(G_s, \mathcal{P}_s, S_s)$ and the hazard matrix \mathcal{H}'_s are defined as follows.

- Let $\mathcal{L} = \{L_1, \dots, L_m\}$ where $L_i \subseteq V$ is a disjoint set of nodes that are merged together due to the contractions. For each $i \in [m]$, denote u_i as the node obtained by merging all the nodes in L_i .
- $G_s = (V_s, E_s)$, where G_s is obtained from G by contracting the active edges and removing the inactive edges in S according the conditioning s , i.e. $V_s = (V \setminus \sqcup_{i=1}^m L_i) \sqcup \{u_i\}_{i=1}^m$, and let $|V_s| = n_s$.
- Let $S_u = \{u_i \in \{u_i\}_{i=1}^m | S_0 \cap L_i \neq \emptyset\}$, the set of nodes that are obtained by merging with at least one seed. Then, $S_s = (S_0 \setminus \sqcup_{i=1}^m L_i) \sqcup S_u$, and let $|S_s| = m_s$.
- \mathcal{P}_s is obtained by computing the transmission probabilities of the edges in E_s .

- The hazard matrix \mathcal{H}'_s is defined as in Definition 3.

$$(\mathcal{H}'_s)_{ij} = \begin{cases} -\ln(1 - (\mathcal{P}_s)_{ij}) & \text{if } (i, j) \in E_s \\ 0 & \text{otherwise.} \end{cases}$$

Note that $\mathbb{E}[X_i|E[S] = s] = \mathbb{E}_{G_s}[X_{i_s}]$, where i_s is the node resulting from merging i with other nodes (possibly none) according to the conditioning s . Applying the bound from [17] on G_s , for all $i \in V_s \setminus S_s$,

$$\mathbb{E}_{G_s}[X_i] \leq 1 - \exp\left(\sum_{j \in V_s} -(\mathcal{H}'_s)_{ji} \mathbb{E}_{G_s}[X_j]\right). \quad (8)$$

Definition 5. For any seeded network $\mathcal{N} = LI(G, \mathcal{P}, S_0)$ and the conditioned seeded network $\mathcal{N}_s = LI(G_s, \mathcal{P}_s, S_s)$, define the vector $\mathbf{a} = (\mathbf{a}_i)_{i \in V_s}$, where $\mathbf{a}_i = k$ represents that node i in G_s results from merging k nodes in G , and call this the count vector.

Then,

$$\sum_{i \in V} \mathbb{E}[X_i|E[S] = s] = \sum_{i_s \in V_s} \mathbf{a}_{i_s} \mathbb{E}_{G_s}[X_{i_s}], \quad (9)$$

since when i_s is active, corresponding \mathbf{a}_{i_s} nodes in G are active.

Definition 6. The modified hazard matrix \mathcal{H}_s of the conditioned graph G_s with count vector \mathbf{a} is obtained from \mathcal{H}'_s by multiplying $\frac{1}{\sqrt{\mathbf{a}_i}}$ to row i and multiplying $\sqrt{\mathbf{a}_i}$ to column i for every $i \in V_s$.

C. Improved bound

Theorem 2. In seeded network $LI(G, \mathcal{P}, S_0)$, for any set S of edges, the expected number of nodes influenced by the seeds satisfies

$$\sigma(S_0) \leq \sum_s (\gamma_s(n - n_s) + n_s) \mathbb{P}(E[S] = s). \quad (10)$$

where γ_s is the smallest solution in $[0, 1]$ of

$$x - 1 + \exp\left(-\rho(S_s)x - \frac{\rho(S_s)m_s}{x(n - m_s)}\right) = 0, \quad (11)$$

and $\rho(S_s)$ is defined as

$$\rho(S_s) = \frac{\mathcal{H}_s(S_s) + \mathcal{H}_s(S_s)^\top}{2} \quad (12)$$

where

$$\mathcal{H}_s(S_s)_{ij} = \mathbf{1}_{j \notin S_s} \{\mathcal{H}_s\}_{ij}. \quad (13)$$

