Diffusion of Real-Time Information in Social-Physical Networks

Dajun Qian*, Osman Yağan†, Lei Yang*, and Junshan Zhang*
* School of ECEE, Arizona State University, Tempe, AZ, USA
† Cybersecurity Laboratory (CyLab), Carnegie Mellon University, Pittsburgh, PA, USA

***Abstract*—We study the diffusion behavior of real-time information. Typically, real-time information is valuable only for a limited time duration, and hence needs to be delivered before its "deadline." Therefore, real-time information is much easier to spread among a group of people with frequent interactions than between isolated individuals. With this insight, we consider a social network which consists of many cliques and information can spread quickly within a clique. Furthermore, information can also be shared through online social networks, such as Facebook, twitter, Youtube, etc.**

We characterize the diffusion of real-time information by studying the phase transition behaviors. Capitalizing on the theory of inhomogeneous random networks, we show that the social network has a critical threshold above which information epidemics are very likely to happen. We also theoretically quantify the fractional size of individuals that finally receive the message. The numerical results indicate that real-time information could be much easier to propagate in a social network when large size cliques exist.

I. Introduction

A. Motivation and Background

In today's modern society, people are becoming increasingly tied together over social networks. Thanks to *online social networks*, such as Facebook and Twitter, people can share messages quickly with their friends. Meanwhile, a *physical information network* [1] based on traditional face-to-face interactions still remains an important medium for message spreading. Very recent work [2] has shown that different social networks are usually *coupled* together, and their conjoining could greatly facilitate information diffusion; see also [3] for a study on the diffusion of *influence* under a similar setting. As a result, today's hot spot news or fashion behaviors are more likely to generate pronounced influence over the population than ever before.

The main thrust of this study is dedicated to understanding the diffusion behavior of real-time information. Typically, the real-time information is valuable only for a limited time duration [4], and hence needs to be delivered before its "deadline." For example, once a limited-time coupon is released from Groupon or Dealsea.com, people can share this news either by talking to friends or posting it on Facebook. However, people would not have much interest on this deal after it is not longer available.

This research was supported in part by the U. S. National Science Foundation grants No. CNS-0905603 and CNS-0917087.

Clearly, due to the timeliness requirement, the potential influenced scale of real-time information in a social network depends on the speed of message propagation. The faster the message passes from one to another, the more people can learn this news before it expires. With this insight, in order to characterize its diffusion behavior, a key step is to quantify how fast the message can spread along different social connections.

In this study, we assume that information could spread amongst people through both face-to-face contacts and online communications. In an online social network, the message can spread quickly over long distance, and hence it is reasonable to treat online connections as the same type of links regardless of real-world distances.

On the other hand, face-to-face communications are largely constrained by distance between individuals. Recent works in [1], [5] have explored the structure of physical information network by tracking in-person interactions over the population. Their findings indicate that such interactions would give rise to a social graph made of a large number of small non-overlapping *cliques*, which are loosely connected. Each clique stands for a group of people who are close to each other. The message can spread quickly within a clique via frequent interactions, but takes longer time to spread across cliques separated by longer distances. Clearly, constrained by its deadline, the real-time information is less likely to propagate across cliques via face-to-face contacts. Needless to say, in order to characterize the diffusion behavior of real-time information, we need to consider the impact of such clique structure, which has not been studied in previous works on general information diffusions [2], [6], [7].

B. Summary of Main Contributions

We explore the diffusion of real-time information in an overlaying social-physical network where the information could spread amongst people through both face-to-face contacts (physical information network) and online communications (online social network). Based on empirical observations in [1], [5], we assume that the physical information network consists of many non-overlapping cliques where each clique represents a group of people with frequent face-to-face interactions, e.g., family in a house or colleagues in an office. Clearly, the face-to-face contacts are less likely to happen across cliques.

We characterize the information diffusion process by studying the *phase transition behaviors* (see Section II-C for details). Specifically, we show that the social network has a *critical threshold* above which *information epidemics* are very likely to happen, i.e., the information can reach a non-trivial fraction of individuals. We also quantify the fraction of individuals that finally receive the message by computing the size of *giant component* (see Section II-C for details). The numerical results in Section V indicate that real-time information could be much easier to propagate in a social network when large size cliques exist. As illustrated in Figure 5, when the average clique size increases from 1 to 2, the fraction of individuals that receive the message can grow sharply from 14% to 80%.

