Journal of Artificial Intelligence Research 51 (2014) 207-226 Submitted 1/14; published 09/14

Sensitivity of Diffusion Dynamics to Network Uncertainty

Abhijin Adiga ABHIJIN@VBI.VT.EDU
Chris J. Kuhlman CKUHLMAN@VBI.VT.EDU
Network Dynamics and Simulation Science Laboratory,
Virginia Bioinformatics Institute,
Virginia Tech, VA 24061

Henning S. Mortveit HMORTVEI@VBI.VT.EDU
Network Dynamics and Simulation Science Laboratory,
Virginia Bioinformatics Institute, and
Department of Mathematics,
Virginia Tech, VA 24061

Anil Kumar S. Vullikanti AKUMAR@VBI.VT.EDU
Network Dynamics and Simulation Science Laboratory,
Virginia Bioinformatics Institute, and
Department of Computer Science,
Virginia Tech, VA 24061

Abstract

Simple diffusion processes on networks have been used to model, analyze and predict diverse phenomena such as spread of diseases, information and memes. More often than not, the underlying network data is noisy and sampled. This prompts the following natural question: how sensitive are the diffusion dynamics and subsequent conclusions to uncertainty in the network structure?

In this paper, we consider two popular diffusion models: *Independent cascade* (IC) model and *Linear threshold* (LT) model. We study how the expected number of vertices that are influenced/infected, for particular initial conditions, are affected by network perturbations. Through rigorous analysis under the assumption of a reasonable perturbation model we establish the following main results. (1) For the IC model, we characterize the sensitivity to network perturbation in terms of the critical probability for phase transition of the network. We find that the expected number of infections is quite stable, unless the transmission probability is close to the critical probability. (2) We show that the standard LT model with uniform edge weights is relatively stable under network perturbations. (3) We study these sensitivity questions using extensive simulations on diverse real world networks and find that our theoretical predictions for both models match the observations quite closely. (4) Experimentally, the transient behavior, i.e., the time series of the number of infections, in both models appears to be more sensitive to network perturbations.

1. Introduction

A number of diverse phenomena are modeled by simple diffusion processes on graphs, such as the spread of epidemics (Newman, 2003), viral marketing (Kempe, Kleinberg, & Tardos, 2005; Goldenberg, Libai, & Muller, 2001) and memes in online social media (Romero, Meeder, & Kleinberg, 2011; Bakshy, Hofman, Mason, & Watts, 2011). It is common to associate with each vertex a state of 0 (denoting "not infected" or "not influenced") or state 1 (denoting "infected" or "influenced") in these models; each node in state 0 switches to state 1 based on a probabilistic rule and nodes in state 1 remain in state 1. We focus on two such models, referred to as the *independent cascade* (IC) model (which is a special case of the SIR process), and the *linear threshold* (LT) model. In most applications, however, the underlying networks are inherently noisy and incomplete, since they are often inferred by indirect measurements, for instance: (i) networks based on Twitter data (González-Bailón, Borge-Holthoefer, Rivero, & Moreno, 2011; Bakshy et al., 2011; Galuba, 2010) are constructed by limited samples available through public APIs, (ii) biological networks are inferred by experimental correlations (Hagmann, 2008; Schwab, Bruinsma, Feldman, & Levine, 2010), which might be incomplete, and (iii) the Internet router/AS level graphs are constructed using traceroutes (Faloutsos, Faloutsos, & Faloutsos, 1999), which are known to give a biased and incomplete structure (Achlioptas, Clauset, Kempe, & Moore, 2005).

This raises a fundamental issue for diffusion processes on networks: *How does the uncertainty in the network affect the conclusions drawn from a study of the diffusion dynamics over that network?* For instance, how robust is an inference that there will be no large outbreak in the network, in the face of noise/uncertainty in the network? Recent statistical and simulation-based studies involving perturbation of the network by "rewiring" pairs of edges (which preserves the degree sequence) show that changes in the network structure significantly alter the dynamics even when aggregate structural properties such as the degree distribution and assortativity are preserved (Eubank, 2010; Chen, 2010). Surprisingly, there is limited mathematically rigorous work to explain the empirical findings in a systematic manner, despite a large body of research on diffusion models.

Our work is motivated by these considerations of sensitivity of the dynamics to noise and the adequacy of sampling of a network $G = (V, E)$. Since there is very limited understanding of how noise should be modeled, we consider two simple *random edge perturbation* models for noise: uniform and degree-assortative. In uniform perturbation, each pair u, v of vertices is selected for edge addition or deletion (or both) with probability $\frac{\epsilon}{n}$, where $\epsilon > 0$ is a parameter, and n is the number of vertices; thus, on average, the perturbed graph differs in $\frac{\epsilon}{n}\binom{n}{2} \approx \frac{\epsilon n}{2}$ edges. This model has been used quite extensively both in social network analysis and computer science for understanding the sensitivity to graph properties (Costenbader & Valente, 2003; Borgatti, Carley, & Krackhardt, 2006; Flaxman & Frieze, 2004; Flaxman, 2007). We study how the expected number of infections, given some initial conditions, is affected by the magnitude of the perturbation parameter, ϵ. In the degree-assortative perturbation, the probability of an edge modification is proportional to the product of the degrees of the end points in G.

1.1 Our Contributions

All the results we obtain are under the assumption of the *uniform edge addition model*, i.e., edges absent in the network are added with probability $\frac{\epsilon}{n}$ to obtain the perturbed network. Later, in Section 5, we compare the addition model with the addition/deletion model and also discuss our results on degree-assortative perturbation. We describe the independent cascade and linear threshold models in Section 2.2.

1.1.1 Independent Cascade Model

We consider networks G which exhibit a phase transition in the infection sizes, with a critical transition probability p_c (see Section 2 for definitions). In Theorem 1, we characterize the expected number of infections in the perturbed graph in terms of transmission probability p and p_c for a single random seed node. When $p < p_c$, we show that there exists a threshold ϵ_t such that for any positive constant $c < 1$: (i) if $\epsilon < (1-c)\epsilon_t$, and in addition, if p_c and the average degree d_{avg} satisfy a technical condition, then, the expected number of infections in the perturbed graph remains close to that in G and, (ii) if $\epsilon > (1+c)\epsilon_t$, there is a phase transition, and the expected number of infections after perturbation is much larger than that in G. The main implication is that the dynamics are quite robust to perturbations, unless the transmission probability is close to p_c, and ϵ is over some critical value. We find this to be consistent with extensive simulations on a large number of real networks—the sensitivity to perturbations is maximized at a point which approximately matches the experimentally determined threshold ϵ_t in many networks. We also examine the transient behavior (i.e., the time series of the number of infections) studying in particular how the time and magnitude of the peak number of new infections are affected by uncertainty. We find these measures to be more sensitive than the expected total number of infections.

1.1.2 Linear Threshold Model

In Theorem 2, we show formally that for any network G with maximum degree $\Delta = O(n/\log n)$, the expected number of infections after perturbation, starting at s random initial infections, is bounded by $O(s(\Delta + \epsilon + \log n)\log n)$. This implies that the dynamics are quite stable for low s and ϵ. Our result is based on the analysis of the random graph model in which each node selects a random in-edge. This is shown to "correspond" to the LT model (Kempe, Kleinberg, & Tardos, 2003). We first show that the diameter is bounded by $O(\Delta' \log n)$, where Δ' is the maximum degree of the perturbed graph, and then prove that the expected number of infections, starting at a random source, is bounded by the diameter. Our theoretical bounds corroborate well with our experimental observations on a large set of real networks, which show a gradual variation with ϵ. We find that the expected number of infections grows more sharply with ϵ, as the number of sources is increased.

1.1.3 Discussion and Implications

From the point of view of dynamical system theory, our work may be regarded as a study of *stability* of dynamics over a network with respect to the edge structure. The existence of the critical value for the parameter ϵ in the IC model can be thought of as a bifurcation point. Admittedly, our results only hold for the specific random edge perturbation model of noise; uncertainty in networks is a much more complex process, and might involve dependencies arising out of the network evolution. Although we focus on specific dynamical properties and the random edge perturbation model, our results give the first rigorous theoretical and empirical analysis of the noise susceptibility of these diffusion models. Further, our analytical and empirical techniques, based on the random graph characterization, are likely to help in the analysis of other more complex noise models, which take dependencies into account.