Proof. For convenience, let $V_s = [n_s] = \{1, \dots, n_s\}$. Define $Z_i = \sqrt{\mathbf{a}_i} \mathbb{E}_{G_s}[X_i]$, $Z = (Z_i)_{i \in [n_s]}$ and $u = (\mathbf{1}_{i \notin S_s} Z_i)_{i \in [n_s]}$. Using (8),

$$\begin{aligned} u^\top Z &= \sum_{i=1}^{n_s} u_i \sqrt{\mathbf{a}_i} \mathbb{E}_{G_s}[X_i] \\ &= \sum_{i=1}^{n_s} u_i \sqrt{\mathbf{a}_i} \left(1 - \exp\left(\sum_{j \in [n_s]} -(\mathcal{H}'_s)_{ji} \mathbb{E}_{G_s}[X_j]\right)\right). \end{aligned}$$

Let $w = (\sqrt{\mathbf{a}_i} u_i)_{i \in [n_s]}$. Then,

$$\begin{aligned} u^\top Z &= \sum_{i=1}^{n_s} w_i \left(1 - \exp\left(\sum_{j \in [n_s]} -(\mathcal{H}'_s)_{ji} \mathbb{E}_{G_s}[X_j]\right)\right) \\ &= |w|_1 \left(1 - \sum_{i=1}^{n_s} \frac{w_i}{|w|_1} \exp\left(-\mathcal{H}'_s{}^\top \left(\frac{1}{\sqrt{\mathbf{a}_i}} Z\right)_{i \in [n_s]}\right)\right). \end{aligned}$$

Let \mathcal{H}''_s be the matrix obtained by multiplying $\frac{1}{\sqrt{\mathbf{a}_i}}$ to i -th row of \mathcal{H}'_s for all rows. Then,

$$u^\top Z = |w|_1 \left(1 - \sum_{i=1}^{n_s} \frac{w_i}{|w|_1} \exp\left(-\mathcal{H}''_s{}^\top Z_i\right)\right) \quad (14)$$

$$\leq |w|_1 \left(1 - \exp\left(-\frac{Z^\top \mathcal{H}''_s w}{|w|_1}\right)\right) \quad (15)$$

Define \mathcal{H}_s be the matrix obtained by multiplying $\sqrt{\mathbf{a}_i}$ to i -th column of \mathcal{H}''_s for all columns, $\mathcal{H}_s(S_s)$ be the matrix obtained by zero-out the columns whose indices are in S_s and $b = (\mathbf{1}_{i \notin S_s} \sqrt{\mathbf{a}_i})_{i \in [n_s]}$. Then,

$$u^\top Z \leq |w|_1 \left(1 - \exp\left(-\frac{Z^\top \mathcal{H}_s(A_s) Z}{|w|_1}\right)\right) \quad (16)$$

and by Cauchy-Schwarz inequality, $\langle u, b \rangle^2 \leq \langle u, u \rangle \langle b, b \rangle$. Therefore,

$$|w|_1^2 \leq (Z^\top Z - m_s)(n - m_s), \quad (17)$$

where $m_s = |S_s|$. The rest of the proof is a repeat of the one from [17]. Let $\rho_s(A_s)$ be the spectral radius of $\frac{\mathcal{H}_s(A_s) + \mathcal{H}_s(A_s)^\top}{2}$ and $y = \frac{u^\top Z}{|w|_1} = \frac{Z^\top Z - m_s}{|w|_1}$. Then, by (17),

$$|w|_1 \leq y(n - m_s). \quad (18)$$

Then,

$$y \leq 1 - \exp\left(-\rho_s(S_s)y - \frac{\rho_s(S_s)m_s}{|w|_1}\right) \quad (19)$$

$$\leq 1 - \exp\left(-\rho_s(S_s)y - \frac{\rho_s(S_s)m_s}{y(n - m_s)}\right). \quad (20)$$