Note that our work here has significant differences from the previous studies on information propagation. In [6], [7], it is assumed that message could propagate at the same speed along different social relationships. Clearly, such assumption would be inappropriate for the diffusion of real-time information that depends on propagation speeds. Very recent work in [2] considered online connections and face-to-face connections for general information diffusion, but did not study the impact of clique structure on information diffusion. To the best of our knowledge, this paper is the first work on the diffusion of real-time information with consideration on the clique structure in social networks. We believe that our work will offer initial steps towards understanding the diffusion behaviors of real-time information.

II. System model

We consider an overlaying social-physical network $\mathbb{H}$ that consists of a physical information network $\mathbb{W}$ and an online social network $\mathbb{F}$. The nodes in the graph $\mathbb{W}$ represent the human beings in the real world. Based on empirical studies in [1], [5], we assume that the graph $\mathbb{W}$ consists of many non-overlapping cliques where each clique represents a group of closely located people, e.g., family in a house or colleagues in an office. Meanwhile, each node in $\mathbb{W}$ can independently participate the online social network $\mathbb{F}$ with probability α, and the nodes in $\mathbb{F}$ stand for their online memberships. Throughout this paper, we also refer to the nodes in $\mathbb{W}$ and $\mathbb{F}$ as "individuals" and "online users," respectively. Furthermore, the links connecting the nodes in $\mathbb{W}$ stand for traditional face-to-face connections, while the links in $\mathbb{F}$ represent online connections.

A. Topology Structure in System Model

In what follows, we specify the topology structure in the system model $\mathbb{H}$ in Fig. 1.

Cliques in the physical information network. The physical information network has N nodes and the nodes set is denoted by $\mathcal{N} = \{1, 2, ..., N\}$. These nodes are gathered into many cliques with different sizes. We expect the clique size follows the distribution $\{\mu_n^w, n = 1, 2, ..., D\}$, where D is the largest possible size. Therefore, an arbitrary clique could contain n nodes with probability μ_n^w. We generate these

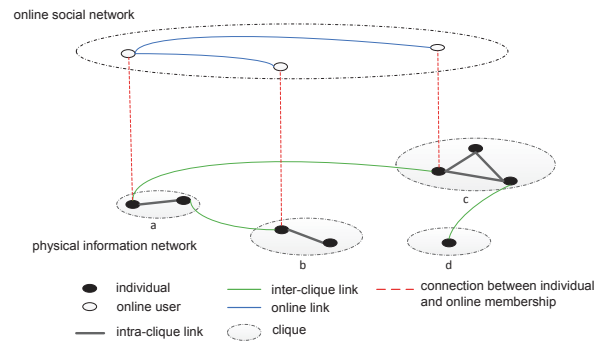

Fig. 1. System model $\mathbb{H}$

cliques as follows: at step $t = 1$, we randomly choose n nodes from the collection $\mathcal{N}$ and create a clique with the selected n nodes, where n is a random number following the distribution $\{\mu_n^w, n = 1, 2, ..., D\}$. We also denote the collection of the remaining nodes in $\mathcal{N}$ by $\mathcal{N}_1$. At each step t, we repeat the above procedure to create a new clique from the collection $\mathcal{N}_{t-1}$[1], and assume that we can finally generate N_c cliques in $\mathbb{W}$ [2]. Generally speaking, the existence of large size cliques indicates that many individuals are close to each other. In other words, the clique size distribution $\{\mu_n^w\}$ offers an abstract characterization of personal distances in $\mathbb{W}$ from a macroscopic perspective.

Type-0 (intra-clique) links in $\mathbb{W}$. Since the nodes within the same cliques could interact to each other frequently, we assume these nodes are fully connected by *type-0 links*. Note that in this study, the concept of clique is different from the well-studied community metric in social networks, in the sense that in the clique all nodes have connections to each other.

Type-1 (inter-clique) links in $\mathbb{W}$. We assume that a face-to-face interaction is still possible between cliques, e.g., a person may talk to a remote friend by walking across a long distance. Suppose each node can randomly connect to k^w nodes from other cliques through *type-1 links* where k^w is a random variable drawn independently from the distribution $\{p_k^w, k = 0, 1, ...\}$.

Online users and type-2 (online) links. The nodes in the online network $\mathbb{F}$ represent the online users. As in [2], we assume each online user randomly connects to k^f online neighbors in $\mathbb{F}$, where k^f is a random variable whose distribution is drawn independently from $\{p_k^f, k = 0, 1, ...\}$. We denote such online connection as *type-2 link*. Furthermore, we draw a virtual *type-3 link* from an online user in $\mathbb{F}$ to the actual person it corresponds to in the physical information network $\mathbb{W}$; this indicates that the two nodes actually correspond to the same individual.