1.2 Related Work

Noise and issues of sampling are well recognized as fundamental challenges in complex networks, and there has been some work on characterizing it and the sensitivity to different parameters, especially in network properties. In some works (Costenbader & Valente, 2003; Borgatti et al., 2006), certain centrality measures are shown to be robust to random edge and node perturbations, and in another (Achlioptas et al., 2005), it is shown that there is an inherent bias in traceroute-based inference of the Internet router network, which might give incorrect degree distributions. Flaxman and Frieze (Flaxman & Frieze, 2004; Flaxman, 2007) formally characterize conditions under which the graph expansion and diameter are highly sensitive to random edge additions; these are among the few analytical results of this type. Some of the approaches to address noise include: (i) the prediction of missing links using clustering properties (Clauset, Moore, & Newman, 2008), and (ii) "property testing" algorithms (Ron, 2010) and "smoothed analysis" (Spielman, 2009) for efficient computation of graph properties.

To our knowledge, most work on the sensitivity of graph dynamical systems to noise in a network is empirical. However, for regular networks such as rings, topics such as synchronization and bifurcations are studied (Kaneko, 1985; Wu, 2005). As discussed earlier, the effects of changes in a network by edge rewirings on epidemic properties are investigated (Eubank, 2010; Chen, 2010). The effect of stochastic changes in the network on influence maximization problems is studied (Lahiri, Maiya, Caceres, Habiba, & Berger-Wolf, 2008). Using simulations, they find that in the LT model, the spread size is quite robust; our techniques help explain some of these observations.

1.2.1 Organization

In Section 2, we introduce notation and describe the noise models and diffusion models in detail. In Sections 3 and 4, we analyze the sensitivity of IC and LT models, respectively. Experimental results are presented in Section 5, and we conclude in Section 6.

2. Preliminaries

We consider only undirected, simple networks. For a network $G = (V, E)$, let Δ denote its maximum degree and d_{avg}, its average degree. For any vertex v, $\deg(v, G)$ and $N(v)$ denote its degree and the set of neighbors respectively. Let λ correspond to the largest eigenvalue of the adjacency matrix of G. We say that an event $A(n)$ occurs *asymptotically almost surely* (a.a.s.) if $P(A(n)) \to 1$ as $n \to \infty$.

2.1 Noise Models

Since there is no consensus on the best way to model uncertainty and noise, we consider two simple models of random edge modifications: (i) uniform and (ii) degree-assortative perturbations. *Uniform perturbation* has been studied quite extensively in social network analysis (Costenbader & Valente, 2003; Borgatti et al., 2006); some problems have also been studied analytically in this model (Flaxman & Frieze, 2004; Flaxman, 2007). Let $G = (V, E)$ be the unperturbed graph; all graphs in this work are undirected and simple. Let $R_u(\epsilon) = (V, E(\epsilon))$ be a random graph on V in which each pair $u, v \in V$ is connected with probability $\frac{\epsilon}{n}$. In our analysis, we consider perturbations involving just addition of edges: This is denoted by $G + R_u(\epsilon)$, and consists of all edges $(u, v) \in E \cup E(R_u(\epsilon))$. For our experimental studies, we also considered addition/deletion of edges. In this case, the perturbation graph $G' = G \oplus R_u(\epsilon)$ is a graph constructed in the following manner: each pair $u, v \in V$ is connected in G' if $(u, v) \in R_u(\epsilon) - E$ or $(u, v) \in E - R_u(\epsilon)$. In other words, each pair u, v is selected for addition/deletion with probability $\frac{\epsilon}{n}$. In the *degree-assortative* perturbation model, we consider random graph $R_d(\epsilon)$ for which $u, v \in V$ are adjacent with probability $\left(\frac{\deg(u,G)\deg(v,G)}{d_{\text{avg}}^2}\right)\frac{\epsilon}{n}$, i.e., the edge probability is proportional to the product of the degrees of the end points in G. In both models, the expected number of edge modifications is approximately $\frac{\epsilon n}{2}$.

2.2 Network Diffusion Models

Let $G = (V, E)$ denote an undirected network. In all the models we study, each vertex $v \in V$ can be in state $x_v \in \{0, 1\}$, with state 0 denoting "inactive/uninfected/uninfluenced" and state 1 denoting "active/infected/influenced," depending on the application. We restrict ourselves to *monotone* or progressive processes, i.e., an infected node stays infected. Each node is associated with an activation function whose inputs include the states of its neighbors. This function computes the next state of the node. The diffusion process starts with a few vertices set as active/infected; we refer to this set as the *initial set* or the seed set. For an initial set of active nodes S, let $\sigma(S)$ denote the expected number of active nodes at termination. These models always reach fixed points. We consider the following models:

(1) Independent Cascade (IC) Model (Kempe et al., 2003): This model is a special case of the SIR model for epidemics. An infected node v infects each neighbor w with probability p (referred to as the *transmission probability*). Equivalently, each edge (v, w) can be *live* with probability p, independently of all other edges. All those nodes which are connected to the initial set through a *live path* are considered infected. In the graph, let (v, x) be an edge. Suppose v gets infected at time t, and x is in state 0. Then v tries to infect x with probability p at time $t + 1$. Irrespective of whether x gets infected by v, v remains in state 1 for all subsequent times, but never again tries to infect x.

(2) Linear Threshold (LT) Model (Kempe et al., 2005): Each node v has a threshold $\theta_v \in [0, 1]$, chosen uniformly at random. Node v is influenced by its neighbor w according to weight $b_{v,w}$ such that $\sum_{w \in N(v)} b_{v,w} \leq 1$. Node v becomes infected if $\sum_{w \in A(v)} b_{v,w} \geq \theta_v$, where $A(v) \subseteq N(v)$ is the set of neighbors of v which is currently infected. In our analysis and experiments, we assume that $b_{v,w} = 1/\deg(v, G)$ for all $w \in N(v)$, where $\deg(v, G)$ is the degree of v in G. This means that v is influenced equally by all its neighbors. This model was considered in (Kempe et al., 2003). In the perturbed graph $G' = G + R_u(\epsilon)$, $b_{v,w} = 1/\deg(v, G')$, where $\deg(v, G')$ is the new degree of v.

3. Analyzing the Sensitivity of the IC Model

In this section, we rigorously analyze the sensitivity of the IC model to edge perturbations. As mentioned earlier, we restrict our attention to uniform edge addition model, i.e., the perturbed graph is obtained by adding an edge between every pair of vertices with probability ϵ/n.

Before stating the main result, we discuss some aspects of the IC model and develop notation which will be used in the analysis. For a transmission probability p, let $G(p)$ denote a random subgraph of G obtained by retaining edges of G with probability p. The relationship between infection spread in the IC model and the structure of $G(p)$ is well-known; the question of whether or not a large spread occurs in G is equivalent to asking if there exists a giant component in $G(p)$. Another interesting aspect is that the IC model exhibits a threshold phenomenon. For many graph families there exists a critical transmission probability p_c at which an abrupt phase transition occurs from a "small" spread at $p < p_c$ to a large one at $p > p_c$ with high probability. A formal definition of p_c follows (see (Bollobás, 1985) for more details).

Definition 3.1. For a graph G on n nodes, p_c is the critical transmission probability if starting with a single random initial infection, for any transmission probability $p \ll p_c$ the total number of infections is a.a.s. $o(n)$ or equivalently, all components in $G(p)$ are of size $o(n)$, while for $p \gg p_c$, the number of infections is a.a.s. $\Theta(n)$ or there exists a giant component in $G(p)$.

Recall from Section 2 that the notion of *asymptotically almost surely* is formally defined for a sequence of graphs. However, we do not state this explicitly, in order to reduce the notational overload. Now we state our main result of this section where we analyze the sensitivity of the IC model for a graph operating at probability $p < p_c$ and satisfying $pd_{\text{avg}} \ll 1$ where, d_{avg} is the average degree of G. We conclude with a discussion on the implications of this theorem.