Consider a solution γ_s in $(0, 1)$ of the equation

$$x = 1 - \exp\left(-\rho_s(S_s)x - \frac{\rho_s(S_s)m_s}{x(n - m_s)}\right). \quad (21)$$

Notice that $|w|_1 \leq \gamma_s(n - m_s)$ and $|w|_1 = \sum_{i \notin S_s} \mathbf{a}_i \mathbb{E}_{G_s}[X_i]$. Thus,

$$\sigma_s(A) = |w|_1 + m_s \quad (22)$$

$$\leq \gamma_s(n - m_s) + m_s \quad (23)$$

and

$$\sigma(A) \leq \sum_s (\gamma_s(n - m_s) + m_s) \mathbb{P}(E[S] = s). \quad (24)$$

□

It is important to compute γ_s over connected graphs when condition s results in creating more than one connected components.

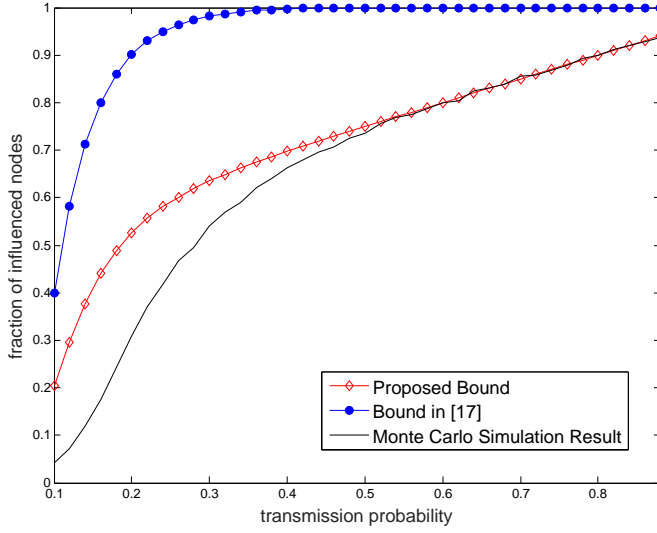


Fig. 1. The proposed bound conditioning on the bottleneck edge, the bound in [17], and the Monte Carlo simulation result for the expected fraction of influenced nodes in the network with a bottleneck.

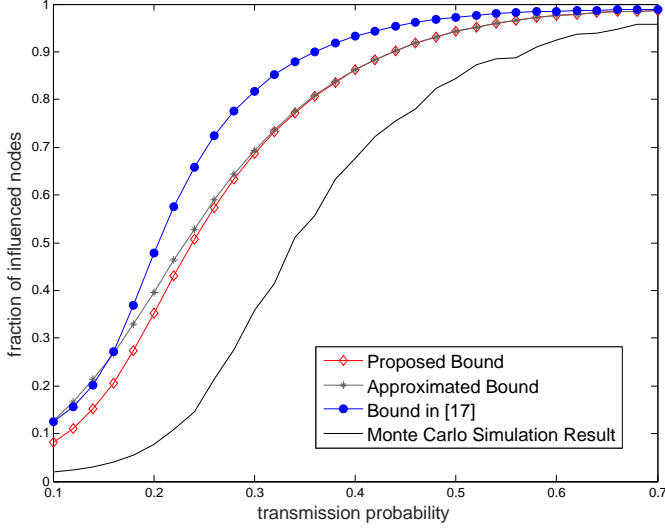


Fig. 2. The proposed bound and the approximated bound conditioning on the seed adjacent edges, the bound in [17], and the Monte Carlo simulation result for the expected fraction of influenced nodes in the Erdős-Rényi network.

V. EXPERIMENTAL RESULTS

We show that the bounds in Theorem 2 have significant improvement in the presence of sensitive edges.