Online users associated with a clique. To avoid confusions, we say *"size-n clique with m online members"* when

[1]Note that the last generated clique may not follow the expected size distribution, since there would be only too few nodes left to choose. However, such impact on clique size distribution will be negligible if the number of cliques is large enough.

[2]Throughout this paper, we use "clique in $\mathbb{W}$" and "clique in $\mathbb{H}$" interchangeably, in the sense that the network $\mathbb{W}$ is also a part of system model $\mathbb{H}$.

referring to the case that a clique contains n individuals and only m of them participate in the online social network $\mathbb{F}$. With this insight, we can also differentiate among the collection of size-n cliques according to their affiliated online users. Specifically, for the collection of size-n cliques with m online members, $m \leq n \leq D$, we assume their fraction size in the whole collection of cliques is μ_{nm}. It is easy to see that

$$\mu_{nm} = \mu_n^w \begin{pmatrix} n \\ m \end{pmatrix} \alpha^m (1-\alpha)^{n-m} \quad \text{and} \quad \mu_n^w = \sum_{m=1}^{n} \mu_{nm}. \tag{1}$$

Furthermore, for the collection of cliques with m online users, their fraction size can be given by

$$\mu_m^f = \sum_{n=m}^{D} \mu_{nm}. \tag{2}$$

B. Information Transmissibility

The message can propagate at different speeds along different types of social connections in $\mathbb{H}$. Due to timeliness requirement, the real-time information is easier to pass over a link that offers fast propagation speed. Therefore, we assign each link with a *transmissibility* as in [2], [7], i.e., the probability that the message can successfully pass through.

From practical scenarios, we set the transmissibility along type-0 link as $T_c = 1$ since the message spreads quickly within a clique. We also define the transmissibilities along type-1 and type-2 links as T_w and T_f, respectively. Throughout this paper, we say a link is *occupied* if the message can successfully pass through that link. Hence, in $\mathbb{H}$ each type-1 link is occupied independently with probability T_w, whereas each type-2 link is occupied independently with probability T_f.

C. Information Cascade

We give a brief description of the information diffusion process in the following. Suppose that the message starts to spread from an arbitrary node i in a clique of $\mathbb{W}$. Then, the other nodes in this clique will quickly receive that message through type-0 links. The message can also propagate to other cliques through occupied type-1 and type-2 links. This in turn may trigger further message propagation and may eventually lead to an information epidemic; i.e., a non-zero fraction of individuals may receive the information in the limit $N \to \infty$.

Clearly, an arbitrary individual can spread the information to nodes that are reachable from itself via the occupied edges of $\mathbb{H}$. Hence, the size of an information outbreak (i.e., the number of individuals that are informed) is closely related to the size of the connected components of $\mathbb{H}$, which contains only the *occupied* type-1 and type-2 links [2], [7], [6] of $\mathbb{H}$. Thus, the information diffusion process considered here is equivalent to a heterogeneous bond-percolation process over the network $\mathbb{H}$; the corresponding bond percolation is heterogeneous since the occupation probabilities are different for type-1 and type-2 links. In this paper, we will exploit this relation and find the condition and the size of information epidemics by studying the phase transition properties of $\mathbb{H}$. A key observation is that the system $\mathbb{H}$ exhibits a *phase transition*

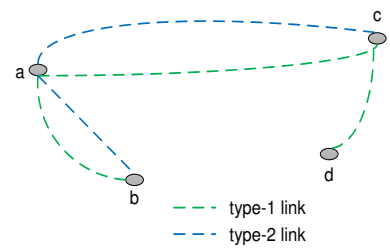

Fig. 2. Equivalent graph $\mathbb{E}$. Nodes {a,b,c,d} in this graph corresponds to the cliques {a,b,c,d} of the system $\mathbb{H}$ in Fig. 1. We assign type-1 and type-2 links in $\mathbb{E}$ according to the same type of links connecting cliques in Fig. 1.

behavior at a *critical threshold*. Specifically, a *giant connected component* G_H that covers a non-trivial fraction of nodes in $\mathbb{H}$ is likely to appear above the critical threshold meaning that *information epidemics* are possible. Below that critical threshold, all components are small indicating that the fraction of influenced individuals tends to zero in the large system size limit.

It is easy to see that the influenced individuals and cliques correspond to the nodes and cliques in $\mathbb{W}$ that are contained inside G_H. Hence, we introduce two parameters to evaluate the scale of information diffusion:

- S_c: The fractional size of influenced cliques in $\mathbb{W}$. Namely, S_c corresponds to the ratio of the number of cliques in G_H to the total number of cliques in $\mathbb{W}$.
- S_n: The fractional size of influenced individuals in $\mathbb{W}$. Namely, S_n corresponds to the ratio of the number of nodes that belong to the cliques in G_H to the total number of nodes in $\mathbb{W}$.

