Tight Bounds for Influence in Diffusion Networks and Application to Bond Percolation and Epidemiology

Authored by:

Nicolas Vayatis
Remi Lemonnier
Kevin Scaman

Abstract

In this paper, we derive theoretical bounds for the long-term influence of a node in an Independent Cascade Model (ICM). We relate these bounds to the spectral radius of a particular matrix and show that the behavior is sub-critical when this spectral radius is lower than 1. More specifically, we point out that, in general networks, the sub-critical regime behaves in O(sqrt(n)) where n is the size of the network, and that this upper bound is met for star-shaped networks. We apply our results to epidemiology and percolation on arbitrary networks, and derive a bound for the critical value beyond which a giant connected component arises. Finally, we show empirically the tightness of our bounds for a large family of networks.

1 Paper Body

The emergence of social graphs of the World Wide Web has had a considerable effect on propagation of ideas or information. For advertisers, these new diffusion networks have become a favored vector for viral marketing operations, that consist of advertisements that people are likely to share by themselves with their social circle, thus creating a propagation dynamics somewhat similar to the spreading of a virus in epidemiology ([1]). Of particular interest is the problem of influence maximization, which consists of selecting the top-k nodes of the network to infect at time t = 0 in order to maximize in expectation the final number of infected nodes at the end of the epidemic. This problem was first formulated by Domingues and Richardson in [2] and later expressed in [3] as an NP-hard discrete optimization problem under the Independent Cascade (IC) framework, a widely-used probabilistic model for information propagation. From an algorithmic point of view, influence maximization has been fairly well

studied. Assuming the transmission probability of all edges are known, Kempe, Kleinberg and Tardos ([3]) derived a greedy algorithm based on Monte-Carlo simulations that was shown to approximate the optimal solution up to a factor 1 ? 1e , building on classical results of optimization theory. Since then, various techniques were proposed in order to significantly improve the scalability of this algorithm ([4, 5, 6, 7]), and also to provide an estimate of the transmission probabilities from real data ([8, 9]). Recently, a series of papers ([10, 11, 12]) introduced continuous-time diffusion networks in which infection spreads during a time period T at varying rates across the different edges. While these models provide a more accurate representation of real-world networks for finite T , they are equivalent to the IC model when T ? ?. In this paper, will focus on such long-term behavior of the contagion. From a theoretical point of view, little is known about the influence maximization problem under the IC model framework. The most celebrated result established by Newman ([13]) proves the equivalence between bond percolation and the Susceptible-Infected-Removed (SIR) model in epidemiology ([14]) that can be identified to a special case of IC model where transmission probability are equal amongst all infectious edges. In this paper, we propose new bounds on the influence of any set of nodes. Moreover, we prove the existence of an epidemic threshold for a key quantity defined by the spectral radius of a given hazard 1

matrix. ? Under this threshold, the influence of any given set of nodes in a network of size n will be O(n), while the influence of a randomly chosen set of nodes will be O(1). We provide empirical evidence that these bounds are sharp for a family of graphs and sets of initial influencers and can therefore be used as what is to our knowledge the first closed-form formulas for influence estimation. We show that these results generalize bounds obtained on the SIR model by Draief, Ganesh and Massouli?e ([15]) and are closely related to recent results on percolation on finite inhomogeneous random graphs ([16]). The rest of the paper is organized as follows. In Sec. 2, we recall the definition of Information Cascades Model and introduce useful notations. In Sec. 3, we derive theoretical bounds for the influence. In Sec. 4, we show that our results also apply to the fields of percolation and epidemiology and generalize existing results in these fields. In Sec. 5, we illustrate our results by applying them on simple networks and retrieving well-known results. In Sec. 6, we perform experiments in order to show that our bounds are sharp for a family of graphs and sets of initial nodes.

2 2.1

Information Cascades Model Influence in random networks and infection dynamics

Let G = (V, E) be a directed network of n nodes and A ? V be a set of n0 nodes that are initially contagious (e.g. aware of a piece of information, infected by a disease or adopting a product). In the sequel, we will refer to A as the influencers. The behavior of the cascade is modeled using a probabilistic framework. The influencer nodes spread the contagion through the network by means of transmission through the edges of the network. More specifically, each contagious node can infect its neighbors with a certain probability. The

influence of A, denoted as ?(A), is the expected number of nodes reached by the contagion originating from A, i.e. X ?(A) = P(v is infected by the contagion —A). (1) v?V