Theorem 1. *Consider a graph G on n nodes with average degree d_{avg} and critical transmission probability p_c. Let the transmission probability p satisfy $pd_{\text{avg}} = o(1)$ and $p = \Omega(1/n^{1-\delta})$ for some constant δ. Let $G + R_u(\epsilon)$ be the perturbed graph obtained by adding edges uniformly at random with factor ϵ. Then, for a single seed node chosen uniformly at random, the following hold:*

(a) *The number of infections in G is $o(n)$ a.a.s. and therefore, $p < p_c$.*

(b) *If $pd_{\text{avg}} = o(1/\log^2 n)$, then, there exists a threshold perturbation factor $\epsilon_t = \frac{1}{p}$ such that for any positive constant $c < 1$, if $\epsilon \leq (1-c)\epsilon_t$, the number of infections in $G + R_u(\epsilon)$ for a transmission probability of p is a.a.s. $o(n)$ while if $\epsilon \geq (1+c)\epsilon_t$, the number of infections is $\Theta(n)$.*

(c) *For any positive constant c, if $\epsilon \geq \frac{1+c}{p}$, the number of infections in $G + R_u(\epsilon)$ is a.a.s. $\Theta(n)$.*

Proof. Let $G' = G + R_u(\epsilon)$ and let $G'(p)$ be the random subgraph of G' obtained by choosing edges with probability p. Since, for any edge $e \notin E(G)$, $\Pr(e \in G'(p)) = p \times \Pr(e \in R_u(\epsilon)) = \frac{\epsilon p}{n}$, it follows that $G'(p)$ can be obtained by adding edges between every pair of nodes in $G(p)$ with probability $\frac{\epsilon p}{n}$. Now, we will prove Statement (a).

Claim 1. The number of components in $G(p)$ with more than one node is $O(n\sqrt{pd_{\text{avg}}})$ a.a.s. and therefore, $G(p)$ has no giant component.

Proof. We use Chebychev's inequality. For any $x > d_{\text{avg}}$, the number of nodes with degree greater than x in G is at most $n\left(\frac{d_{\text{avg}}}{x+1}\right) < n\left(\frac{d_{\text{avg}}}{x}\right)$. Let n_{iso} denote the number of isolated nodes in $G(p)$. Then,

$$\begin{aligned}\mathbb{E}[n_{\text{iso}}] &= \sum_{v \in V(G)} \Pr(v \text{ is isolated}) \geq \sum_{v \in V(G), \deg(v,G) \leq x} \Pr(v \text{ is isolated}) \\ &\geq (1-p)^x n\left(1 - \frac{d_{\text{avg}}}{x}\right) \geqslant (1-px)n\left(1 - \frac{d_{\text{avg}}}{x}\right).\end{aligned}$$

Choosing x to be $\sqrt{d_{\text{avg}}/p}$, $\mathbb{E}[n_{\text{iso}}] \geq (1 - \sqrt{pd_{\text{avg}}})^2 n \geq (1 - 2\sqrt{pd_{\text{avg}}})n$. Since by assumption, $d_{\text{avg}} = o(1/p)$, $\mathbb{E}[n_{\text{iso}}] = (1 - o(1))n$. For any node $v \in V(G)$, let $I_v = 1$ if v has no neighbors in $G(p)$ and 0 otherwise. Let $\text{Var}[\cdot]$ and $\text{Cov}[\cdot]$ denote the variance and covariance, respectively. Using the bounds $\text{Var}[I_v] \leq \mathbb{E}[I_v]$ and $\text{Cov}[I_a I_b] = P(I_a \cap I_b) \leq 1$,

$$\begin{aligned}\text{Var}[n_{\text{iso}}] &= \sum_{v \in V(G)} \text{Var}[I_v] + 2 \sum_{(a,b) \in E(G)} \text{Cov}[I_a I_b] \\ &\leq \mathbb{E}[n_{\text{iso}}] + 2 \sum_{(a,b) \in E(G)} \text{Cov}[I_a I_b] \leq \mathbb{E}[n_{\text{iso}}] + nd_{\text{avg}} = O(nd_{\text{avg}}).\end{aligned}$$

Since we have assumed that $p = \Omega(1/n^{1-\delta})$, it follows that $\text{Var}[n_{\text{iso}}] = O(n^{2-\delta} pd_{\text{avg}})$. Applying Chebychev's inequality,

$$P\left(|n_{\text{iso}} - \mathbb{E}[n_{\text{iso}}]| > n\sqrt{pd_{\text{avg}}}\right) \leq P\left(|n_{\text{iso}} - \mathbb{E}[n_{\text{iso}}]| > \sqrt{n^\delta \text{Var}[n_{\text{iso}}]}\right) \leq \frac{1}{n^\delta}.$$

Therefore, a.a.s., $n - n_{\text{iso}} \leq 3n\sqrt{d_{\text{avg}} p}$. ∎

Let $\{C_i \mid i \in N\}$ be the set of connected components in $G(p)$, where N is the number of components and let n_i denote the size of C_i. The probability that components C_i and C_j are connected by at least one edge of $R_u(\epsilon)$ in $G'(p)$ is at most $\frac{n_i n_j \epsilon p}{n}$. Let S_{iso} denote the set of isolated nodes in $G(p)$ and $\overline{S}_{\text{iso}}$ the set of remaining nodes.

Let H be the graph obtained by adding a special vertex v to $G'(p)$ and making it adjacent to all nodes in $\overline{S}_{\text{iso}}$. Clearly, $\overline{S}_{\text{iso}}$ belongs to a component in H and any component in $G'(p)$ is contained in a component of H. This implies that $G'(p)$ has a giant component only if H has one. Now, we will prove the first part of Statement (b).

Claim 2. *H has components of size $o(n)$ if $pd_{\text{avg}} = o(1/\log^2 n)$ and $\epsilon \leq \frac{(1-c)}{p}$.*

Proof. Let $H[S_{\text{iso}}]$ and $H[\overline{S}_{\text{iso}}]$ be graphs induced by S_{iso} and $\overline{S}_{\text{iso}}$, respectively, in H. Note that since S_{iso} is a set of isolated nodes in $G(p)$, $H[S_{\text{iso}}]$ is an Erdős-Rényi graph on $|S_{\text{iso}}|$ nodes with edge probability $\frac{\epsilon p}{|S_{\text{iso}}|} = \frac{\epsilon p}{n(1-o(1))}$. Since $\epsilon p < 1 - c$, it follows that $H[S_{\text{iso}}]$ has components of size $O(\log n)$ (see, e.g., Bollobás, 1985). Now we will show that the component containing $\overline{S}_{\text{iso}}$ is of size $o(n)$ a.a.s., thus completing the proof.

Let $N(\overline{S}_{\text{iso}})$ denote the size of the neighborhood of $\overline{S}_{\text{iso}}$ in H. The probability that a node in S_{iso} is a neighbor of $\overline{S}_{\text{iso}}$ is at most $\frac{|\overline{S}_{\text{iso}}|\epsilon p}{n}$. Therefore,

$$\mathbb{E}\Big[N(\overline{S}_{\text{iso}})\Big] \leq |S_{\text{iso}}| \times \frac{|\overline{S}_{\text{iso}}|\epsilon p}{n} \leq |\overline{S}_{\text{iso}}|\epsilon p.$$

Applying a version of Chernoff bound (Chung & Lu, 2002), we have the following:

$$P\Big(N(\overline{S}_{\text{iso}}) - \mathbb{E}\Big[N(\overline{S}_{\text{iso}})\Big] > \alpha\Big) \leq \exp\left(-\frac{\alpha^2}{2\Big(\mathbb{E}\Big[N(\overline{S}_{\text{iso}})\Big] + \alpha/3\Big)}\right).$$

We have two regimes to consider: (i) $|\overline{S}_{\text{iso}}|\epsilon p = \omega(1)$ and (ii) $|\overline{S}_{\text{iso}}|\epsilon p = O(1)$. If $|\overline{S}_{\text{iso}}|\epsilon p = \omega(1)$, setting $\alpha = |\overline{S}_{\text{iso}}|\epsilon p$, it follows that $N(\overline{S}_{\text{iso}}) \leq 2|\overline{S}_{\text{iso}}|\epsilon p$ a.a.s. Since $H[S_{\text{iso}}]$ has components of size $O(\log n)$, the size of the component containing $\overline{S}_{\text{iso}}$ is

$$O\big(|\overline{S}_{\text{iso}}| + N(\overline{S}_{\text{iso}})\log n\big) = O\big(|\overline{S}_{\text{iso}}| + |\overline{S}_{\text{iso}}|\epsilon p \log n\big) = O\Big(n\sqrt{pd_{\text{avg}}}(1 + \epsilon p \log n)\Big).$$