With this insight, we can explore the information diffusion process by characterizing the phase transition behavior of the giant component G_H.

III. Equivalent graph: a clique level approach

In this study, we are particularly interested in the following two questions:

- What is the critical threshold of the system $\mathbb{H}$? In other words, under what condition, the information reaches a non-trival fraction of the network rather than dying out quickly?
- What is the *expected* size of an information epidemic? In other words, to what fraction of nodes and cliques does the information reach? Or, equivalently, what are the sizes S_c and S_n?

These two questions can be answered by quantifying the phase transition behaviors of the random graph $\mathbb{H}$. Due to the clique structure in our system model, the techniques employed in existing works [2], [6], [7] cannot be directly applied here. To tackle this challenge, we develop an equivalent random graph $\mathbb{E}$ that exhibits the same phase transition behavior as the original model $\mathbb{H}$. Then, we characterize the phase transition behaviors in the graph $\mathbb{E}$ by capitalizing on the recent results in *inhomogeneous random graph* [8], [9].

We first construct an equivalent graph $\mathbb{E}$ based on the topology structure of $\mathbb{H}$. Since the nodes within the same

clique can immediately share the message, we treat each clique including affiliated online users as a single virtual node in $\mathbb{E}$. Furthermore, we assign type-1 and type-2 links between two virtual nodes according to the original connections in $\mathbb{H}$. To get a more concrete sense, we depict the equivalent graph in Fig. 2 that corresponds to the original model in Fig. 1. It is easy to see that the (type-1 and type-2) link degree of a virtual node equals the total number of (type-1 and type-2) links that are incident on the nodes within the corresponding clique. The equivalent graph $\mathbb{E}$ is expected to exhibit the same phase transition behavior as the original model $\mathbb{H}$ since both graphs have the similar connectivity structure. In particular, the fractional size of the giant component G_E in the equivalent graph $\mathbb{E}$ (the ratio of the number of nodes in G_E to the number of nodes in $\mathbb{E}$) is equal to the aforementioned quantity S_c. Thus, with a slight abuse of notation, we use S_c to denote the fractional size of G_E.

The degree of an arbitrary node in $\mathbb{E}$ can be represented by a two-dimensional vector $\boldsymbol{d} = [d^w \;\; d^f]$ where d^w and d^f correspond to the numbers of type-1 and type-2 links incident on that node, respectively. For a node in $\mathbb{E}$ that corresponds to a size-n clique in $\mathbb{W}$, we use K_n^w to denote its type-1 link degree, where K_n^w is a random variable following the distribution $\{P_{nk}^w, k = 0, 1, 2, ...\}$. Similarly, for a node in $\mathbb{E}$ that corresponds to a clique with m online users, we use K_m^f to denote its type-2 link degree where K_m^f follows the distribution $\{P_{mk}^f, k = 0, 1, 2, ...\}$. It is clear to see that an arbitrary node in $\mathbb{E}$ has link degree $[i \;\; j]$ with probability

$$p(i,j) = \sum_{n=1}^{D}\sum_{m=0}^{n} \mu_{nm} P_{ni}^w P_{mj}^f \quad i,j \in N. \tag{3}$$

Let $\mathrm{E}[d_w]$ and $\mathrm{E}[d_f]$ be the mean numbers of type-1 and type-2 links for a node in $\mathbb{E}$, i.e., $\mathrm{E}[d_w] = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} p(i,j)i$ and $\mathrm{E}[d_f] = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} p(i,j)j$. We also define $\mathrm{E}[d_w d_f] = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} p(i,j)ij$. Furthermore, let $\mathrm{E}[(d_w)^2]$ and $\mathrm{E}[(d_f)^2]$ denote the second moments of the number of type-1 and type-2 links for a node in $\mathbb{E}$, respectively; i.e., $\mathrm{E}[(d_w)^2] = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} p(i,j)i^2$ and $\mathrm{E}[(d_f)^2] = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} p(i,j)j^2$.

IV. Analytical solutions

In this section, we analyze information diffusion process by characterizing the phase transition behaviors in the equivalent random graph $\mathbb{E}$. We present our analytical results in the following two steps. We first quantify the conditions for the emergence of a giant component as well as the fractional sizes S_c and S_n for the special case $T_w = 1$ and $T_f = 1$. We next show that these results can be easily extended to a more general case with $0 \le T_w \le 1$ and $0 \le T_f \le 1$.