We consider three infection dynamics that we will show in the next section to be equivalent regarding the total number of infected nodes at the end of the epidemic. Discrete-Time Information Cascades [DT IC(P)] At time t = 0, only the influencers are infected. Given a matrix P = (pij)ij ? [0, 1]n?n , each node i that receives the contagion at time t may transmit it at time t + 1 along its outgoing edge (i, j) ? E with probability pij . Node i cannot make any attempt to infect its neighbors in subsequent rounds. The process terminates when no more infections are possible. Continuous-Time Information Cascades [CT IC(F, T)] At time t = 0, only the influencers are infected. Given a matrix F = (fij)ij of non-negative integrable functions, each node i that receives the contagion at time t may transmit it at time s ¿ t along its outgoing edge (i, j) ? E with stochastic rate of occurrence fij (s ? t). The process terminates at a given deterministic time T ¿ 0. This model is much richer than Discrete-time IC, but we will focus here on its behavior when T = ?. Random Networks [RN (P)] Given a matrix P = (pij)ij ? [0, 1]n?n , each edge (i, j) ? E is removed independently of the others with probability 1 ? pij . A node i ? V is said to be infected if i is linked to at least one element of A in the spanning subgraph G 0 = (V, E 0) where E 0 ? E is the set of non-removed edges. For any v ? V, we will designate by influence of v the influence of the set containing only v, i.e. ?({v}). We will show in Section 4.2 that, if P is symmetric and G undirected, these three infection processes are equivalent to bond percolation and the influence of a node v is also equal to the expected size of the connected component containing v in G 0 . This will make our results applicable to percolation in arbitrary networks. Following the percolation literature, we will denote as sub-critical a cascade whose influence is not proportional to the size of the network n. 2

2.2

The hazard matrix

In order to linearize the influence problem and derive upper bounds, we introduce the concept of hazard matrix, which describes the behavior of the information cascade. As we will see in the following, in the case of Continuous-time Information Cascades, this matrix gives, for each edge of the network, the integral of the instantaneous rate of transmission (known as hazard function). The spectral radius of this matrix will play a key role in the influence of the cascade. Definition. For a given graph G = (V, E) and edge transmission probabilities pij , let H be the n ? n matrix, denoted as the hazard matrix, whose coefficients are

? ln(1 ? pij) if (i, j) ? E Hij = . (2) 0 otherwise Next lemma shows the equivalence between the three definitions of the previous section. Lemma 1. For a given graph G = (V, E), set of influencers A, and transmission probabilities matrix P, the distribution of the set of infected nodes is equal under R ?the infection dynamics DT IC(P), CT IC(F, ?) and RN (P), provided that for any (i, j) ? E, 0 fij (t)dt = Hij . Definition. For a given set of influencers A ? V,

we will denote as H(A) the hazard matrix except for zeros along the columns whose indices are in A: H(A)ij = 1{j ?A} Hij . /

(3)

We recall that for any square matrix M , its spectral radius ?(M) is defined by ?(M) = maxi (—?i —) where ?1 , ..., ?n are the (possibly repeated) eigenvalues of matrix M . We will also use that, when ¿ ¿ M is a real square matrix with positive entries, ?(M +M) = supX XX ¿MXX . 2 Remark. When the pij are small, the hazard matrix is very close to the transmission matrix P. This implies that, for low pij values, the spectral radius of H will be very close to that of P. More specifically, a simple calculation holds ?(P) ? ?(H) ?

? ln(1 ? kPk?) ?(P), kPk?

(4)

for x ? 1? implies that the where kPk? = maxi,j pij . The relatively slow increase of ? ln(1?x) x behavior of ?(P) and ?(H) will be of the same order of magnitude even for high (but lower than 1) values of kPk? .

3

Upper bounds for the influence of a set of nodes

Given A ? V the set of influencer nodes and —A— = n0 ¡ n, we derive here two upper bounds for the influence of A. The first bound (Proposition 1) applies to any set of influencers A such that —A— = n0 . Intuitively, this result correspond to a best-case scenario (or a worst-case scenario, depending on the viewpoint), since we can target any set of nodes so as to maximize the resulting contagion. ¿

Proposition 1. Define ?c (A) = ?(H(A)+H(A)). Then, for any A such that —A— = n0 ¡ n, denoting 2 by ?(A) the expected number of nodes reached by the cascade starting from A: ?(A) ? n0 + ?1 (n ? n0), where ?1 is the smallest solution in [0, 1] of the following equation:

?c (A)n0 ?1 ? 1 + exp ??c (A)?1 ? = 0. ?1 (n ? n0) Corollary 1. Under the same assumptions: 3

(5)

(6)

s ? if ?c (A) ¡ 1,
? if ?c (A) ? 1,
?(A) ? n0 +
?c (A) p n0 (n ? n0), 1 ? ?c (A) 2?c (A)
!

