Information Diffusion in Overlaying Social-Physical Networks

Osman Yağan, Dajun Qian, Junshan Zhang, and Douglas Cochran
{oyagan, dqian, junshan.zhang, cochran}@asu.edu
School of Electrical, Computer and Energy Engineering
Arizona State University, Tempe, AZ 85287-5706 USA

***Abstract*—We study the diffusion of information in an overlaying social-physical network. Namely, we consider a physical information network where information spreads amongst people through *conventional* communication media (e.g., face-to-face communication, phone calls), *and* conjoint to this physical network, there are online social networks where information spreads via web sites such as Facebook, Twitter, FriendFeed, YouTube, etc. Capitalizing on the theory of *inhomogeneous random graphs*, we quantify the size and the critical threshold of information epidemics in this conjoint social-physical network by assuming that information diffuses according to the SIR epidemic model. One interesting finding is that even if there is no percolation in the individual networks, percolation (i.e., information epidemics) can take place in the conjoint social-physical network. We also show, both analytically and experimentally, that the fraction of individuals who receive an item of information (started from an arbitrary node) is significantly larger in the conjoint social-physical network case, as compared to the case where the networks are disjoint. These findings reveal that conjoining the physical network with online social networks can have a dramatic impact on the speed and scale of information diffusion.**

Key words: Information Diffusion, Coupled Networks, Random Graphs.

I. Introduction

A. Motivation and Background

Modern society relies on basic physical network infrastructures, such as power stations, telecommunication networks and transportation systems. Recently, due to advances in communication technologies and cyber-physical systems, these infrastructures have become increasingly dependent on one another and have emerged as *interdependent networks* [9]. One archetypal example of such coupled systems is the smart grid where the power stations and the communication network controlling them are coupled together. See the pioneering work of Buldyrev et al. [5] (see also [6], [7], [20]) for a diverse set of models on coupled networks.

Apart from physical infrastructure networks, coupling can also be observed between different types of social networks. Traditionally, people are tied together in a *physical* information network through old-fashioned communication media, such as face-to-face interactions. On the other hand, recent advances of Internet and mobile communication technologies have enabled people to be connected more closely through *online* social networks. Indeed, people can now interact through e-mail or online chatting, or communicate through a Web 2.0 website such as Facebook, Twitter, FriendFeed, YouTube, etc. Clearly, the physical information network and online social networks are *not* completely separate since people may participate in two or more of these networks at the same time. For instance, a person can forward a message to his/her online friends at Facebook and Twitter upon receiving it from someone via face-to-face communication. As a result, the information spread in one network may trigger the propagation in another network, and may result in a possible cascade of information. One conjecture is that due to this *coupling* between the physical and online social networks, today's hot spot news (and information in general) can spread at an unprecedented speed throughout the population, and this is the main subject of the current study.

Information cascades over coupled networks can deeply influence the patterns of social behaviors. Indeed, people have become increasingly aware of the fundamental role of the coupled social-physical network as a medium for the spread of not only information but also ideas and influence. Twitter has emerged as an ultra-fast source of news [19] and Facebook has attracted major businesses and politicians for advertising products or candidates. Several music groups or singers have gained international fame by uploading videos to YouTube. In almost all cases, a new video uploaded to YouTube, a rumor started in Facebook or Twitter, or a political movement advertised through online social networks, either dies out quickly or reaches a significant proportion of the population. In order to fully understand the extent to which these events happen, it is of great interest to consider the combined behavior of the physical information network and several online social networks.

Despite the increasingly important role that coupled networks play in many aspects of modern society, there has been little study on information diffusions across such networks. Most existing works consider information (or disease) propagation only on a single network. For instance, the spread of diseases was studied using the classical susceptible-infectious-recovered (SIR) and susceptible-infectious-susceptible (SIS) models [1], [2], [10], for small-world networks [12], [14], scale-free networks [17], and networks with arbitrary degree distributions [15]. Among them, Newman [15] has shown that SIR epidemics can be quantified for a wide variety of networks by using percolation theory. In particular, the threshold and the size of epidemics are characterized by studying the *phase transition* properties of the underlying random graph models.

B. Summary of Main Contributions

In this paper, we aim to develop a new theoretic framework towards understanding the characteristics of information diffusion across *multiple* networks. For illustration purposes, the model definitions are given in the context of an overlaying social-physical network. Specifically, there is a physical information network where the information spreads amongst people through *conventional* communication media (e.g., face-to-face communication, phone calls), and *conjoint to this physical network*, there are online social networks offering alternative platforms for information diffusion, such as Facebook, Twitter, FriendFeed, YouTube, etc.

In the interest of easy exposition, we focus on the case where there exists only one online social network along with the physical information network; see [21] for an extension to the multiple social networks case. We model the physical network and the social network as random graphs with different topology properties. We assume that each individual in the population is a member of the physical network, and becomes a member of the social network independently with a certain probability. It is also assumed that information is transmitted between two nodes (that are connected by a link in any one of the graphs) according to the SIR model; see Section II for precise definitions.