The last expression follows from Claim 1. Since $\epsilon \leq (1-c)\epsilon_t$ and $pd_{\text{avg}} = o(1/\log^2 n)$, it follows that $n\sqrt{pd_{\text{avg}}}\epsilon p \log n \leq n\sqrt{pd_{\text{avg}}}\log n(1-c) = o(n)$. Now we consider regime (ii): $|\overline{S}_{\text{iso}}|\epsilon p = O(1)$. By setting $\alpha = \log n$, the component size is $O\big(\log^2 n\big) = o(n)$. Hence, we have proved the claim. ■

The proofs of the second part of Statement (b) and Statement (c) are straightforward. Let R_p denote the random subgraph of the perturbation network $R_u(\epsilon)$ obtained by sampling its edges with probability p. It is easy to see that R_p is an Erdős-Rényi graph with edge probability $\frac{\epsilon p}{n} = \frac{1+c}{n}$ which implies that it has a giant component (Bollobás, 1985). This in turn implies that $G'(p)$ has a giant component. Hence, we have proved the theorem. □

Theorem 1 indicates that for a large class of networks, the closer we operate to p_c, the more sensitive the dynamics is to structural perturbation. This is indeed true if the conditions of Statement (b) are met: for any $p_1 < p_2 < p_c$, $\epsilon_t(p_1) > \epsilon_t(p_2)$. It implies that greater perturbation is required in the case of p_1 (compared to p_2) to observe a significant difference in the expected infection size after perturbation. We have observed the same in our experiments; see the last three columns of Table 1. In the AstroPh graph, for example, at $p = 0.03$, phase transition occurs for $\epsilon_t = 8$ while for $p = 0.02$, ϵ_t has to be greater than 20 for a transition to occur.

4. Analyzing the Sensitivity of the LT Model

We now analyze the impact of edge perturbations on the LT model on a graph $G = (V, E)$. As in the previous section, we restrict our attention to the uniform edge addition model. Recall that in the specific version of the LT model we consider here, we set $b_{v,w} = 1/\deg(v)$ for each node $v \in V$ and $w \in N(v)$.

The fixed points and the number of infected nodes can be studied through an elegant random graph model (Kempe et al., 2003) which we describe here. Construct a random directed graph $H_{LT} = (V, E')$ in the following manner: for each node $v \in V$, a neighbor w is chosen with probability $b_{v,w}$ and a directed edge is added from w to v. Figure 1 illustrates a graph G and an instance of H_{LT}. Note that even though G is undirected, H_{LT} is a directed graph. For a set $S \subset V$, let $\sigma(S, H_{LT})$ denote the number of nodes reachable from S in H_{LT} (including those in S). Then, $\sigma(S)$, the expected number of infections with a starting set S, satisfies $\sigma(S) = \sum_{H_{LT}} \Pr[H_{LT}]\sigma(S, H_{LT})$ (Kempe et al., 2003). We use this characterization to analyze the impact of edge perturbations.

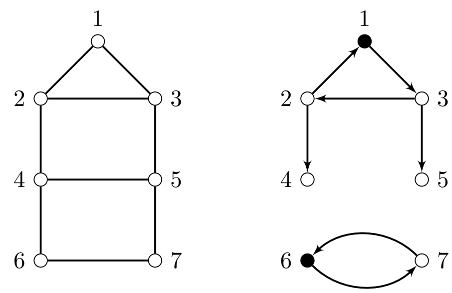

Figure 1: A graph (on the left) and an instance of the random graph H_{LT} (on the right) corresponding to the LT model. For the component T induced by $\{1, 2, 3, 4, 5\}$, 1 is chosen as the root and as a result, $T_0 = \{1\}$, $T_1 = \{3\}$, $T_2 = \{2, 5\}$ and $T_3 = \{4\}$.

The random graph H_{LT} constructed by the above process has the following structure: in each connected component T of H_{LT}, every vertex has one incoming edge and therefore, there exists exactly one directed cycle. If we choose a vertex in the cycle as the *root* r and remove its incoming edge, then, T corresponds to a tree rooted at r with all edges oriented away from r. T can be partitioned into sets $T_0, \dots, T_k$ such that for each $i > 0$, a vertex $v \in T_i$ has an incoming edge from some vertex $u \in T_{i-1}$. The set T_0 is a singleton consisting of the root vertex r. The incoming edge for r is from some neighbor in $\cup_{i=1}^{k} T_i$. This is illustrated in Figure 1. First, we will show the following:

Lemma 4.1. *In the LT model, let $\delta = \min_{v \in V, w \in N(v)} b_{v,w}$. Each component in the random subgraph H_{LT} has depth $O\left(\frac{1}{\delta} \log n\right)$ with probability at least $1 - \frac{1}{n^3}$.*

Proof. Consider a component T in H_{LT}, which is partitioned into sets $T_0, \dots, T_k$, as mentioned above. For any $i = 1, \dots, k-1$, a vertex $v \in T_i$ would become a root if it chooses an incoming edge from one of its descendants. The probability of this event is at least $\min_{w \in N(v)} b_{v,w} \geq \delta$. Therefore, the probability that none of the vertices in T_i becomes a root is at most $1 - \delta$, which in turn implies that the probability that none of the vertices in T_i, for $i = 1, \dots, k-1$ becomes a root is at most $(1-\delta)^{k-1}$. Hence, the probability that T has depth more than $k = 4 \cdot \frac{1}{\delta} \log n + 1$ is at most $\sum_{k \geq 4 \cdot \frac{1}{\delta} \log n + 1}^{n} (1-\delta)^{k-1} \leq \frac{1}{n^4}$. Since there are at most n such components in H_{LT}, the probability that any of these has depth more than $O\left(\frac{1}{\delta} \log n\right)$ is at most $\frac{1}{n^3}$. □

Consider a vertex v contained in a component T. Let $n(v, T)$ denote the number of vertices reachable from v in T. Then, the number of infections resulting from v is the expected value of $n(v, T)$, averaged over all random subgraphs H_{LT} and components containing v. Let $A(T) = \frac{1}{|T|} \sum_{v \in T} n(v, T)$. Conditioned on a random subgraph H_{LT}, the average number of infections starting at a random source is $\sum_{T \in H_{LT}} A(T) \frac{|T|}{n}$. The average number of infections starting at a random source is $\sum_{H_{LT}} \Pr[H_{LT}] \sum_{T \in H_{LT}} A(T) \frac{|T|}{n}$.

Lemma 4.2. *For each component T in a random subgraph H_{LT}, $A(T) \leq 2d$, where d is the depth of T.*

Proof. Define $\hat{T}$ to be the tree obtained by removing the incoming edge for the root in T. As described above, $\hat{T}$ is an out-tree. For each $v \in \hat{T}$, we define $n(v, \hat{T})$ as the number of vertices reachable from v in $\hat{T}$ — this corresponds to the size of the subtree rooted at v in $\hat{T}$. We define $A(\hat{T}) = \frac{1}{|\hat{T}|}\sum_{v\in\hat{T}} n(v, \hat{T})$, and prove that $A(\hat{T}) \leq d$ by induction on the depth of the out-tree. The base case is a leaf node u, for which $A(u) = 1$.

Let r be the root of $\hat{T}$. Suppose it has children $v_1, \ldots, v_a$. Let $\hat{T}_i$ be the subtree rooted at v_i, and let n_i be the number of vertices in $\hat{T}_i$. By induction, $A(\hat{T}_i) = \frac{1}{n_i}\sum_{v\in\hat{T}_i} n(v, \hat{T}_i) \leq d-1$.

$$\begin{aligned}
A(\hat{T}) &= \frac{1}{|\hat{T}|}\sum_{v\in\hat{T}} n(v, \hat{T}) \\
&= \frac{1}{|\hat{T}|} n(r, \hat{T}) + \sum_{i=1}^{a} \frac{1}{|\hat{T}|}\sum_{v\in\hat{T}_i} n(v, \hat{T}_i) \\
&= 1 + \sum_{i=1}^{a} \frac{n_i}{|\hat{T}|} A(\hat{T}_i) \leq 1 + \sum_{i=1}^{a} \frac{n_i}{|\hat{T}|}(d-1) \\
&\leq 1 + \frac{|\hat{T}|-1}{|\hat{T}|}(d-1) \leq d
\end{aligned}$$

The third equality follows because $n(r, \hat{T}) = |\hat{T}|$, and by definition, $A(\hat{T}_i) = \frac{1}{n_i}\sum_{v\in\hat{T}_i} n(v, \hat{T}_i)$. The first inequality follows by the induction hypothesis, since the depth of each $\hat{T}_i \leq d-1$. The second inequality follows because $\sum_{i=1}^{a} n_i = |\hat{T}| - 1$.