. ?(A) ? n ? (n ? n0) exp ??c (A) ? p 4n/n0 ? 3 ? 1

? In particular, when ?c (A) ¡ 1, ?(A) = O(n) and the regime is sub-critical. The second result (Proposition 2) applies in the case where A is drawn from a uniform distribution over the ensemble of sets of n0 nodes chosen amongst n (denoted as Pn0 (V)). This result corresponds to the average-case scenario in a setting where the initial influencer nodes are not known and drawn independently of the transmissions over each edge. ¿

Proposition 2. Define ?c = ?(H+H). Assume the set of influencers A is drawn from a uniform 2 distribution over Pn0 (V). Then, denoting by ?uniform the expected number of nodes reached by the cascade starting from A: ?uniform

? n0 + ?2 (n ? n0), where ?2 is the unique solution in [0, 1] of the following equation:

?c n 0 ?2 ? 1 + exp ??c ?2 ? = 0. n ? n0

(7)

(8)

Corollary 2. Under the same assumptions: ? if ?c ¡ 1, ? if ?c ? 1,

n0 , 1 ? ?c

?c ?uniform ? n ? (n ? n0) exp ? . 1 ? nn0 ?uniform ?

In particular, when ?c ¡ 1, ?uniform = O(1) and the regime is sub-critical. ? The difference in the sub-critical regime between O(n) and O(1) for the worst and average case influence is an important feature of our results, and is verified in our experiments (see Sec. 6). Intuitively, when the network is inhomogeneous and contains highly central nodes (e.g. scale-free networks), there will be a significant difference between specifically targeting the most central nodes and random targeting (which will most probably target a peripheral node).

4

Application to epidemiology and percolation

Building on the celebrated equivalences between the fields of percolation, epidemiology and influence maximization, we show that our results generalize existing results in these fields. 4.1

Susceptible-Infected-Removed (SIR) model in epidemiology

We show here that Proposition 1 further improves results on the SIR model in epidemiology. This widely used model was introduced by Kermac and McKendrick ([14]) in order to model the propagation of a disease in a given population. In this setting, nodes represent individuals, that can be in one of three possible states, susceptible (S), infected (I) or removed (R). At t = 0, a subset A of n0 nodes is infected and the epidemic spreads according to the following evolution. Each infected node transmits the infection along its outgoing edge (i, j) ? E at stochastic rate of occurrence ? and is removed from the graph at stochastic rate of occurrence ?. The process ends for a given T ¿ 0. It is straightforward that, if the removed events are not observed, this infection process is equivalent to CT IC(F, T) where for any (i, j) ? E,fij (t) = ? exp(??t). The hazard matrix H is therefore

equal to ?? A where A = 1{(i,j)?E} ij is the adjacency matrix of the underlying network. Note 4

that, by Lemma 1, our results can be used in order to model the total number of infected nodes in a setting where infection and recovery rates of a given node exhibit a non-exponential behavior. For instance, incubation periods for different individuals generally follow a log-normal distribution [17], which indicates that continuous-time IC with a log-normal rate of removal might be well-suited to model some kind of infections. It was recently shown by Draief, Ganesh and Massouli?e ([15]) that, in the case of undirected networks, and if ??(A) ¡ ?, ? nn0 ?(A) ? . (9) ? 1 ? ? ?(A) ? This result shows, that, when ?(H) = ?? ?(A) ¡ 1, the influence of set of nodes A is O(n). We show in the next lemma that this result is a direct consequence of Corollary 1: the condition ?c (A) ¡ 1 is weaker than ?(H) ¡ 1 and, under these conditions, the bound of

Corollary 1 is tighter. Lemma 2. For any symmetric adjacency matrix A, initial set of influencers A such that —A— = n0 ¡ ? n, ? ¿ 0 and ? ¡ ?(A) , we have simultaneously ?c (A) ? ?? ?(A) and s ? nn0 ?c (A) p , (10) n0 (n ? n0) ? n0 + ? 1 ? ?c (A) 1 ? ? ?(A) where the condition ? ¡

? ?(A)

imposes that the regime is sub-critical.