The problem under consideration is intricate since the relevant random graph model corresponds to a *union* of coupled random graphs, and the techniques employed in [15], [16] for single networks fall short of characterizing its phase transition properties. Capitalizing on the recent progress in *inhomogeneous random graphs* [4], [18], we show that the overlaying social-physical network exhibits a "critical point" above which *information epidemics* are possible; i.e., a single node can spread an item of information (a rumor, an advertisement, a video, etc.) to a positive fraction of individuals in the asymptotic limit. Below the critical point, only *small* information outbreaks can occur and the fraction of influenced individuals always tends to zero.

We consider two different models for the individual networks. First, we assume that both the physical information network and the online social network are Erdős-Rényi (ER) graphs [3], and then we consider the case where both networks are random graphs with arbitrary degree distributions [16]; i.e., both networks are generated according to the *configuration* model [3], [13]. In each case, we quantify the aforementioned critical point by computing the phase transition *threshold* of the conjoint random graph model, and show that it depends on both the degree distributions of the networks and the number of individuals that are members of the online social network. Further, we compute the probability for an information originating from an arbitrary individual to yield an epidemic along with the resulting fraction of individuals that are influenced; this is done for both cases by computing the *giant component* size of the corresponding models.

The results show that the conjoint social-physical network can spread an item of information to a significantly larger fraction of the population as compared to the case where the two networks are disjoint. For instance, consider a physical information network $\mathbb{W}$ and an online social network $\mathbb{F}$ that are ER graphs with respective mean degrees λ_w and λ_f, and assume that each node in $\mathbb{W}$ is a member of $\mathbb{F}$ independently with probability α. If $\lambda_w = 0.6$ and $\alpha = 0.2$, we show that information epidemics are possible in the overlaying social-physical network $\mathbb{H} = \mathbb{W} \cup \mathbb{F}$ whenever $\lambda_f \geq 0.77$. In stark contrast, this happens only if $\lambda_w > 1$ or $\lambda_f > 1$ when the two networks are disjoint. Furthermore, in a single ER network $\mathbb{W}$ with $\lambda_w = 1.5$, an information item originating from an arbitrary individual gives rise to an epidemic with probability 0.58 (i.e., can reach at most 58% of the individuals). However, if the same network $\mathbb{W}$ is conjoined with an ER network $\mathbb{F}$ with $\alpha = 0.5$ and $\lambda_f = 1.5$, the probability of an epidemic becomes 0.82 (indicating that up to 82% of the population can be influenced).

To the best of our knowledge, this paper is the first work that characterizes the information diffusion across multiple networks. The techniques (and the model) presented in this paper can also pave the way in studying the influence maximization [11] problem over multiple networks. We believe that our findings along this line shed light on the understanding on information (and influence) propagation across coupled social-physical networks.

A word on the notation and conventions in use: All limiting statements, are understood with n going to infinity. We use the notation $\xrightarrow{a.s.}$ to indicate almost sure convergence and $\xrightarrow{p}$ to indicate convergence in probability. The mean value of a random variable k is denoted by $< k >$. For a random graph $\mathcal{G}$ we write $C_i(\mathcal{G})$ for the number of nodes in its ith largest connected component. We say that an event holds *with high probability* (whp) if it holds with probability 1 as $n \to \infty$. In comparing the asymptotic behaviors of the sequences $\{a_n\}, \{b_n\}$, we use $a_n = o(b_n)$, $a_n = O(b_n)$, $a_n = \Omega(b_n)$, and $a_n = \Theta(b_n)$, with their meaning in the standard Landau notation.

II. System Model

We consider the following model for an overlaying social-physical network. Let $\mathbb{W}$ stand for the *physical* information network of human beings on the node set $\mathcal{N} = \{1, \ldots, n\}$. We assume that the graph $\mathbb{W}$ characterizes the possible spread of information amongst people through *old-fashioned* communication media; e.g., face-to-face communication, phone calls, etc. Next, let $\mathbb{F}$ stand for the network that characterizes the information spread through an online social networking web site, e.g., Facebook. We assume that each node in $\mathcal{N}$ is a *member* of this auxiliary network with probability $\alpha \in (0, 1]$ independently from any other node. In other words, let

$$\mathbb{P}\left[i \in \mathcal{N}_F\right] = \alpha, \qquad i = 1, \ldots, n, \tag{1}$$

with $\mathcal{N}_F$ denoting the set of human beings that are members of Facebook. With this assumption, it is clear that the vertex set $\mathcal{N}_F$ of $\mathbb{F}$ satisfies $\frac{|\mathcal{N}_F|}{n} \xrightarrow{a.s.} \alpha$ by the law of large numbers.

We assume that information diffuses amongst human beings in an overlaying graph $\mathbb{H}$ that is constructed by taking the union of $\mathbb{W}$ and $\mathbb{F}$. In other words, for any distinct pair of nodes i, j, we say that i and j are adjacent in the network $\mathbb{H}$, denoted $i \sim_{\mathbb{H}} j$, as long as at least one of the conditions $\{i \sim_{\mathbb{W}} j\}$ or $\{i \sim_{\mathbb{F}} j\}$ holds. This is intuitive since a node i can forward information to another node j either by using old-fashioned communication channels (i.e., links in $\mathbb{W}$) or by using Facebook (i.e., links in $\mathbb{F}$). Of course, for the latter to be possible, both i and j should be Facebook users.