A. Special Case: $T_w = T_f = 1$

We characterize the phase transition behavior of the giant component in $\mathbb{E}$ by capitalizing the theory of inhomogeneous random graphs [8], [9], [10]. Specifically, we define $a_{11} = \mathrm{E}[(d_w)^2]/\mathrm{E}[d_w] - 1$, $a_{12} = \mathrm{E}[d_w \mathrm{d}_f]/\mathrm{E}[d_w]$, $a_{21} = \mathrm{E}[d_w \mathrm{d}_f]/\mathrm{E}[d_f]$ and $a_{22} = \mathrm{E}[(d_f)^2]/\mathrm{E}[d_f] - 1$. Along the same line in [2], [8], [10], we have the following result.

Lemma 4.1: Let

$$\sigma = \frac{1}{2}\left(a_{11} + a_{22} + \sqrt{(a_{11} - a_{22})^2 + 4a_{12}a_{21}}\right) \tag{4}$$

if $\sigma > 1$, with high probability (whp) there exists a giant component in $\mathbb{E}$, i.e., a non-trival fraction of nodes in $\mathbb{E}$ are connected; otherwise, a giant component does no exist in $\mathbb{E}$ whp.

The proof of Lemma 4.1 is relegated to Appendix in [11]. As we discussed in Section II-C, the existence of a giant component in $\mathbb{E}$ indicates that the information can reach a non-trival fraction of cliques in $\mathbb{H}$ rather than dying out quickly.

Next, let h_1 and h_2 in $(0, 1]$ be given by the smallest solution to the following recursive equations:

$$h_1 = \frac{1}{\mathrm{E}[d_w]}\sum_{n=1}^{D}\sum_{m=0}^{n} \mu_{nm}\mathrm{E}[K_n^w h_1^{K_n^w - 1}]\mathrm{E}[h_2^{K_m^f}], \tag{5}$$

$$h_2 = \frac{1}{\mathrm{E}[d_f]}\sum_{n=1}^{D}\sum_{m=0}^{n} \mu_{nm}\mathrm{E}[h_1^{K_n^w}]\mathrm{E}[K_m^f h_2^{K_m^f - 1}]. \tag{6}$$

We have the following results on the size and probability of an information epidemic.

Lemma 4.2: The fractional size of the giant component in $\mathbb{E}$ (equivalently, the fractional size of influenced cliques in $\mathbb{W}$) is given by

$$S_c = \sum_{n=1}^{D}\sum_{m=0}^{n} \mu_{nm}\left(1 - \mathrm{E}[h_1^{K_n^w}]\mathrm{E}[h_2^{K_m^f}]\right). \tag{7}$$

The fractional size of influenced nodes in $\mathbb{W}$ is given by

$$S_n = \frac{1}{C}\sum_{n=1}^{D}\sum_{m=0}^{n} n\mu_{nm}\left(1 - \mathrm{E}[h_1^{K_n^w}]\mathrm{E}[h_2^{K_m^f}]\right), \tag{8}$$

with the normalization term $C = \sum_{n=1}^{D} n\mu_n$.

The proof of Lemma 4.2 is relegated to Appendix in [11]. For any given set of parameters, Lemma 4.2 reveals the fraction of individuals and cliques that are likely to receive an information that is started from an arbitrary individual. Namely, an information started from an arbitrary individual gives rise to an information epidemic with probability S_n (attributed to the possibility that the arbitrary node belongs to the giant component G_H), and reaches a fraction S_n of nodes in the network. Similar conclusions can be drawn in terms of S_c for the fraction of cliques that receive the information.

Note that the condition (4) in Lemma 4.1 depends on the first/second moments of d_w and d_f, which boils down to the linear combinations of the first/second moments of k^w and k^f in the following manner:

$$\mathrm{E}[d_w] = \sum_{n=1}^{D}\mu_n^w n\mathrm{E}[k^w] \quad \mathrm{E}[d_f] = \sum_{m=1}^{D}\mu_m^f m\mathrm{E}[k^f], \tag{9}$$

$$\mathrm{E}[d_w d_f] = \sum_{n=1}^{D}\sum_{m=1}^{n}\mu_{nm} nm\mathrm{E}[k^w]\mathrm{E}[k^f], \tag{10}$$

$$\mathrm{E}[(d_w)^2] = \sum_{n=1}^{D} \mu_n^w \left(n\mathrm{E}[(k^w)^2] + (n^2 - n)(\mathrm{E}[k^w])^2 \right), \quad (11)$$

$$\mathrm{E}[(d_f)^2] = \sum_{m=1}^{D} \mu_m^f \left(m\mathrm{E}[(k^f)^2] + (m^2 - m)\big(\mathrm{E}[k^f]\big)^2 \right). \quad (12)$$