Moreover, these new bounds capture with more accuracy the behavior of the influence in extreme cases. In the limit ? ? 0, the difference between the two? bounds is significant, because Proposition 1 yields ?(A) ? n0 whereas (9) only ensures ?(A) ? nn0 . When n = n0 , Proposition 1 also 0 ensures that ?(A) = n0 whereas (9) yields ?(A) ? 1? ?n?(A) . Secondly, Proposition 1 gives also ?

bounds in the case ??(A) ? ?. Finally, Proposition 1 applies to more general cases that the classical homogeneous SIR model, and allows infection and recovery rates to vary across individuals. 4.2

Bond percolation

Given a finite undirected graph G = (V, E), bond percolation theory describes the behavior of connected clusters of the spanning subgraph of G obtained by retaining a subset E 0 ? E of edges of G according to a given distribution.When these removals occur independently along each edge with same probability 1 ? p, this process is called homogeneous percolation and is fairly well known (see e.g [18]). The inhomogeneous case, where the independent edge removal probabilities 1 ? pij vary across the edges, is more intricate and has been the subject of recent studies. In particular, results on critical probabilities and size of the giant component have been obtained by Bollobas, Janson and Riordan in [16]. However, these bounds hold for a particular class of asymptotic graphs (inhomogeneous random graphs) when n ? ?. In the next lemma, we show that our results can be used in order to obtain bounds that hold in expectation for any fixed graph. Lemma 3. Let P = (pij)ij ? [0, 1]n?n be a symmetric matrix. Let G 0 = (V, E 0) be the undirected subgraph of G such that each edge {i, j} ? E is removed independently with probability 1 ? pij . Let Gd = (V, Ed) be the directed graph such that (i, j) ? Ed ?? {i, j} ? E. Then, for any v ? V, the expected size of the connected component containing v in G 0 is equal to the influence of v in Gd under the infection process DT IC(P). We now derive an upper bound for C1 (G 0), the size of the largest connected component of the spanning subgraph G 0 = (V, E 0). In the following, we will denote by E[C1 (G 0)] the expected value of this random variable, given P = (pij)ij . Proposition 3. Let G = (V, E) be an undirected network where each edge {i, j} ? E has an independent probability 1 ? pij of being removed. The expected size of the largest connected component of the resulting subgraph G 0 is upper bounded by: ? (11) E[C1 (G 0)] ? n ?3 , where ?3 is the unique solution in [0, 1] of the following equation:

n?1 n ?3 ? 1 + exp ? ?(H)?3 = 0. n n?1 5

(12)

Moreover, the resulting network has a probability of being connected upper bounded by: P(G 0 is connected) ? ?3 .

(13)

In the case $\rho(H) < 1$, we can further simplify our bounds in the same way than for Propositions 1 and 2. q n Corollary 3. In the case $\rho(H) < 1$, $E[C_1 (G'_0)] \le \frac{1}{1-\rho(H)}$. Whereas our results hold for any $n \in N$, classical results in percolation theory study the asymptotic behavior of sequences of graphs when $n \to \infty$. In order to further compare our results, we therefore consider sequences of spanning subgraphs $(G'_{0n})_{n \in N}$, obtained by removing each edge of graphs of n nodes $(G_n)_{n \in N}$ with probability $1 - p_{nij}$. A previous result ([16], Corollary 3.2 of section 5) states that, for particular sequences known as inhomogeneous random graphs and under a given sub-criticality condition, $C_1(G'_{0n}) = o(n)$ asymptotically almost surely (a.a.s.), i.e with probability going to 1 as $n \to \infty$. Using Proposition 3, we get for our part the following result:

Corollary 4. Assume the sequence $H_n = - \ln(1 - p_{nij})_{ij}$ is such that $n \in N$

$$\limsup_{n \to \infty} \rho(H_n) < 1. \tag{14}$$

Then, for any $\epsilon > 0$, we have asymptotically almost surely when $n \to \infty$, $C_1(G_{n0}) = o(n^{1/2+\epsilon}).$

$$(15)$$

This result is to our knowledge the first to bound the expected size of the largest connected component in general arbitrary networks.