Next, we consider $A(T)$. We recall that T is a tree with a cycle of length at most d. Let the cycle consist of vertices $u_0 = r, u_1, \ldots, u_b$, with $b \leq d-1$. For each u_i, $n(u_i, T) = |T|$, since there is a path from u_i to r. For every other vertex $u \neq u_i$ in T, $n(u, T) = n(u, \hat{T})$. This implies that $A(T) \leq \frac{d|T|}{|T|} + A(\hat{T}) \leq 2d$. □

Finally, we bound the number of infections in the perturbed graph below.

Theorem 2. *Let $G(V, E)$ be a graph with maximum degree Δ and $G + R_u(\epsilon)$ be the perturbed graph obtained by adding edges uniformly at random with factor ϵ. For the LT model where $b_{v,w} = 1/\deg(v)$ for each node $v \in V$ and $w \in N(v)$, the expected number of infected vertices starting with an initial random seed set of size s in the perturbed graph $G + R_u(\epsilon)$ is $O(s(\Delta + \epsilon + \log n)\log n)$.*

Proof. By a direct application of the Chernoff bound, it can be shown that with probability at least $1 - \frac{1}{n^3}$, the maximum degree of $G' = G + R_u(\epsilon)$ is at most $\Delta + \epsilon + c \cdot \log n$ for a constant c with the remaining probability of $\frac{1}{n^3}$, the maximum degree is $O(n)$. We consider the random graph process to generate a subgraph H_{LT} of G'. Since $b_{v,w} = 1/\deg(v)$ for each node $v \in V$ and $w \in N(v)$, for this model, the value of δ of Lemma 4.1 is $1/\Delta(G')$ and therefore, each component in H_{LT} has depth at most $O((\Delta + \epsilon + \log n)\log n)$, with probability at least $1 - \frac{1}{n^3}$. Conditioned on H_{LT} satisfying this bound on the depth, $A(T) = O((\Delta + \epsilon + \log n)\log n)$ for all $T \in H_{LT}$. If H_{LT} does not satisfy the depth bound, $A(T) = O(n)$ for all $T \in H_{LT}$. Therefore, the expected number of infections for a single random seed is $O((\Delta + \epsilon + \log n)\log n) + O(\frac{n}{n^3}) = O((\Delta + \epsilon + \log n)\log n)$. The result extends to $s > 1$ by submodularity of the expected number of infections. □

Theorem 2 implies that the LT model with uniform edge weights is in general very robust to perturbation. Note that the final part of the proof of Theorem 2 is essentially based on the maximum degree of G'. Replacing G' with G and retracing the steps, one can show that the expected spread in the unperturbed graph G is $O(s\Delta \log n)$, where Δ is its maximum degree. Hence, we see that for a reasonably high value of Δ (say $\Omega(\log n)$), the difference between the total number of infections in G and G' is at most a linear function of ϵ and therefore there is no abrupt change in the outcome. Moreover, the bound suggests that the higher the Δ, the lower the effect of perturbation. However, making this formal would require obtaining good lower bounds on the expected number of infections, which we leave as an interesting research question.

5. Experimental Results

We study the sensitivity to edge perturbations on twenty diverse real-world networks (Leskovec, 2011) with varying degrees of perturbation and other factors for both IC and LT models. They are listed in Table 1 along with some of their properties, the first four of which are number of nodes, average degree, maximum eigenvalue, and maximum degree. Other properties will be discussed subsequently. We present representative results for selected networks, with other networks exhibiting the same behavior unless stated otherwise. We focus primarily on the sensitivity of expected number of infections and transient behavior to edge perturbation. Of course, all of our observations are restricted to the conditions for which the experiments were performed.

5.1 Experimental Setup and Methodology

Each network G in Table 1 was perturbed with values of ϵ ranging from 0 to 100, where $\epsilon = 0$ corresponds to the unperturbed network. For each ϵ, we generated ten graph instances $G' = G + R$ or $G \oplus R$. Here, R may be R_u or R_d, depending on whether the perturbation is a uniform edge approach or a degree-assortative approach. For each graph instance, we performed a simulation run, which consists of 100 separate diffusion instances. A diffusion instance is the process of setting all node states initially to zero, assigning relevant properties to graph entities (e.g., transmission probability on edges for the IC model and edge weights and node thresholds for the LT model), selecting a seed node set S whose element states are changed to 1 (i.e., are initially infected), and marching time forward in discrete units, continuing the simulation until a fixed point is reached. We record each node's time of infection in each diffusion instance. Thus, for example, the experimental data displayed as average and variance quantities are based on 1000 values.

5.2 Edge Additions vs. Deletions

We find that perturbations involving both edge additions and deletions do not alter the results by much, compared to perturbations involving just edge additions. This is due to the sparsity of the graphs considered. For example, for uniform perturbations, the expected number of edges deleted is $|E|\epsilon/n = \epsilon d_{\text{avg}}/2$ while the expected number of edges modified is $\epsilon n/2$. Therefore, the remainder of the paper focuses on perturbation by addition of edges only; i.e., only graphs of the form $G' = G + R$.

Network	$n = \|V\|$	d_{avg}	λ	Δ	$\approx p_c$	$(p;\epsilon_t)$ pairs by experiments		
Synthetic graphs								
Random-Regular-20	10000	20.0	20.0	20	0.05	0.04;9	0.03;>10	0.02;>20
Autonomous Systems								
AS-2000-01-02	6474	3.88	46.31	1458	0.2	0.15;2	0.1;5	0.05;>10
Oregon1-01-03-31	10670	4.12	58.72	2312	0.2	0.15;<2	0.1;5	0.05;10
Oregon2-01-03-31	10900	5.72	70.74	2343	0.2	0.15;<1	0.1;5	0.05;10
Co-authorship								
AstroPh	17903	22.0	94.42	504	0.04	0.03;8	0.02;>20	0.01;>40
CondMat	21363	8.54	37.88	279	0.12	0.1;2	0.08;4	0.06;8
Grqc	4158	6.45	45.61	81	0.2	0.15;<2	0.1;5	0.05;10
HepPh	11204	20.99	244.93	491	0.1	0.07;<2	0.04;8	0.01;>20
HepTh	8638	5.74	31.03	65	0.2	0.15;<1	0.1;5	0.05;10
Citation								
HepPh	34546	24.46	76.58	846	0.04	0.03;6	0.02;>20	0.01;>40
HepTh	27770	25.37	111.25	2468	0.05	0.04;<2	0.03;8	0.02;>40
Communication								
Email-Enron	33696	10.02	118.41	1383	0.1	0.07;4	0.04;10	0.01;>20
Email-EuAll	265214	2.74	102.53	7636	0.3	0.2;<2	0.1;6	0.05;10
Social								
Epinion	75877	10.69	184.17	3044	0.15	0.12;<1	0.1;2	0.08;4
Slashdot0811	77360	12.12	131.34	2539	0.1	0.07;3	0.04;9	0.01;>20
Slashdot0902	82168	12.27	134.62	2552	0.1	0.07;3	0.04;9	0.01;>20
Twitter	22405	5.34	54.08	888	0.2	0.15;<2	0.1;4	0.05;10
Wiki-Vote	7066	28.51	138.15	1065	0.04	0.03;3	0.02;10	0.01;>20
Internet peer-to-peer								
Gnutella04	10876	7.35	15.7	103	0.125	0.1;2	0.07;7	0.04;>10
Gnutella24	26518	4.93	19.06	355	0.2	0.15;1	0.1;6	0.05;>10

Table 1: Some relevant properties of the networks used in our simulations and results from experiments.

5.3 The Independent Cascade Model

We have experimental results for both uniform and degree-assortative perturbation, about which we make observations relating to the theoretical results and about the behavior of the networks in general.

5.3.1 Uniform Perturbation

Effect of ϵp on final infection size. Figure 2 consists of plots of variation in the average and variance of the fraction of infected nodes with (i) the transmission probability p for various levels of perturbation and (ii) the perturbation level ϵ for various p values for three networks. The plots for the remaining networks are in the full version (Adiga, Kuhlman, Mortveit, & Vullikanti, 2014). We note that when $\epsilon p > 1$ (and $p \geq p_c$), the average infection size is generally high, agreeing with Statement (c) of Theorem 1. We observe that in all the plots the final fraction of infections is at least 0.5 when $\epsilon p > 1$.