According to Lemma 4.2, S_c and S_n are determined by $\mathrm{E}[h_1^{K_n^w}]$, $\mathrm{E}[K_n^w h_1^{K_n^w - 1}]$, $\mathrm{E}[h_2^{K_m^f}]$ and $\mathrm{E}[K_m^f h_2^{K_m^f - 1}]$ in (5)-(8), i.e., the integrals with respect to the distributions of K_n^w and K_m^f for different n and m. The calculations can be simplified by utilizing the following transformations:

$$\mathrm{E}[h_1^{K_n^w}] = (\mathrm{E}[h_1^{k^w}])^n \quad \mathrm{E}[h_2^{K_m^f}] = (\mathrm{E}[h_2^{k^f}])^m, \quad (13)$$

$$\mathrm{E}[K_n^w h_1^{K_n^w - 1}] = n\Big(\mathrm{E}[h_1^{k^w}]\Big)^{n-1} \mathrm{E}[k^w h_1^{k^w - 1}], \quad (14)$$

$$\mathrm{E}[K_m^f h_2^{K_m^f - 1}] = m\Big(\mathrm{E}[h_2^{k^f}]\Big)^{m-1} \mathrm{E}[k^f h_2^{k^f - 1}]. \quad (15)$$

With the help of (13)-(15), we only need to calculate the integrals with respect to the distributions of k^w and k^f. In this way, we can find h_1 and h_2 by numerically solving the recursive equations (5)-(6) and compute S_c and S_n from (7)-(8), respectively. Due to the space limitations, the detailed derivations of (9)-(15) are omitted (see details in Appendix of [11]).

B. General Case: $0 \le T_w \le 1$ *and* $0 \le T_f \le 1$

We next generalize Lemma 4.1 and Lemma 4.2 to the case $0 \le T_w \le 1$ and $0 \le T_f \le 1$. To this end, we maintain the *occupied* links in the equivalent graph $\mathbb{E}$ by deleting each type-1 and type-2 edge with probability $1 - T_w$ and $1 - T_f$, respectively. Let $\tilde{k}^w$ and $\tilde{k}^f$ be the occupied link degrees (instead of k^w and k^f) with the distributions $\{\tilde{p}_k^w, k = 0, 1, ...\}$ and $\{\tilde{p}_k^f, k = 0, 1, ...\}$. According to [7], the generating functions corresponding to $\tilde{k}^w$ and $\tilde{k}^f$ can be given by

$$\tilde{g}(x) = g\left(1 + T_w(x-1)\right) \quad \tilde{q}(x) = q\left(1 + T_f(x-1)\right). \quad (16)$$

From (9)-(15), we observe that the critical threshold and the giant component size are determined by the distributions of k^w and k^f. Therefore, Lemma 4.1 and Lemma 4.2 still hold if we replace the terms associated with k^w and k^f in (9)-(15) by those associated with $\tilde{k}^w$ and $\tilde{k}^f$, respectively. To this end, by using the generating functions (16), we find

$$\begin{aligned} \mathrm{E}[\tilde{k}^w] &= T_w \mathrm{E}[k^w], \\ \mathrm{E}[(\tilde{k}^w)^2] &= T_w^2 \left(\mathrm{E}[(k^w)^2] - \mathrm{E}[k^w] \right) + T_w \mathrm{E}[k^w]. \end{aligned}$$

In the same manner, we can compute $\mathrm{E}[\tilde{k}^f]$ and $\mathrm{E}[(\tilde{k}^f)^2]$. The critical threshold (in the general case) can now be computed by replacing $\mathrm{E}[k^w]$, $\mathrm{E}[k^f]$, $\mathrm{E}[(k^w)^2]$, $\mathrm{E}[(k^f)^2]$ with $\mathrm{E}[\tilde{k}^w]$, $\mathrm{E}[\tilde{k}^f]$, $\mathrm{E}[(\tilde{k}^w)^2]$, $\mathrm{E}[(\tilde{k}^f)^2]$, respectively, in (9)-(12).