5

Application to particular networks

In order to illustrate our theoretical results, we now apply our bounds to three specific networks and compare them to existing results, showing that our bounds are always of the same order than these specific results. We consider three particular networks: 1) star-shaped networks, 2) Erdős-Rényi networks and 3) random graphs with an expected degree distribution. In order to simplify these problems and exploit existing theorems, we will consider in this section that $p_{ij} = p$ is fixed for each edge $\{i, j\} \in E$. Infection dynamics thus only depend on p, the set of influencers A, and the structure of the underlying network. 5.1

Star-shaped networks

For a star shaped network centered around a given node v_1, and $A = \{v_1\}$, the exact influence is computable and writes $\sigma(\{v_1\}) = 1 + p(n-1)$. As $H(A)_{ij} = - \ln(1 - p)1_{\{i=1, j \neq 1\}}$, the spectral radius is given by

$$\frac{H(A) + H(A)_\top}{2} \Rightarrow -\frac{\ln(1-p)}{2}\sqrt{?} = \sqrt{n-1}. \tag{16}$$

Therefore, Proposition 1 states that $\sigma(\{v_1\}) \le 1 + (n-1)\gamma_1$ where γ_1 is the solution of equation

$$-\frac{\ln(1-p)}{2}\sqrt{\frac{1}{n-1}}\gamma_1 = \exp\left(\frac{\gamma_1}{n-1} + ?\right). \tag{17}$$

It is worth mentionning that, when $p = \frac{c}{n-1}$, $\gamma_1 = \frac{c}{n-1}$ is solution of (17) and therefore the σ bound is $\sigma(\{v_1\}) \le 1 + n - 1$ which is tight. Note that, in the case of star-shaped networks, the influence does not present a critical behavior and is always linear with respect to the total number of nodes n.

5.2

Erdős-Rényi networks

For Erdős-Rényi networks $G(n, p)$ (i.e. an undirected network with n nodes where each couple of nodes $(i, j) \in V^2$ belongs to E independently of the

others with probability p), the exact influence 6

of a set of nodes is not known. However, percolation theory characterizes the limit behavior of the giant connected component when n ? ?. In the simplest case of Erd?os-R?enyi networks G(n, nc) the following result holds: Lemma 4. (taken from [16]) For a given sequence of Erd?os-R?enyi networks G(n, nc), we have: ? if c ¡ 1, C1 (G(n, nc)) ?

3 (1?c)2

log(n) a.a.s.

? if c ¿ 1, C1 (G(n, nc)) = (1 + o(1))?n a.a.s. where ? ? 1 + exp(??c) = 0. As previously stated, our results hold for any given graph, and not only asymptotically. However, we get an asymptotic behavior consistent with the aforementioned result. Indeed, using notations of n section 4.2, Hij = ? ln(1 ? nc)1{i6=j} and ?(Hn) = ?(n ? 1) ln(1 ? nc). Using Proposition 3, and noting that ?3 = (1 + o(1))?, we get that, for any ¿ 0: ? if c ¡ 1, C1 (G(n, nc)) = o(n1/2+) a.a.s. ? if c ¿ 1, C1 (G(n, nc)) ? (1 + o(1))?n1+ a.a.s., where ? ? 1 + exp(??c) = 0. 5.3

Random graphs with given expected degree distribution

In this section, we apply our bounds to random graphs whose expected degree distribution is fixed (see e.g [19], section 13.2.2). More specifically, let w = (wi)i?{1,...,n} be the expected degree of each node of the network. For a fixed w, let G(w) be a random graph whose edges are selected independently and randomly with probability 1{i6=j} wi wj . (18) qij = P k wk For these graphs, results on the volume of connected components (i.e the expected sum of degrees of the nodes in these components) were derived in [20] but our work gives to our knowledge the first result on the size of the giant component. Note that Erd?os-R?enyi G(n, p) networks are a special case of (18) where wi = np for any i ? V. In order to further compare our results, we note that these graphs are also very similar to the widely used configuration model where node degrees are fixed to a sequence w, the main difference being that the occupation probabilities pij are in this P case not independent anymore. For configuration P models, a giant component exists if and only if i wi2 ¿ 2 i wi ([21, 22]). In theP case of graphs P with given expected degree distribution, we retrieve the key role played by the ratio i wi2 / i wi ¿) ¡ 1 where in our criterion of non-existence of the giant component given by ?(H+H 2 P

w2 H + H¿ ? ? ?((qij)ij) ? Pi i . (19) 2 i wi The left-hand approximation is particularly good when the qij are small. This is for instance the case ? as soon as there exists ? ¡ 1 such that, for any i ? V, wi = o(n P). The right-hand side P is based P on 2 the fact that the spectral radius of the matrix (qij + 1{i=j} wi / k wk)ij is given by i wi2 / i wi .