We assume that information spreads among the population according to the SIR model. Namely, an individual is either *susceptible* meaning that he/she has not yet heard a particular information, or *infectious* meaning that he/she is aware of the information and is capable of spreading it to his/her contacts, or *recovered* meaning that he/she is no longer spreading the information. As in [15], we assume that an infectious individual i transmits the information to a susceptible contact j with probability $T_{ij} = 1 - e^{-r_{ij}\tau_i}$. Here, r_{ij} denotes the average rate of being in contact over the link from i to j, and τ_i is the time i keeps spreading the information; i.e., the time it takes for i to become recovered.

It is expected that the information propagates over the physical and social network at different speeds, which boil down to different probabilities T_{ij} across links in this case. Specifically, let T_{ij}^{w} stand for the probability of information transmission over a link (between and i and j) in $\mathbb{W}$ and let T_{ij}^{f} denote the probability of information transmission over a link in $\mathbb{F}$. For simplicity, we assume that T_{ij}^{w} and T_{ij}^{f} are independent for all distinct pairs $i, j = 1, \ldots, n$. Furthermore, we assume that the random variables r_{ij}^{w} and τ_i^{w} are independent and identically distributed (i.i.d.) with probability densities $P_w(r)$ and $P_w(\tau)$, respectively. In that case, it was shown in [15] that the information propagates over $\mathbb{W}$ as if all transmission probabilities were equal to T_w, where T_w is the mean value of T_{ij}^{w}, i.e.,

$$T_w := < T_{ij}^{w} > = 1 - \int_0^\infty \int_0^\infty e^{-r\tau} P_w(r) P_w(\tau) dr d\tau.$$

We refer to T_w as the *transmissibility* of the information over the physical network $\mathbb{W}$. Analogously, we let T_f be the information transmissibility over the online social network $\mathbb{F}$.

Under these assumptions, information diffusion becomes equivalent to the (bond) percolation problem of the conjoint network $\mathbb{H} = \mathbb{W} \cup \mathbb{F}$. More specifically, we assume (as in [15]) that each edge in $\mathbb{W}$ is *occupied*, meaning that it can be used in spreading the information, with probability T_w independently from all other edges. Similarly, each edge in $\mathbb{F}$ is deemed occupied (independently) with probability T_f. Then, the size of the information epidemic in $\mathbb{H}$ is equal to the number of individuals that can be reached from an arbitrary node by using only the *occupied* links of $\mathbb{H}$. Hence, the threshold and the size of the information epidemic can be computed by studying the phase transition problem in $\mathbb{H}(T_w, T_f)$ where $\mathbb{H}(T_w, T_f)$ is the random graph containing only the occupied edges of $\mathbb{H}$.

III. Main Results

In what follows, we present the main results of the paper. Due to space limitations, we do not give any proofs here; all the details can be found in [21].

A. Information Diffusion in coupled ER graphs

We first consider a basic scenario where both the physical information network $\mathbb{W}$ and the online social network $\mathbb{F}$ are Erdős-Rényi graphs [3]. More specifically, let $\mathbb{W} = \mathbb{W}(n; \lambda_w/n)$ be an ER network on the vertices $\{1, \ldots, n\}$ such that there exists an edge between any distinct nodes $i, j = 1, \ldots, n$ with probability λ_w/n; this ensures that mean degree of each node is asymptotically equal to λ_w. Next, obtain a set of vertices $\mathcal{N}_F$ by picking each node $1, \ldots, n$ independently with probability $\alpha \in (0, 1]$. Now, let $\mathbb{F} = \mathbb{F}(n; \alpha, \lambda_f/(\alpha n))$ be an ER graph on the vertex set $\mathcal{N}_F$ with edge probability given by $\frac{\lambda_f}{\alpha n}$. The mean degree of a node in $\mathbb{F}$ is given (asymptotically) by λ_f. Assume further that each edge in $\mathbb{W}$ (resp. in $\mathbb{F}$) is occupied with probability T_w (resp. with probability T_f), independently from all other edges.

The overall system model $\mathbb{H}$ can now be obtained by *conjoining* the physical information network $\mathbb{W}$ and the online social network $\mathbb{F}$. In other words, $\mathbb{H}$ is constructed on the vertices $1, \ldots, n$ by conjoining the *occupied* edges of $\mathbb{W}$ and $\mathbb{F}$, i.e., we have $\mathbb{H}(n; \alpha, T_w\lambda_w, T_f\lambda_f) = \mathbb{W}(n; T_w\lambda_w/n) \cup \mathbb{F}(n; \alpha, T_f\lambda_f/(\alpha n))$. Next, we present the first main result that characterizes the critical threshold and the size of the information epidemic in the overlaying social-physical network.