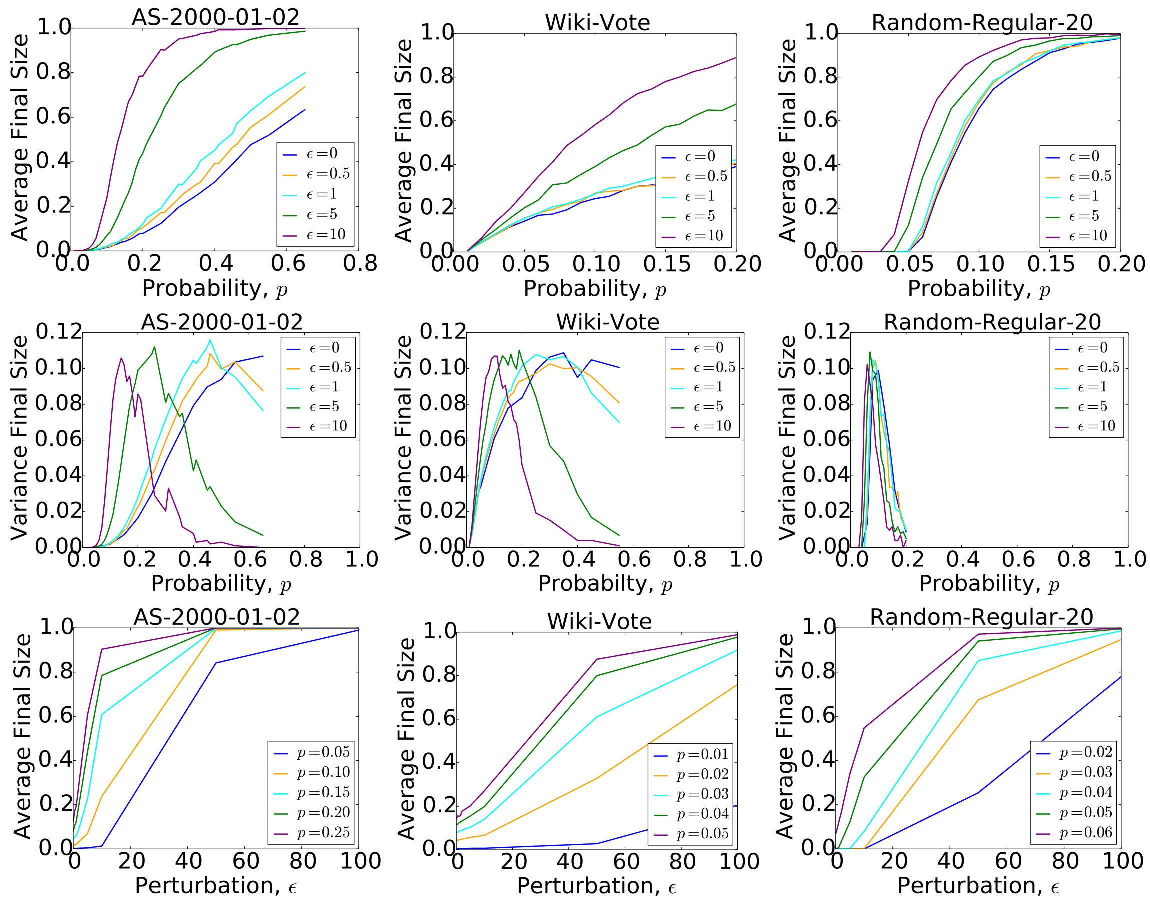

Figure 2: Uniform perturbation (IC model): Average and variance of fraction of infections for a single random seed (i) vs. transmission probability p for various ϵ values and (ii) vs. ϵ for various values of p. The plots in the first and third rows are average final fraction of infected nodes and those in the second row are average variance in final fraction of infected nodes. The remaining plots are in the full version (Adiga et al., 2014).

The sensitivity of final infection size with ϵ is modulated by d_{avg}. We first consider the relationship between p_c and d_{avg} of the unperturbed networks. Since for finite networks, there is no clear definition of p_c, we chose it to be that value of p for which the average total number of infections is 10% of the number of nodes in the network. Table 1 contains p_c for each network. The dependence of p_c on d_{avg} is shown in Figure 3 where each data point corresponds to one graph and is colored according to the type of graph. Clearly, the plot is indicating an inverse relationship between the two values. This implies that the higher the d_{avg}, the lower the p for which the spread is small. Note that Theorem 1 shows that the minimum ϵ for which there is a large spread is inversely proportional to p. This strongly suggests that the greater the edge density, the greater the perturbation required to achieve significant change in dynamics. This is supported by our experiments, too. For example, see plots for AS-2000-01-02 and Wiki-Vote in the first row of Figure 2. Consider $p = 0.2$ and note the change in spread in going from $\epsilon = 0$ to 10. For Wiki-Vote, it goes from 0.38 to 0.89, for a ratio of 2.3. For AS-2000-01-02, it changes from 0.08 to 0.82, for a ratio of 10.3. Thus, AS-2000-01-02, with a factor of 7 smaller d_{avg} than Wiki-Vote, has much greater sensitivity of spread for changes in ϵ. Note the numbers of nodes in the two graphs are comparable. We see this same behavior in other (high d_{avg}, low d_{avg}) network pairs: (AstroPh, CondMat), (HepPh, HepTh), and (Email-Enron, Email-EuAll).

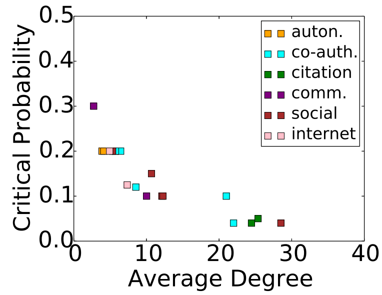

Figure 3: Dependence of p_c on d_{avg}. Data are for the graphs in Table 1, and colors correspond to graph types in the table.

Variance of final infection size. In Figure 2 (and more plots in the full version Adiga et al., 2014), the middle row of plots contains variance in final infection size as a function of transmission probability. There are several observations. First, for all graphs whose dynamics exhibit a phase transition to larger infection sizes with increasing ϵ for fixed p (and for increasing p for fixed ϵ), the variance of the infection size qualitatively increases with p, peaks in the region of phase transition and again decreases. The peaks in variance correspond to the regimes where the change in final infected fractions are changing the most with p, as expected (cf., the first row of figures). Second, the greater the perturbation, the lesser the range of p values for which the variance is high. This is because the greater the value of ϵ, the faster the contagion spreads, thus driving down variance. The last observation is non-intuitive. That is, the peak of the variance does not seem to vary for combinations of ϵ and p where $0 \leq \epsilon \leq 10$ and $0 \leq p \leq 0.6$. For these ranges, only the value of p where the peak occurs decreases with increasing ϵ. In fact, the peak of the variance is around 0.1 for all graphs for these conditions.

Effect of regular network structure on numbers of infected nodes. There are several reasons to investigate random networks with uniform degree. First, in an investigation of voter model dynamics (Kuhlman, Kumar, & Ravi, 2013), it was shown that uniform degree networks generated results very near to those of realistic graphs whose degree distribution was exponential decay, but far from those for graphs with scale free degree distributions. So the question arises here as to how close the behavior of uniform degree networks is to those of realistic networks. Second, since the perturbing subgraph R in this study is a random graph, the perturbed graph G' in some sense maintains its random structure compared to G, when G is the random 20-regular graph (a random graph where each node has degree 20). Figure 2 shows that this latter consideration dominates. That is, the upper right plot shows that the curves are basically shifted left as ϵ increases. So, too, in the plot at the right of the middle row, the variance curves shift slightly left as ϵ increases, but this effect is much smaller than in Wiki-Vote or AS-2000-01-02.

Effect of ϵ on the average time-to-peak number of new infections. The average time histories for 1000 diffusion instances that comprise each curve for one graph of each type are provided in the top row of Figure 4. The transmission probabilities used for each plot are such that $p > p_c$. The average time at which the maximum number of new infections occurs can decrease, stay the same, or increase as a function of ϵ, depending on the graph, moving left to right. Over all graphs of Table 1, there seems to be a fairly even split in that nine of the graphs show an increase in average time-to-peak with increasing ϵ and eight show a decrease in average time-to-peak. The remainder show no change in time-to-peak.