6

Experimental results

In this section, we show that the bounds given in Sec. 3 are tight (i.e. very close to empirical results in particular graphs), and are good approximations of the influence on a large set of random networks. Fig. 1a compares experimental simulations of the influence to the bound derived in proposition 1. The considered networks have n = 1000 nodes and are of 6 types (see e.g [19] for further

8

details on these different networks): 1) Erd?os-R?enyi networks, 2) Preferential attachment networks, 3) Smallworld networks, 4) Geometric random networks ([23]), 5) 2D regular grids and 6) totally connected networks with fixed weight b ? [0, 1] except for the ingoing and outgoing edges of the influencer node A = {v1 } having weight a ? [0, 1]. Except for totally connected networks, edge probabilities are set to the same value p for each edge (this parameter was used to tune the spectral radius ?c (A)). All points of the plots are averages over 100 simulations. The results show that the bound in proposition 1 is tight (see totally connected networks in Fig. 1a) and close to the real influence for a large

7

class of random ? networks. In particular, the tightness of the bound around ?c (A) = 1 validates the behavior in n of the worst-case influence in the sub-critical regime. Similarly, Fig. 1b compares 1000

900

900

800

800

700

700 uniform

600

influence (?

influence (?(A))

)

1000

500 400 totally connected erdos renyi preferential attachment small World geometric random 2D grid upper bound

300 200 100 0

0

2

4 6 8 spectral radius of the Hazard matrix (?c(A))

(a) Fixed set of influencers

600 500 400 totally connected erdos renyi preferential attachment small World geometric random 2D grid upper bound

300 200 100 0

10

0

2

4 6 8 spectral radius of the Hazard matrix (?c)

10

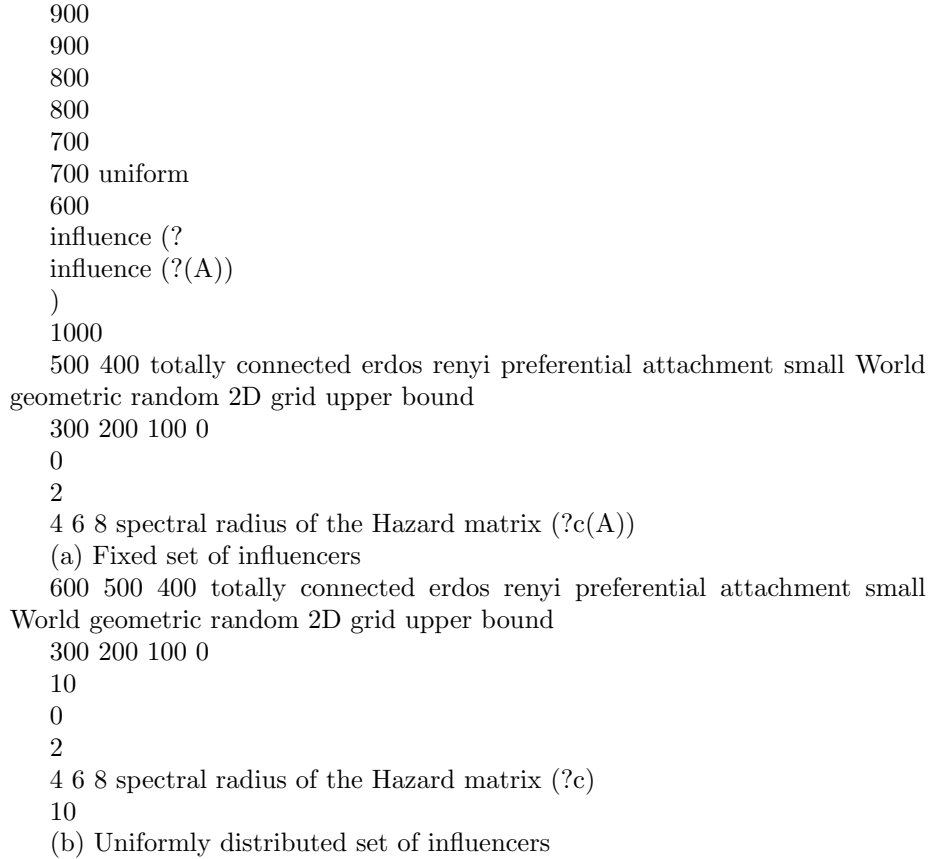(b) Uniformly distributed set of influencers