Let $\lambda_{fw}^{\star}$ be defined by

$$\lambda_{fw}^{\star} := \frac{1}{2}\left(T_f\lambda_f + T_w\lambda_w\right) + \frac{1}{2}\sqrt{\left(T_f\lambda_f + T_w\lambda_w\right)^2 - 4(1-\alpha)T_f\lambda_f T_w\lambda_w}. \tag{2}$$

Also, let ρ_1 be the largest solution of the equation

$$(1-\alpha)T_w\lambda_w\left((1-\rho_1)e^{\rho_1 T_f\lambda_f} - 1\right) - \log(1-\rho_1) = \rho_1(T_f\lambda_f + \alpha T_w\lambda_w) \tag{3}$$

with ρ_1 in $[0, 1]$, and let ρ_2 be given by

$$\rho_2 = \frac{-\log(1-\rho_1) - \rho_1(\alpha T_w\lambda_w + T_f\lambda_f)}{(1-\alpha)T_w\lambda_w}. \tag{4}$$

Theorem 3.1: With the above assumptions, we have

(i) If $\lambda_{fw}^{\star} \leq 1$, then with high probability, the size of the largest component satisfies $C_1(\mathbb{H}(n; \alpha, T_w\lambda_w, T_f\lambda_f)) = O(\log n)$; in contrast, if $\lambda_{fw}^{\star} > 1$ we have $C_1(\mathbb{H}(n; \alpha, T_w\lambda_w, T_f\lambda_f)) = \Theta(n)$ whp, while the size of the second largest component satisfies $C_2(\mathbb{H}(n; \alpha, T_w\lambda_w, T_f\lambda_f)) = O(\log n)$.

(ii) Also, $\frac{1}{n}C_1(\mathbb{H}(n; \alpha, T_w\lambda_w, T_f\lambda_f)) \xrightarrow{p} \alpha\rho_1 + (1-\alpha)\rho_2$.

Theorem 3.1 quantifies the number of individuals in the overlaying social-physical network that are likely to receive an item of information which starts spreading from a single individual. Specifically, the "critical point" of the information epidemic is marked by $\lambda_{fw}^{\star} = 1$, with the critical

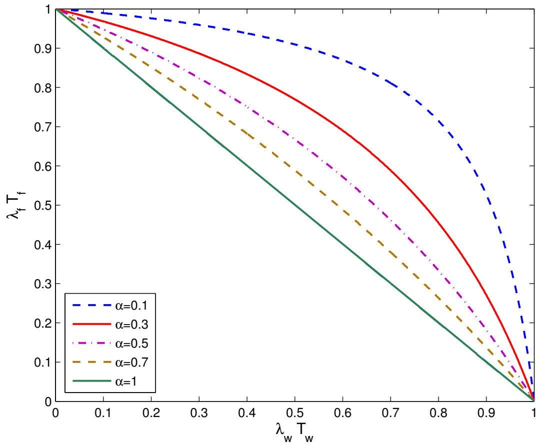

Fig. 1. *The minimum $\lambda_f T_f$ required for the existence of a giant component in $\mathbb{H}(n; \alpha, T_w\lambda_w, T_f\lambda_f)$ versus $\lambda_w T_w$ for various α values.*

threshold $\lambda^\star_{fw}$ given by (2). We conclude from Theorem 3.1 that for any parameter set that yields $\lambda^\star_{fw} \leq 1$ (the subcritical regime), the largest possible number of individuals who receive the information is $O(\log n)$, meaning that only small (non-epidemic) information outbreaks can take place. On the other hand, if $\lambda^\star_{fw} > 1$ (the supercritical regime), the information has a positive probability of reaching to a linear fraction of the individuals; i.e, information epidemics can occur. In that case, an information (started spreading from an arbitrary individual) gives rise to an information epidemic with probability $\alpha\rho_1 + (1-\alpha)\rho_2$ and reaches to a fraction $\alpha\rho_1 + (1-\alpha)\rho_2$ of individuals in the network; here ρ_1 is obtained by the largest solution of (3) and ρ_2 is given by (4).

We observe that the threshold function $\lambda^\star_{fw}$ is symmetric in $T_f\lambda_f$ and $T_w\lambda_w$ meaning that both networks have identical effect on carrying the network to the supercritical regime where information can reach to a linear fraction of the nodes. To get a more concrete sense, we depict in Figure 1 the minimum $\lambda_f T_f$ required to have a giant component in $\mathbb{H}(n; \alpha, T_w\lambda_w, T_f\lambda_f)$ versus $\lambda_w T_w$ for various α values. Each curve in the figure corresponds to a phase transition boundary above which information epidemics are possible. If $T_f = T_w = 1$, the same plot shows the boundary of the giant component existence with respect to the mean degrees λ_f and λ_w. This clearly shows how two networks that are in the subcritical regime can yield an information epidemic when they are conjoined together. For instance, we see that for $\alpha = 0.1$, it suffices to have $\lambda_f = \lambda_w = 0.76$ for the existence of an information epidemic. Yet, if the two networks were disjoint, it would be needed [3] to have $\lambda_f > 1$ and $\lambda_w > 1$.