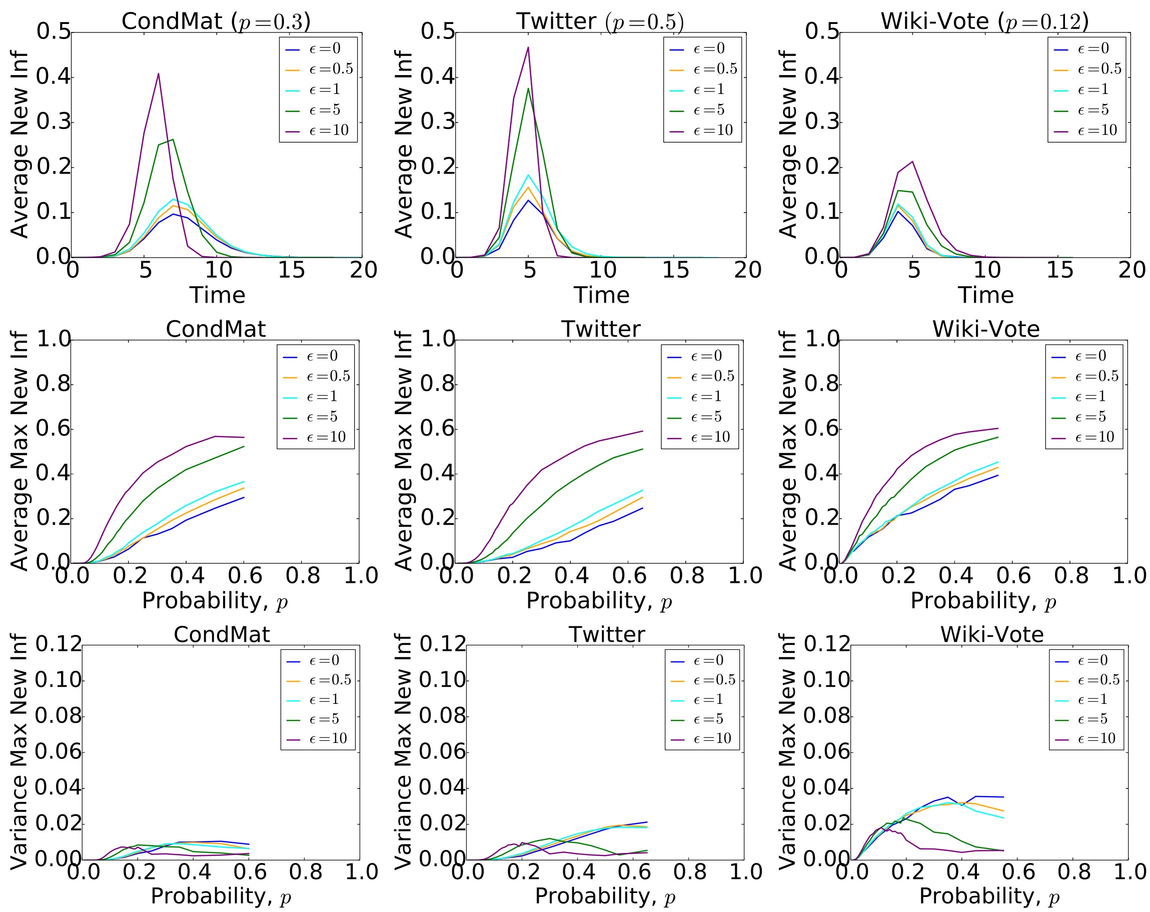

Figure 4: Sensitivity of temporal characteristics to uniform perturbation (IC model): Plots of (i) average number of new infections for each time step for selected p values, (ii) average of maximum number of new infections at any time vs. p. and (iii) variance of maximum number of new infections at any time vs. p. The remaining plots are in the full version (Adiga et al., 2014).

Effect of ϵ on the average peak number of new infections. The second row of plots in Figure 4 depicts the average maximum number of new infections at any one time as a function of transmission probability, for different ϵ. As ϵ increases, the maximum number of new infections increase.

Variance in average maximum number of new infections for any discrete time. The last row of plots in Figure 4 provides variance in the 1000 experimentally determined values used to compute the average maximum number of new infections; that is, in the peak value of each curve in the plots of the top row. For these particular graphs, the variances are very roughly on the order of 0.01, which is, interestingly, much smaller than those for the final number of infections in Figure 2.

5.3.2 Degree-Assortative Perturbation

Here, we focus on results that are different from those for uniform perturbation.

Effect of ϵ and degree-assortativity on average final number of infections and variance. The first row of plots in Figure 5 shows the average final number of infections, from 1000 measurements, as a function of p and ϵ. The $\epsilon = 0$ curves are the same as those in Figure 2. By comparison with the top row of plots in that figure, it is clear that degree-assortativity perturbations significantly reduce the effect of ϵ on changes in the average final number of infections. We find, over all networks, that AS-2000-01-02 and Wiki-Vote provide two bounding cases, i.e., least effect and most effect of degree-assortative perturbations, respectively. From comparison of the second row of plots in each of Figure 5 and Figure 2, it is also clear that degree-assortative perturbations correspondingly collapse the variances across ϵ values for Wiki-Vote, but those for AS-2000-01-02 are less affected.

Effect of perturbation method on average final number of infections. Since the $\epsilon = 0$ curves are the same in Figure 2 and Figure 5, when all other factors are the same, uniform perturbations generate greater numbers of infections than do degree-assortative perturbations (compare the top rows of plots in the two figures).

Effect of network structure for degree-assortative perturbations. The effects of degree assortative perturbations appear to be more network-specific than those for uniform perturbations. This is expected since the perturbation instances inherit some of the network properties.

Effect of degree-assortative perturbations on the random regular graph. The effects of the two perturbation methods on the random regular graphs are the same because of the uniform node degrees.

5.4 Linear Threshold Model

We have results for the effect of uniform perturbation in the LT model. Figure 6 shows the plots of average number of infections vs. ϵ for three representative networks for different seed probabilities s. The remaining plots are in the full version (Adiga et al., 2014). In each diffusion instance, the seed set was constructed by sampling the vertex set uniformly with

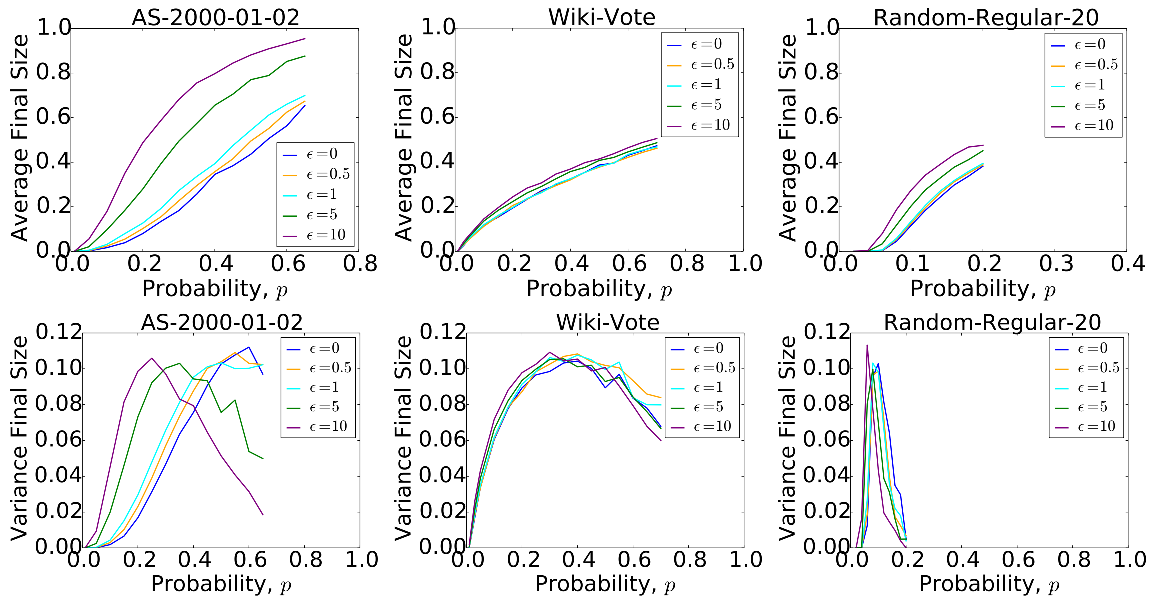

Figure 5: Degree-assortative perturbation (IC model): Average and variance of fraction of infections for a single random seed vs. transmission probability p for various ϵ values. The remaining plots are in the full version (Adiga et al., 2014).

probability s, and every node was assigned a threshold chosen uniformly at random in the interval $[0, 1]$.