Figure 1: Empirical influence on random networks of various types. The solid lines are the upper bounds in propositions 1 (for Fig. 1a) and 2 (for Fig. 1b). experimental simulations of the influence to the bound derived in proposition 2 in the case of random initial influencers. While this bound is not as tight as the previous one, the behavior of the bound agrees with experimental simulations, and proves a relatively good approximation of the influence under a

random set of initial influencers. It is worth mentioning that the bound is tight for the subcritical regime and shows that corollary 2 is a good approximation of ?uniform when ?c ¡ 1. In order to verify the criticality of ?c (A) = 1, we compared the behavior of ?(A)?w.r.t the size of the network n. When ?c (A) ¡ 1 (see Fig. 2a in which ?c (A) = 0.5), ?(A) = O(n), and the bound is tight. On the contrary, when ?c (A) ¿ 1 (see Fig. 2b in which ?c (A) = 1.5), ?(A) = O(n), and ?(A) is linear w.r.t. n for most random networks. totally connected erdos renyi preferential attachment small World geometric random 2D grid upper bound

influence (?(A))

25

20

400

15

300

200

10

100

5

0

totally connected erdos renyi preferential attachment small World geometric random 2D grid upper bound

500

influence (?(A))

30

0

200

400 600 size of the network (n)

800

1000

(a) Sub-critical regime: ?c (A) = 0.5

0

0

200

400 600 size of the network (n)
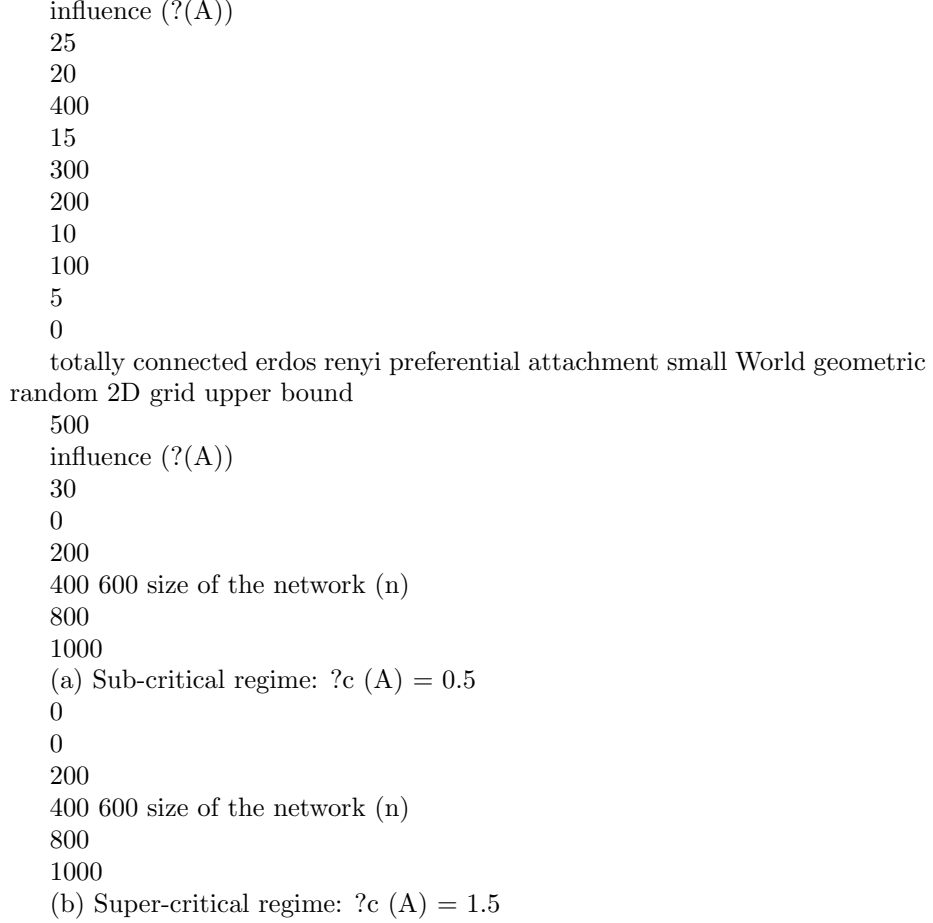
800

1000

(b) Super-critical regime: ?c (A) = 1.5

Figure 2: Influence w.r.t. the size of the network in the sub-critical and super-critical regime. The solid line is the upper bound in proposition 1. Note the square-root versus linear behavior.