It is of interest to state whether or not Theorem 3.1 can be deduced from the phase transition results for random graphs with arbitrary degree distributions (e.g., see [13], [15], [16]). It is well-known [13] that for these graphs the critical point of the phase transition is given by $\frac{\mathbb{E}[d_i(d_i-1)]}{\mathbb{E}[d_i]} = 1$, where d_i is the degree of an arbitrary node. We next show that this condition is not equivalent (indeed, is not even a good approximation) to $\lambda^\star_{fw} = 1$. First, note that the degree of an arbitrary node i in $\mathbb{H}$ follows a Poisson distribution with mean $T_w\lambda_w$ if $i \notin \mathcal{N}_F$ (happens with probability $1-\alpha$), and it follows a Poisson distribution with mean $T_f\lambda_f + T_w\lambda_w - \frac{T_f\lambda_f T_w\lambda_w}{n}$ if $i \in \mathcal{N}_F$ (happens with probability α). For n large, this leads to

$$\frac{\mathbb{E}\left[d_i(d_i-1)\right]}{\mathbb{E}\left[d_i\right]} = \frac{\alpha(T_f\lambda_f + T_w\lambda_w)^2 + (1-\alpha)(T_w\lambda_w)^2}{\alpha T_f\lambda_f + T_w\lambda_w}. \quad (5)$$

It can be seen that the above expression is not equal to $\lambda^\star_{fw}$ given by (2). For instance, with $\alpha = 0.2$, $T_w\lambda_w = 0.6$ and $T_f\lambda_f = 0.8$, we have $\lambda^\star_{fw} = 1.03$ while (5) yields 0.89 signalling a significant difference between the exact threshold $\lambda^\star_{fw}$ and the mean field approximation given by (5). We conclude that the results established above go beyond the classical results for random graphs with arbitrary degree distributions.

B. Information Diffusion in Coupled Graphs with Arbitrary Degree Distributions

We now expand the previous results to a more general and in fact more practically relevant class of graphs usually known as random graphs with arbitrary degree distribution [15], [16]. In particular, we specify a degree distribution that gives the properly normalized probabilities $\{p_k^w, k = 0, 1, \ldots\}$ that an arbitrary node in $\mathbb{W}$ has degree k. Namely, we let each node $i = 1, \ldots, n$ in $\mathbb{W} = \mathbb{W}(n; \{p_k^w\})$ has a random degree drawn from the distribution $\{p_k^w\}$ independently from any other node. Similarly, we assume that the degrees of all nodes in $\mathbb{F}$ are drawn independently from the distribution $\{p_k^f\}$; see [15], [16], [18] for details about the construction of random graphs with given degree distributions. Finally, the vertex set of $\mathbb{F} = \mathbb{F}(n; \alpha, \{p_k^f\})$ is obtained in the usual manner by picking each node $1, \ldots, n$ independently with probability α. In what follows we shall assume that the degree distributions are well-behaved in the sense that all moments of arbitrary order are finite.

As in the previous section, let T_w be the information transmissibility, i.e., the mean probability of information transfer between any two nodes, in the physical network $\mathbb{W}$, and let T_f be the information transmissibility in the online social network $\mathbb{F}$. In other words, each edge in $\mathbb{W}$ is deemed *occupied* meaning that it can be used in spreading the information with probability T_w. Similarly, we let each edge in $\mathbb{F}$ be occupied with probability T_f independently from all the other edges. The overall system model can now be obtained by taking a union of the occupied edges of $\mathbb{W}$ and $\mathbb{F}$. Namely, we let $\mathbb{H}(n; \alpha, \{p_k^w\}, T_w, \{p_k^f\}, T_f) = \mathbb{W}(n; \{p_k^w\}, T_w) \cup \mathbb{F}(n; \alpha, \{p_k^f\}, T_f)$ be the corresponding social-physical network over which the information diffuses.

We now present the second main result that characterizes the threshold and the size of the information epidemic in $\mathbb{H}$ by revealing its phase transition properties. First, for notational convenience, let k_f and k_w be random variables independently drawn from the distributions $\{p_k^f\}$ and $\{p_k^w\}$, respectively, and let $< k_f > := \lambda_f$ and $< k_w > := \lambda_w$. Furthermore, assume that β_f and β_w are given by

$$\beta_f := \left(< k_f^2 > - \lambda_f\right)/\lambda_f \quad \text{and} \quad \beta_w := \left(< k_w^2 > - \lambda_w\right)/\lambda_w,$$

and let the threshold function $\sigma^\star_{fw}$ be defined through

$$\sigma^\star_{fw} = \frac{T_f\beta_f + T_w\beta_w + \sqrt{(T_f\beta_f - T_w\beta_w)^2 + 4\alpha T_f T_w \lambda_f \lambda_w}}{2} \quad (6)$$