Recall from Theorem 2 that the spread is bounded by a linear function of ϵ. From Figure 6, we observe that the model is generally robust to perturbation and the spread is a linear or sublinear function of ϵ depending on the seed probability s. In particular, when the seed probability is low ($s = 0.0001$ for example), there is hardly any change in the average spread. Note that in the plots, ϵ ranges from 0 all the way up to 100 which is an extreme value for perturbation considering the sizes of the networks. However, for larger values of s, especially when it is comparable to the maximum degree Δ, the effect of ϵ is more pronounced, even for low values of ϵ. This observation is worth investigating and could lead to better bounds for spread than what we have in Theorem 2.

6. Conclusions and Open Problems

We give the first rigorous results on the stability of the independent cascade and linear threshold models with respect to edge perturbations. We considered two popular noise models namely, uniform and degree-assortative perturbations, and studied the sensitivity of the final outbreak size and temporal characteristics to these perturbations. Our analysis was supported by experimental observations on 20 diverse real networks. We showed that the sensitivity of the independent cascade model depends on the transmission probability and perturbation can lead to abrupt changes in the outcome, while the linear threshold model with uniform edge weights is in general stable to network perturbations. Also, our experiments suggest that dynamics are more sensitive to uniform perturbation than

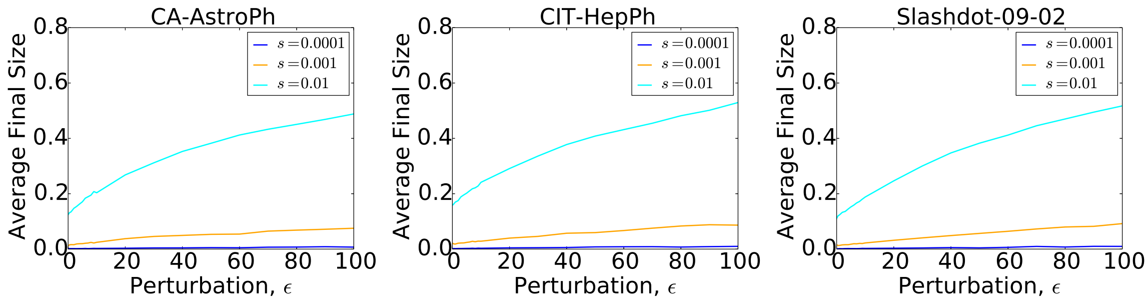

Figure 6: Uniform perturbation (LT model): Average fraction of infected nodes vs. perturbation ϵ for various seed probabilities s. The remaining plots are in the full version (Adiga et al., 2014).

degree-assortative. Extending our results to other models of noise, especially those involving dependencies, sensitivity to the number of sources, and examining the sensitivity of other dynamical properties in more general diffusion models (including the IC and LT models with heterogeneous probabilities or weights) are natural open problems for future research.

7. Acknowledgments

This work has been partially supported by the following grants: DTRA Grant HDTRA1-11-1-0016, DTRA CNIMS Contract HDTRA1-11-D-0016-0010, NSF Career CNS 0845700, NSF ICES CCF-1216000, NSF NETSE Grant CNS-1011769 and DOE DE-SC0003957. Also supported by the Intelligence Advanced Research Projects Activity (IARPA) via Department of Interior National Business Center (DoI/NBC) contract number D12PC000337, the US Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright annotation thereon.
Disclaimer: The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of IARPA, DoI/NBC, or the US Government.

A preliminary version of this paper appeared in Twenty-Seventh AAAI Conference on Artificial Intelligence (2013).

References

Achlioptas, D., Clauset, A., Kempe, D., & Moore, C. (2005). On the bias of traceroute sampling: or, power-law degree distributions in regular graphs. In *Proceedings of the thirty-seventh annual ACM Symposium on Theory of Computing*, pp. 694–703. ACM.

Adiga, A., Kuhlman, C., Mortveit, H. S., & Vullikanti, A. K. S. (2014). Sensitivity of diffusion dynamics to network uncertainty. Technical report, available at http://ndssl.vbi.vt.edu/supplementary-info/vskumar/sensitivity-jair.pdf.

Bakshy, E., Hofman, J. M., Mason, W. A., & Watts, D. J. (2011). Everyone's an influencer: quantifying influence on Twitter. In *Proceedings of the fourth ACM international* *conference on Web search and data mining*, pp. 65–74. ACM.

Bollobás, B. (1985). *Random graphs*. Academic Press.

Borgatti, S., Carley, K., & Krackhardt, D. (2006). On the robustness of centrality measures under conditions of imperfect data. *Social Networks*, *28*, 124–136.

Chen, J. (2010). The effects of demographic and spatial variability on epidemics: A comparison between Beijing, Delhi and Los Angeles. In *Conf. on Crit. Inf.*

Chung, F., & Lu, L. (2002). Connected components in random graphs with given expected degree sequences.. *Annals of Combinatorics*, *6*, 125–145.

Clauset, A., Moore, C., & Newman, M. (2008). Hierarchical structure and the prediction of missing links in networks. *Nature*, *453*, 98–101.

Costenbader, E., & Valente, T. (2003). The stability of centrality measures when networks are sampled. *Social Networks*, *25*, 283–307.

Eubank, S. (2010). Detail in network models of epidemiology: are we there yet?. *Journal of Biological Dynamics*, *4(5)*, 446–455.

Faloutsos, M., Faloutsos, P., & Faloutsos, C. (1999). On power-law relationships of the internet topology. In *SIGCOMM*, Vol. 29, pp. 251–262.

Flaxman, A., & Frieze, A. M. (2004). The diameter of randomly perturbed digraphs and some applications. In *APPROX-RANDOM*, pp. 345–356.

Flaxman, A. (2007). Expansion and lack thereof in randomly perturbed graphs. *Internet Mathematics*, *4*(2-3), 131–147.

Galuba, W. (2010). Outtweeting the twitterers - predicting information cascades in microblogs. In *WOSN*.

Goldenberg, J., Libai, B., & Muller, E. (2001). Talk of the network: A complex systems look at the underlying process of word-of-mouth. *Marketing Letters*.

González-Bailón, S., Borge-Holthoefer, J., Rivero, A., & Moreno, Y. (2011). The dynamics of protest recruitment through an online network. *Scientific Reports*, *1*.

Hagmann, P. (2008). Mapping the structural core of human cerebral cortex. *PLoS Biol*, *6(7)*.

Kaneko, K. (1985). Spatiotemporal intermittency in coupled map lattices. *Progress of Theoretical Physics*, *74*(5), 1033–1044.

Kempe, D., Kleinberg, J., & Tardos, É. (2003). Maximizing the spread of influence through a social network. In *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 137–146. ACM.

Kempe, D., Kleinberg, J., & Tardos, É. (2005). Influential nodes in a diffusion model for social networks. In *Automata, languages and programming*, pp. 1127–1138. Springer.

Kuhlman, C., Kumar, V., & Ravi, S. (2013). Controlling opinion propagation in online networks. *Journal of Computer Networks*, *57*, 2121–2132.

Lahiri, M., Maiya, A. S., Caceres, R. S., Habiba, & Berger-Wolf, T. Y. (2008). The impact of structural changes on predictions of diffusion in networks. In *ICDM*, pp. 939–948.

Leskovec, J. (2011). Stanford network analysis project. http://snap.stanford.edu/index.html.

Newman, M. (2003). The structure and function of complex networks. *SIAM Review*, *45*(2), 167–256.

Romero, D., Meeder, B., & Kleinberg, J. (2011). Differences in the mechanics of information diffusion across topics: idioms, political hashtags, and complex contagion on twitter. In *Proc. of WWW*, pp. 695–704. ACM.

Ron, D. (2010). Algorithmic and analysis techniques in property testing. *Foundations and Trends in TCS*, *5*(2), 73–205.

Schwab, D. J., Bruinsma, R. F., Feldman, J. L., & Levine, A. J. (2010). Rhythmogenic neuronal networks, emergent leaders, and k-cores. *Phys. Rev. E*, *82*, 051911.

Spielman, D. (2009). Smoothed analysis: An attempt to explain the behavior of algorithms in practice. *Communications of the ACM*, 76–84.

Wu, C. W. (2005). Synchronization in networks of nonlinear dynamical systems coupled via a directed graph. *Nonlinearity*, *18*, 1057–1064.