In Figure 1, we consider a network $\mathcal{N}_1 = LI(G_1, \mathcal{P}_1, S_1)$, where S_1 contains a single seed v_1 , randomly selected from $V(G_1)$, and all edges have the same transmission probabilities. The graph G_1 consists of two Erdős-Rényi random graphs with 100 nodes where edges are drawn with probability 0.2, and the two graphs are connected to each other by a single edge e_c . Notice that the edge e_c (bottleneck edge) has a significant impact on the influence. Figure 1 compares the bounds given in Theorem 2 obtained by conditioning on the edge e_c to the

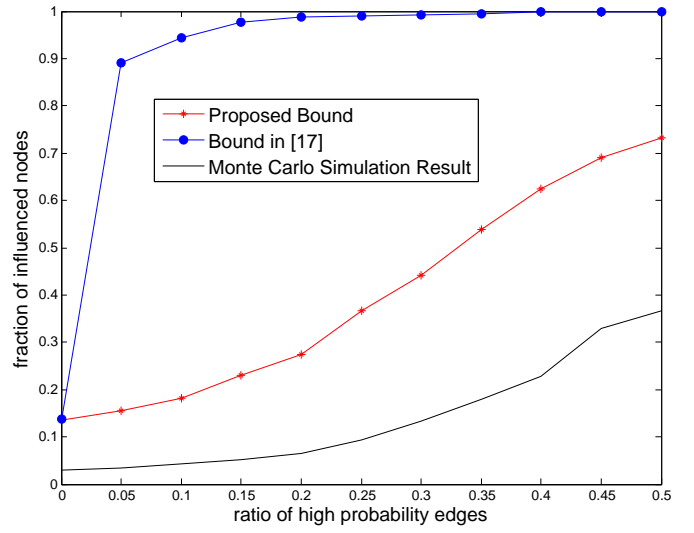


Fig. 3. The proposed approximated bound conditioning on the high transmission probability edges, the bound in [17], and the Monte Carlo simulation result for the expected fraction of influenced nodes in the Erdős-Rényi network.

existing bounds from [17], for the transmission probability ranging from 0.1 to 0.9. We also provide the expected fraction of influenced nodes obtained by Monte Carlo simulation. The result shows that the bound of Theorem 2 is tighter than the existing bound introduced in [17].

Figure 3 considers a network $\mathcal{N}_2 = LI(G_2, \mathcal{P}_2, S_2)$, where S_2 contains a single seed v_2 , randomly selected from $V(G_2)$, and all edges have the same transmission probabilities. The graph G_2 is an Erdős-Rényi random graph with 100 nodes where edges are drawn with probability 0.05. Figure 3 compares the bounds given in Theorem 2 obtained by conditioning on the seed adjacent edges to the existing bounds from [17], for the transmission probability ranging from 0.1 to 0.7. Since the bounds in Theorem 2 are often more computationally expensive than the bounds in [17], we also include an approximated version of the bounds. The approximated bounds take the conditioning s from $\{0^{|S|}, 1^{|S|}\}$, i.e. one case where all edges in S are active and the other where all edges in S are inactive, and assign probability $1 - \prod_{e \in S} (1 - p_e)$ to the first and $\prod_{e \in S} (1 - p_e)$ to the second case. As shown in the Figure 3, the approximated bound for the experimented network is tighter than the bound in [17] across transmission probabilities greater than 0.2.

Figure 3 considers a network $\mathcal{N}_3 = LI(G_3, \mathcal{P}_3, S_3)$, where S_3 contains a single seed v_3 , randomly selected from $V(G_3)$. The graph G_2 is an Erdős-Rényi random graph with 50 nodes where edges are drawn with probability 0.06. The transmission probability for each edge is drawn randomly from $\{0.1, 0.9\}$. The graph G_3 and the transmission probability matrix \mathcal{P}_3 are drawn 100 times, and the points in the figure are the averages. Figure 3 compares the bounds given in Theorem 2 obtained by conditioning on the edges with high transmission probability to the existing bounds from [17], for the ratio of the high

transmission probability edges ranging from 0 to 0.5. Note that since transmission probability 0.9 results in trivial bound, we use the approximated bound instead. As shown in the Figure 3, the approximated bound for the experimented network shows a similar trend to the simulated result, whereas the bound in [17] gives almost trivial bound.