Finally, let h_1, h_2 in $(0,1]$ be given by the pointwise smallest solution of the recursive equations

$$h_1 = \frac{1}{\lambda_f}\mathbb{E}\left[k_f\left(1+T_f(h_1-1)\right)^{k_f-1}\right] \times \mathbb{E}\left[\left(1+T_w(h_2-1)\right)^{k_w}\right] \quad (7)$$

$$h_2 = \frac{1}{\lambda_w}\mathbb{E}\left[\alpha(1+T_f(h_1-1))^{k_f}+1-\alpha\right] \times \mathbb{E}\left[k_w\left(1+T_w(h_2-1)\right)^{k_w-1}\right]. \quad (8)$$

Theorem 3.2: Under the assumptions just stated, we have

(i) *If* $\sigma^\star_{fw} \le 1$ *then with high probability the size of the largest component satisfies* $C_1\left(\mathbb{H}(n;\alpha,\{p_k^w\},T_w,\{p_k^f\},T_f)\right) = o(n)$. *On the other hand, if* $\sigma^\star_{fw} > 1$, *then* $C_1\left(\mathbb{H}(n;\alpha,\{p_k^w\},T_w,\{p_k^f\},T_f)\right) = \Theta(n)$ *whp.*

(ii) *Also,*

$$\frac{1}{n}C_1\left(\mathbb{H}(n;\alpha,\{p_k^w\},T_w,\{p_k^f\},T_f)\right) \xrightarrow{p} 1-\mathbb{E}\left[\alpha(1+T_f(h_1-1))^{k_f}+1-\alpha\right] \times \mathbb{E}\left[\left(1+T_w(h_2-1)\right)^{k_w}\right]. \quad (9)$$

Theorem 3.2 can be viewed as a counterpart of Theorem 3.1. It quantifies the number of individuals in the overlaying social-physical network likely to receive a particular information when the physical network $\mathbb{W}$ and the online social network $\mathbb{F}$ have arbitrary degree distributions $\{p_k^w\}$ and $\{p_k^f\}$, respectively. Specifically, for $\{p_k^w\}$ and $\{p_k^f\}$ with finite moments, Theorem 3.2 shows that the critical point of the information epidemic is marked by $\sigma^\star_{fw} = 1$, with the critical threshold $\sigma^\star_{fw}$ given by (6). In other words, for any parameter set that yields $\sigma^\star_{fw} > 1$ (supercritical regime), an information has a positive probability of giving rise to an information epidemic, i.e., reaching to a linear fraction of the individuals. In that case, the asymptotic fraction of the individuals who receive the information can be found by first solving the recursive equations (7)-(8) for the smallest h_1, h_2 in $(0,1]$ and then computing the expression given in (9). On the other hand, if it holds that $\sigma^\star_{fw} \le 1$ (subcritical regime), we conclude from Theorem 3.2 that the largest number of individuals who receive the information will be $o(n)$ with high probability, meaning that all outbreaks are non-epidemic.

We have some further remarks on the applicability of Theorem 3.2. Consider the case where both $\mathbb{W}$ and $\mathbb{F}$ are ER graphs, i.e., let $p_k^w = e^{-\lambda_w}\frac{\lambda_w^k}{k!}$ and $p_k^f = e^{-\lambda_f}\frac{\lambda_f^k}{k!}$. We have that $\beta_f = \lambda_f$, $\beta_w = \lambda_w$, and it is easy to check that $\sigma^\star_{fw} = \lambda^\star_{fw}$ so that part (i) of Theorem 3.2 is compatible with part (i) of Theorem 3.1. Also, we find (numerically) that the second parts of Theorem 3.1 and Theorem 3.2 yield the same asymptotic giant component size. Nevertheless, it is worth noting that, although ER graphs constitute a special case of the random graphs with arbitrary degree distributions, Theorem 3.1 is not a mere consequence of Theorem 3.2. This is because, through a different technique used in the proofs, Theorem 3.1 provides sharper bounds $C_1(\mathbb{H}(n;\alpha,T_w\lambda_w,T_f\lambda_f) = O(\log n)$ (subcritical case) and $C_2(\mathbb{H}(n;\alpha,T_w\lambda_w,T_f\lambda_f) = O(\log n)$ (supercritical case) that go beyond Theorem 3.2.

IV. Numerical Results

We next present numerical results for the case where the networks $\mathbb{W}$ and $\mathbb{F}$ are random graphs with arbitrary degree distributions. In order to gain more insight on the consequences of Theorem 3.2, we consider a specific example of information diffusion over the physical information network $\mathbb{W}$ and the online social network $\mathbb{F}$ with degree distributions $\{p_k^w\}$ and $\{p_k^f\}$, respectively. Here, we use power-law distributions with exponential cutoff. Namely, we let $p_0^w = p_0^f = 0$, and

$$p_k^w = \left(\mathrm{Li}_{\gamma_w}(e^{-1/\Gamma_w})\right)^{-1} k^{-\gamma_w} e^{-k/\Gamma_w}, \quad k = 1,2,\ldots \quad (10)$$