In order to compute the giant component size, we only need to replace the corresponding terms in (13)-(15) with $\mathrm{E}[h_1^{\tilde{k}^w}]$, $\mathrm{E}[h_2^{\tilde{k}^f}]$, $\mathrm{E}[\tilde{k}^w h_1^{\tilde{k}^w - 1}]$ and $\mathrm{E}[\tilde{k}^f h_2^{\tilde{k}^f - 1}]$. By using (16), we have

$$\mathrm{E}[h_1^{\tilde{k}^w}] = \tilde{g}(h_1) = \mathrm{E}[(1 + T_w(h_1 - 1))^{k^w}],$$

TABLE I
THE CLIQUE SIZE DISTRIBUTION IN FOUR SCENARIOS

scenario	size-1	size-2	size-3	average clique size
1	100%	0	0	1
2	66.7%	33.3%	0	1.333
3	33.3%	66.7%	0	1.666
4	33.3%	33.3%	33.3%	2

$$\mathrm{E}[\tilde{k}^w h_1^{\tilde{k}^w - 1}] = [\tilde{g}(h_1)]' = T_w \mathrm{E}[k_w(1 + T_w(h_1 - 1))^{k^w - 1}].$$

Similar relations can be obtained for $\mathrm{E}[h_1^{\tilde{k}^f}]$ and $\mathrm{E}[\tilde{k}^f h_1^{\tilde{k}^f - 1}]$. The size of the giant component (in the general case) can now be computed by reporting the updated (13)-(15) into (5)-(8).

V. NUMERICAL RESULTS AND SIMULATIONS

In this section, we numerically study the diffusion of real-time information by utilizing the analytical results derived in Section IV. In particular, we are interested in how the clique structure can impact the scale of information epidemic. To get a more concrete sense, we compare four system scenarios, each with different clique size distribution as illustrated in Table I.

For the sake of fair comparison, the total number of nodes in $\mathbb{W}$ is fixed at 12000 in each scenario. From Table I, we can see that the average clique size increases from scenario 1 to scenario 4, indicating that individuals are getting closer to each other. We assume that the type-1 link degree for each node in $\mathbb{W}$ follows a poisson distribution, i.e., $p_k^w = \frac{\lambda^k}{k!} \cdot e^{-\lambda}$, $k = 0, 1, 2, ...$, where λ is the average type-1 link degree. Meanwhile, the type-2 link degree for each online user in $\mathbb{F}$ follows a power-law distribution with exponential cutoff, i.e., $p_0^f = 0$ and

$$p_k^f = \frac{1}{C} k^{-\gamma} e^{-\frac{k}{\Gamma}}, \quad k = 1, 2, \ldots, \quad (17)$$

with the normalization factor $C = \sum_{k=1}^{\infty} k^{-\gamma} e^{-\frac{k}{\Gamma}}$.

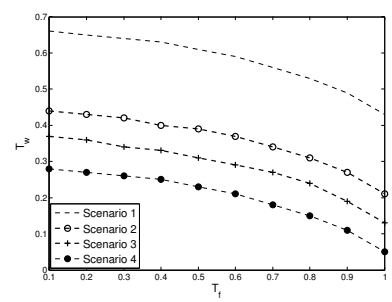

Fig. 3. The minimum T_w required for the existence of a giant component in $\mathbb{E}$ versus T_f in four scenarios. We let $\lambda = 1.5$, $\alpha = 0.1$, $\gamma = 3$ and $\Gamma = 10$. Each curve corresponds to the boundary of the phase transition in one scenario. A giant component is very likely to emerge above the boundary.

We first compare these scenarios in term of the required conditions for the existence of a giant component; in other words, in terms of the minimum conditions for an information epidemic to take place. We let $\lambda = 1.5$, $\alpha = 0.1$, $\gamma = 3$ and $\Gamma = 10$. By computing the system's critical threshold, we

depict in Figure 3 the minimum required value of T_w to have a giant component in $\mathbb{E}$ versus T_f. Each scenario corresponds to a curve in the figure that stands for the boundary of a phase transition; above the boundary a giant component is very likely to emerge. By the definition of transmissibility, a lower required T_w indicates that the system is more likely to give rise to a giant component (and thus, to an information epidemic). From scenario 1 to scenario 4, this figure clearly shows that larger clique sizes lead to smaller values for the minimum T_w required for an information epidemic, meaning that information epidemics are more likely to take place for larger clique sizes. The analytical results of Figure 3 can also be verified by simulations. Due to the space limitations, we provide the corresponding numerical results in [11].