VI. CONCLUSION

In this paper, we propose upper bounds for the expected number of influence nodes in independent cascade models. The bounds improve the bounds in [17], in particular, in the presence of sensitive edges, such as bottlenecks, seed adjacent, and high probability edges. We also present approximated bounds with low computational complexity. Finally, the results are supported by empirical data.

REFERENCES

- [1] M. Granovetter, "Threshold models of collective behavior," *American journal of sociology*, pp. 1420–1443, 1978.
- [2] D. Lopez-Pintado and D. J. Watts, "Social influence, binary decisions and collective dynamics," *Rationality and Society*, vol. 20, no. 4, pp. 399–443, 2008.
- [3] D. J. Watts, "A simple model of global cascades on random networks," *Proceedings of the National Academy of Sciences*, vol. 99, no. 9, pp. 5766–5771, 2002.
- [4] B. Shulgin, L. Stone, and Z. Agur, "Pulse vaccination strategy in the sir epidemic model," *Bulletin of Mathematical Biology*, vol. 60, no. 6, pp. 1123–1148, 1998.
- [5] A. Khelil, C. Becker, J. Tian, and K. Rothermel, "An epidemic model for information diffusion in manets," in *Proceedings of the 5th ACM international workshop on Modeling analysis and simulation of wireless and mobile systems*. ACM, 2002, pp. 54–60.
- [6] A. Banerjee, A. G. Chandrasekhar, E. Duflo, and M. O. Jackson, "The diffusion of microfinance," *Science*, vol. 341, no. 6144, p. 1236498, 2013.
- [7] J. Yang and S. Counts, "Predicting the speed, scale, and range of information diffusion in twitter," 2010.
- [8] J. Leskovec, L. A. Adamic, and B. A. Huberman, "The dynamics of viral marketing," *ACM Transactions on the Web (TWEB)*, vol. 1, no. 1, p. 5, 2007.
- [9] D. Kempe, J. Kleinberg, and É. Tardos, "Maximizing the spread of influence through a social network," in *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2003, pp. 137–146.
- [10] W. Chen, Y. Wang, and S. Yang, "Efficient influence maximization in social networks," in *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2009, pp. 199–208.
- [11] A. Goyal, W. Lu, and L. V. Lakshmanan, "Celf++: optimizing the greedy algorithm for influence maximization in social networks," in *Proceedings of the 20th international conference companion on World wide web*. ACM, 2011, pp. 47–48.
- [12] J. Leskovec, A. Krause, C. Guestrin, C. Faloutsos, J. VanBriesen, and N. Glance, "Cost-effective outbreak detection in networks," in *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2007, pp. 420–429.
- [13] C. Wang, W. Chen, and Y. Wang, "Scalable influence maximization for independent cascade model in large-scale social networks," *Data Mining and Knowledge Discovery*, vol. 25, no. 3, pp. 545–576, 2012.
- [14] M. E. Newman, "Spread of epidemic disease on networks," *Physical review E*, vol. 66, no. 1, p. 016128, 2002.
- [15] W. O. Kermack and A. G. McKendrick, "A contribution to the mathematical theory of epidemics," in *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, vol. 115, no. 772. The Royal Society, 1927, pp. 700–721.
- [16] M. Draief, A. Ganesh, and L. Massoulié, "Thresholds for virus spread on networks," in *Proceedings of the 1st international conference on Performance evaluation methodologies and tools*. ACM, 2006, p. 51.
- [17] R. Lemonnier, K. Scaman, and N. Vayatis, "Tight bounds for influence in diffusion networks and application to bond percolation and epidemiology," in *Advances in Neural Information Processing Systems*, 2014, pp. 846–854.
- [18] W. Chen, L. V. Lakshmanan, and C. Castillo, "Information and influence propagation in social networks," *Synthesis Lectures on Data Management*, vol. 5, no. 4, pp. 1–177, 2013.