$$p_k^f = \left(\mathrm{Li}_{\gamma_f}(e^{-1/\Gamma_f})\right)^{-1} k^{-\gamma_f} e^{-k/\Gamma_f}, \quad k = 1,2,\ldots \quad (11)$$

where γ_w, γ_f, Γ_w and Γ_f are positive constants and the normalizing constant $\mathrm{Li}_m(z)$ is the mth polylogarithm of z, i.e., $\mathrm{Li}_m(z) = \sum_{k=1}^\infty \frac{z^k}{k^m}$. Power law distributions with exponential cutoff are chosen here because they are applied to a variety of real-world networks (e.g., see [8] for a detailed empirical study on the degree distributions of real-world networks). Moreover, the distributions (10)-(11) ensure that all moments of arbitrary order are finite as required by Theorem 3.2.

To that end, we depict in Figure 2 the minimum T_f value required to have a giant component in $\mathbb{H}(n;\alpha,\{p_k^w\},T_w,\{p_k^f\},T_f)$ versus T_w, for various α values. In other words, each curve corresponds to a phase transition boundary above which information epidemics are possible, in the sense that an information has a positive probability of reaching out to a linear fraction of individuals in the overlaying social-physical network. In all plots, we set $\gamma_f = \gamma_w = 2.5$ and $\Gamma_f = \Gamma_w = 10$. The T_f and T_w values are multiplied by the corresponding β_f and β_w values to make a fair comparison with the disjoint network case where it is required [15] to have $\beta_w T_w > 1$ (or $\beta_f T_f > 1$) for the existence of an epidemic; under the current setting we have $\beta_f = \beta_w = 1.545$. Figure 2 illustrates (in the arbitrary distribution case) how conjoining two networks can speed up the information diffusion. Namely, it can be seen that even for small α values, two networks, albeit having no giant component individually, can yield an information epidemic when they are conjoined together. To give an example, we see that for $\alpha = 0.1$, it suffices to have that $\beta_f T_f = \beta_w T_w = 0.774$ for the existence of an information epidemic in the conjoint network $\mathbb{H}$, whereas if the networks $\mathbb{W}$ and $\mathbb{F}$ are disjoint, an information epidemic can occur only if $\beta_w T_w > 1$ or $\beta_f T_f > 1$.

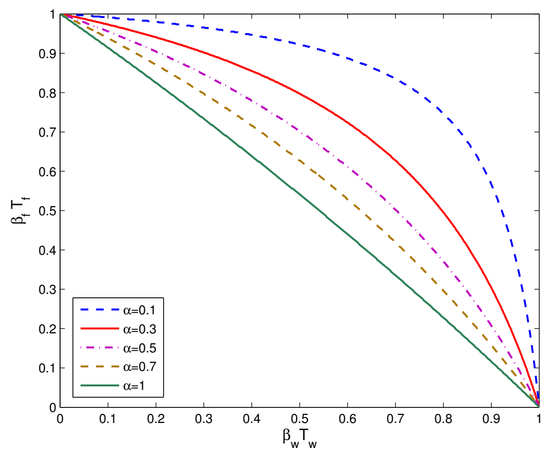

Fig. 2. *The minimum T_f required for the existence of a giant component in $\mathbb{H}(n; \alpha, \{p_k^w\}, T_w, \{p_k^f\}, T_f)$ versus T_w. The distributions $\{p_k^w\}$ and $\{p_k^f\}$ are given by (10) and (11), with $\gamma_f = \gamma_w = 2.5$ and $\Gamma_f = \Gamma_w = 10$.*

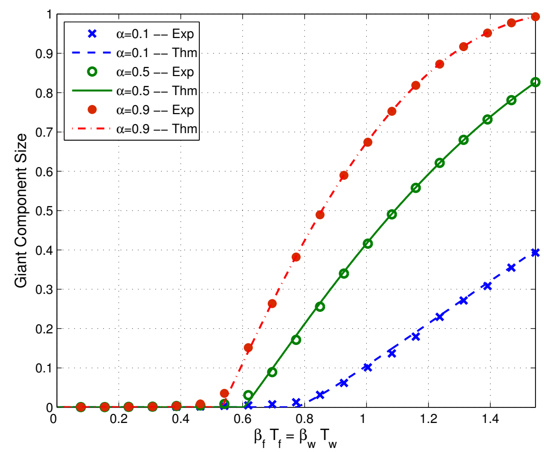

Fig. 3. *The fractional size of the giant component in $\mathbb{H}(n; \alpha, \{p_k^w\}, T_w, \{p_k^f\}, T_f)$ versus $T_f\beta_f = T_w\beta_w$. The parameters are as in Figure 2. The curves correspond to analytical results obtained via Theorem 3.2, whereas marked points stand for the experimental results obtained for $n = 20,000$ by averaging 200 experiments in each parameter set.*

Next, we turn to the computation of the giant component size. In Figure 3, we depict the fractional size of the giant component in $\mathbb{H}(n; \alpha, \{p_k^w\}, T_w, \{p_k^f\}, T_f)$ versus $T_f\beta_f = T_w\beta_w$, for various α values; as before, we set $\gamma_f = \gamma_w = 2.5$ and $\Gamma_f = \Gamma_w = 10$ yielding $\beta_f = \beta_w = 1.545$. In other words, the plots stand for the largest fraction of individuals in the social-physical network who can receive an information item that has started spreading from a single individual. The curves stand for the analytical results obtained via Theorem 3.2 whereas marked points stand for the experimental results obtained with $n = 20,000$ nodes by averaging 200 experiments for each parameter set. We see that there is a good agreement between the theory and experiment even for such a small number of vertices; the discrepancy near the phase transition is clearly due to the finite size effect.