We next compare these scenarios in terms of the fractional sizes of influenced cliques and influenced individuals. For each scenario, we plot the fractional size of the giant component in $\mathbb{E}$ versus T_f in Figure 4, which indicates the fraction of cliques that will receive the information. We set $T_w = 0.3$, $\lambda = 2$, $\alpha = 0.3$, $\gamma = 3$ and $\Gamma = 10$. In this Figure, the curves stand for analytical results obtained by (7), while the marked points stand for the simulation results obtained by averaging 200 experiments for each set of parameter. It is easy to check that the analytical results are in good agreement with the simulations. Obviously, the fractional size of influenced cliques in scenario 4 (with average clique size 2) is much larger than that in scenario 1 (with average clique size 1), which indicates that large cliques in the social network could greatly facilitate the message propagation.

We finally compare the fractional size of influenced individuals in Figure 5. In this figure, the curves stand for the fractional size of influenced nodes obtained via (8), whereas the marked points stand for the simulation results. Similar to Figure 4, the information is much easier to propagate in a social network when larger size cliques exist. For instance, when $T_f = 1$, the fractional size of individuals that receive the message grows sharply from 14% (scenario 1 with average clique size 1) to 80% (scenario 4 with average clique size 2). In conclusion, the above results agree with a natural conjecture that the messages are more influential (i.e., more likely to reach a large portion of the population) when people are close to each other.

VI. CONCLUSION

In this study, we explore the diffusion of real-time information in social networks. We develop an overlaying social-physical network that consists of an online social network and a physical information network with clique structure. We theoretically quantify the condition and the size of information epidemics. To the best of our knowledge, this paper is the first work on the diffusion of real-time information with consideration on the clique structure in social networks. We believe that our studies will offer initial steps towards understanding the diffusion behaviors of real-time information.

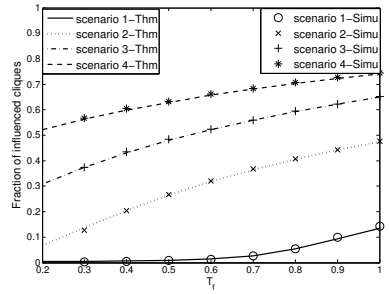

Fig. 4. Fraction size of influenced cliques in $\mathbb{W}$ versus T_f. We set $T_w = 0.3$, $\lambda = 2$, $\alpha = 0.3$, $\gamma = 3$ and $\Gamma = 10$. The curves stand for analytical results obtained by (7), while the marked points stand for the simulation results with $N = 12000$ by averaging 200 experiments for each set of parameter. The analytical results are in good agreement with the simulations.

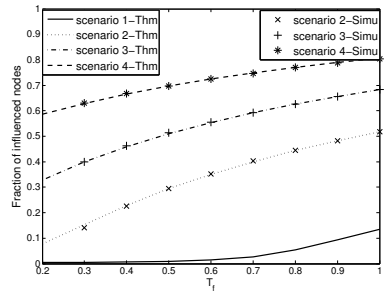

Fig. 5. Fraction size of influenced nodes in $\mathbb{W}$ versus T_f. All the parameters follow the same setup as in Figure 4. The curves stand for analytical results obtained by (8) and the marked points stand for the simulation results with $N = 12000$. The analytical results are in good agreement with the simulations. For comparison, we also plot the fraction size of influenced cliques in scenario 1 where the each clique has only one node.

REFERENCES

[1] K. Zhao, J. Stehle, G. Bianconi, and A. Barrat. Social network dynamics of face-to-face interactions. *Phys. Rev. E*, 83(5):056109, 2011.

[2] O. Yağan, D. Qian, J. Zhang, and D. Cochran. Conjoining speeds up information diffusion in overlaying social-physical networks. Technical report, Available online at arXiv:1112.4002v2[cs.SI].

[3] O. Yağan and V. Gligor. Analysis of complex contagions in random multiplex networks. Technical report, Available online at arXiv:1204.0491v3[physics.soc-ph].

[4] J. Yang and J. Leskovec. Modeling information diffusion in implicit networks. In *Proceedings of the 10th IEEE International Conference on Data Mining*, 2010.

[5] J. Stehlé, A. Barrat, and G. Bianconi. Dynamical and bursty interactions in social networks. *Phys. Rev. E*, 81(3):035101, 2010.

[6] 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), 2001.

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

[8] B. Söderberg. General formalism for inhomogeneous random graphs. *Phys. Rev. E*, 66(066121), 2002.

[9] B. Söderberg. Random graphs with hidden color. *Phys. Rev. E*, 68(015102(R)), 2003.

[10] B. Söderberg. Random graph models with hidden color. *Acta Phys. Pol. B*, 34:5085–5102, 2003.

[11] D. Qian, O. Yağan, L. Yang, and J. Zhang. Diffusion of real-time information in social-physical networks. Technical report, Available online at arXiv:1203.6122[cs.SI].