7

Conclusion

In this paper, we derived the first upper bounds for the influence of a given set of nodes in any finite graph under the Independent Cascade Model (ICM) framework, and relate them to the spectral radius of a given hazard matrix. We show that these bounds can also be used to generalize previous results

in the fields of epidemiology and percolation. Finally, we provide empirical evidence that these bounds are close to the best possible for general graphs. Acknowledgments This research is part of the SODATECH project funded by the French Government within the program of ?Investments for the Future ? Big Data?. 8

2 References

[1] Justin Kirby and Paul Marsden. Connected marketing: the viral, buzz and word of mouth revolution. Elsevier, 2006. [2] Pedro Domingos and Matt Richardson. Mining the network value of customers. In Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining, pages 57?66. ACM, 2001. ? Tardos. Maximizing the spread of influence through a social [3] David Kempe, Jon Kleinberg, and Eva network. In Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD ?03, pages 137?146, New York, NY, USA, 2003. ACM. [4] Wei Chen, Yajun Wang, and Siyu Yang. Efficient influence maximization in social networks. In Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining, pages 199?208. ACM, 2009. [5] Wei Chen, Chi Wang, and Yajun Wang. Scalable influence maximization for prevalent viral marketing in large-scale social networks. In Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining, pages 1029?1038. ACM, 2010. [6] Amit Goyal, Wei Lu, and Laks VS Lakshmanan. Celf++: optimizing the greedy algorithm for influence maximization in social networks. In Proceedings of the 20th international conference companion on World wide web, pages 47?48. ACM, 2011. [7] Kouzou Ohara, Kazumi Saito, Masahiro Kimura, and Hiroshi Motoda. Predictive simulation framework of stochastic diffusion model for identifying top-k influential nodes. In Asian Conference on Machine Learning, pages 149?164, 2013. [8] Manuel Gomez Rodriguez, Jure Leskovec, and Andreas Krause. Inferring networks of diffusion and influence. In Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining, pages 1019?1028. ACM, 2010. [9] Seth A. Myers and Jure Leskovec. On the convexity of latent social network inference. In NIPS, pages 1741?1749, 2010. [10] Manuel Gomez-Rodriguez, David Balduzzi, and Bernhard Sch?olkopf. Uncovering the temporal dynamics of diffusion networks. In ICML, pages 561?568, 2011. [11] Manuel G Rodriguez and Bernhard Sch?olkopf. Influence maximization in continuous time diffusion networks. In Proceedings of the 29th International Conference on Machine Learning (ICML-12), pages 313?320, 2012. [12] Nan Du, Le Song, Manuel Gomez-Rodriguez, and Hongyuan Zha. Scalable influence estimation in continuous-time diffusion networks. In NIPS, pages 3147?3155, 2013. [13] Mark EJ Newman. Spread of epidemic disease on networks. Physical review E, 66(1):016128, 2002. [14] William O Kermack and Anderson G McKendrick. Contributions to the mathematical theory of epidemics. ii. the problem of endemicity. Proceedings of the Royal society of London. Series

A, 138(834):55? 83, 1932. [15] Moez Draief, Ayalvadi Ganesh, and Laurent Massouli?e. Thresholds for virus spread on networks. In Proceedings of the 1st international conference on Performance evaluation methodolgies and tools, page 51. ACM, 2006. [16] B?ela Bollob?as, Svante Janson, and Oliver Riordan. The phase transition in inhomogeneous random graphs. Random Structures & Algorithms, 31(1):3?122, 2007. [17] Kenrad E Nelson. Epidemiology of infectious disease: general principles. Infectious Disease Epidemiology Theory and Practice. Gaithersburg, MD: Aspen Publishers, pages 17?48, 2007. [18] Svante Janson, Tomasz Luczak, and Andrzej Rucinski. Random graphs, volume 45. John Wiley & Sons, 2011. [19] Mark Newman. Networks: An Introduction. Oxford University Press, Inc., New York, NY, USA, 2010. [20] Fan Chung and Linyuan Lu. Connected components in random graphs with given expected degree sequences. Annals of combinatorics, 6(2):125?145, 2002. [21] Michael Molloy and Bruce Reed. A critical point for random graphs with a given degree sequence. Random structures & algorithms, 6(2-3):161?180, 1995. [22] Michael Molloy and Bruce Reed. The size of the giant component of a random graph with a given degree sequence. Combinatorics probability and computing, 7(3):295?305, 1998. [23] Mathew Penrose. Random geometric graphs, volume 5. Oxford University Press Oxford, 2003.

9