V. Conclusion

In this paper, we characterize the critical threshold and the asymptotic size of information epidemics in an overlaying social-physical network. To capture the spread of information, we consider a physical information network that characterizes the face-to-face interactions of human beings, and some overlaying online social networks (e.g., Facebook, Twitter, etc.) that are defined on a subset of the population. Assuming that information is transmitted between individuals according to the SIR model, we show that the critical point and the size of information epidemics on this overlaying social-physical network can be precisely determined by employing the approaches on inhomogeneous random graphs. We believe that our findings here shed light on the further studies on information (and influence) propagation across social-physical networks.

Acknowledgments

This research was supported in part by the U.S. National Science Foundation grants No. CNS-0905603, CNS-0917087, and the DTRA grant HDTRA1-09-1-0032.

References

[1] R. M. Anderson and R. M. May, *Infectious Diseases of Humans*, Oxford University Press, Oxford (UK), 1991.

[2] N. T. J. Bailey, *The Mathematical Theory of Infectious Diseases and its Applications*, Hafner Press, New York, 1975.

[3] B. Bollobás, *Random Graphs*, Cambridge Studies in Advanced Mathematics, Cambridge University Press, Cambridge (UK), 2001.

[4] B. Bollobás, S. Janson, and O. Riordan, "The phase transition in inhomogeneous random graphs," *Random Structures and Algorithms* **33**:1, 2007.

[5] S. V. Buldyrev, R. Parshani, G. Paul, H. E. Stanley, and S. Havlin, "Catastrophic cascade of failures in interdependent networks," *Nature* **464***:1025–1028, 2010.*

[6] S. V. Buldyrev, N. W. Shere, and G. A. Cwilich, "Interdependent networks with identical degrees of mutually dependent nodes," Phys. Rev. E **83***:016112, 2011.*

[7] W. Cho, K. I. Goh and I. M. Kim, "Correlated couplings and robustness of coupled networks," Available online at arXiv:1010.4971v1.

[8] A. Clauset, C. R. Shalizi, and M. E. J. Newman, "Power-law distributions in empirical data," Available online at arXiv:0706.1062v2, 2009.

[9] CPS Steering Group, "Cyber-physical systems executive summary, 2008," Available online at http://varma.ece.cmu.edu/summit/CPS-Executive-Summary.pdf.

[10] H. W. Hethcote, "Mathematics of infectious diseases," SIAM Review **42:** *599–673, 2000.*

[11] D. Kempe, J. Kleinberg, E. Tardos, "Maximizing the spread of influence through a social network," Proceedings of the KDD-2003.

[12] M. Kuperman and G. Abramson, "Small world effect in an epidemiological model," Phys. Rev. Lett. **86***, 2001, pp. 2909–2912.*

[13] M. Molloy and B. Reed, "A critical point for random graphs with a given degree sequence," Random Structures and Algorithms **6***:161–179, 1995.*

[14] C. Moore and M.E.J. Newman, "Epidemics and percolation in small-world networks," Phys. Rev. E **61(5)***, 2000, pp. 5678–5682.*

[15] M. E. J. Newman, "Spread of epidemic disease on networks," Phys. Rev. E **66(1)***:16128, 2002.*

[16] M. E. J. Newman, S. H. Strogatz, and D. J. Watts, "Random graphs with arbitrary degree distributions and their applications," Phys. Rev. E **64(2):***26118, 2001.*

[17] R. Pastor-Satorras and A. Vespignani, "Epidemic spreading in scale-free networks," Phys. Rev. Lett. **86***, 2001, pp. 3200-3203.*

[18] B. Söderberg, "Random graph models with hidden color," Acta Phys. Pol B **34** *(2003), 5085–5102.*

[19] Wall Street Journal article appeared at http://online.wsj.com/article/SB10001424052970204138204576598942105167646.html.

[20] O. Yağan, D. Qian, J. Zhang, and D. Cochran, "Optimal Allocation of Interconnecting Links in Cyber-Physical Systems: Interdependence, Cascading Failures and Robustness." Special Issue of IEEE Transactions on Parallel and Distributed Systems on Cyber-Physical Systems. Accepted for publication (2012). Available online at arXiv:1201.2698v1.

[21] O. Yağan, D. Qian, J. Zhang, and D. Cochran, "Conjoining speeds-up information diffusion in overlaying social-physical networks," Available online at arXiv:1112.4002v1[cs.SI].