Networking Science manuscript No.
(will be inserted by the editor)

Diffusion of Real-Time Information in Overlaying Social-Physical Networks: Network Coupling and Clique Structure

Dajun Qian · Osman Yağan · Lei Yang · Junshan Zhang and Kai Xing

Received: date / Accepted: date

Abstract We study the diffusion behavior of real-time information in an overlaying social-physical network. Typically, real-time information is valuable only for a limited time duration and needs to be delivered before its deadline, indicating that real-time information is more likely to spread among friends within a "social proximity." With this insight, we consider a *physical information network* which consists of many cliques and assume that real-time information can spread quickly within a clique. Conjoint to this physical information network, there are *online social networks* where the information can propagate via websites such as Facebook, twitter, Youtube, etc.

Capitalizing on the theory of inhomogeneous random graph, we analytically characterize the size of information epidemic. One interesting finding is that a larger size online social network, with the same degree distribution, may not necessarily yield a larger size of information epidemic in this overlaying social-physical network. In fact, under certain conditions, the size of information epidemic could even decrease with the growing size of the online social network. This is in stark contrast to that in a single network.

Keywords Social network · Real-time information · Information diffusion · Random graph theory

Dajun Qian · Lei Yang · Junshan Zhang
School of ECEE, Arizona State University, Tempe, AZ, USA.
E-mail: dqian@asu.edu

Osman Yağan
Cybersecurity Laboratory (CyLab), Carnegie Mellon University, Pittsburgh, PA, USA.

Kai Xing
School of Computer Science and Technology, University of Science and Technology of China, Anhui, P.R. China.

This work was supported in part by the US National Science Foundation under CNS-1218484 and in part by the US DTRA under Grant HDTRA1-13-1-0029.

1 Introduction

1.1 Motivation and Background

In today's modern society, people are becoming increasingly connected over social networks. Thanks to *online social networks*, such as Facebook and Twitter, people can share messages quickly with their friends. Meanwhile, *physical information networks* [2], based on traditional face-to-face interactions, still remain as an important medium for message spreading. These networks are increasingly becoming *coupled* together [3] due to individuals that participate in multiple of them. As a result of such coupling, the information propagation in one network can trigger further propagation in another, and vice versa, greatly facilitating the diffusion of information [4]. This leads today's hot spot news or fashion behaviors to generate pronounced influence over the population more than ever before.

The main thrust of this study is dedicated to understanding the diffusion behavior of real-time information over coupled networks. Typically, the real-time information is valuable only for a limited time duration [5] and hence needs to be delivered before its deadline. For example, once a time-limited coupon is released from Groupon or Dealsea.com, people can share this news either by talking to friends or posting it on Facebook. However, the interest on this deal would die down after it expires.

Clearly, due to the timeliness requirement, the influence of real-time information depends on its propagation speed. The faster the message passes from one to another, the more people can learn this news before it expires, indicating that its diffusion behavior hinges heavily on 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. Observe that the efficiency of face-to-face communications depends on the physical distance between individuals, but in an online social network, message spreading depends mainly on online connections (not on physical distance). Recent works [2,6] have explored the structure of physical information network by tracking in-person interactions over the population, and their findings indicate that such interactions would give rise to a social graph consisting of a large number of small *cliques*, which are somewhat loosely connected to each other. Each clique therein stands for a group of people who are close to each other. The message can spread quickly within a clique via frequent face-to-face interactions, but takes longer time to spread across cliques separated by longer distances. Clearly, constrained by its limited propagation time, 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 the clique structure, which is missing in other related works [3,7,8].

1.2 Summary of Main Contributions

We study the diffusion of real-time information in an overlaying social-physical network. In this study, we consider a *physical information network* where the message could spread amongst people through face-to-face contacts. Furthermore, the information could also propagate via an *online social network* conjoint to this physical information network. For convenience, we refer to the physical information network simply as the *physical network* and refer to the online social network simply as the *social network*. Hence, the overall system is termed as the coupled (or overlaying) social-physical network.

Specifically, we investigate the information diffusion under two scenarios, namely, coupled-network model I and coupled-network model II, as illustrated in Fig. 1. In model I, we assume that all nodes in the social network are also in the physical network, i.e., the collection of online users is a proper subset of the individuals in the physical network. In model II, we consider a more general case where the social network also has online users who do not belong to the physical network. As illustrated in Fig. 1, the social and physical networks are "partially overlapping" and the overlapping fraction represents the collection of online users who are also in the physical network.

In both models, we characterize the information diffusion process by studying the *phase transition behaviors* of the underlying random graph models (see Section 2.3 for details). Specifically, we show that the system model has a *critical threshold* above which *information epidemics* can take place, i.e., the information can reach a non-trivial fraction of individuals. We also quantify the number of individuals that finally receive the message by computing the size of the *giant component* in the induced random graph model. One interesting finding is that a larger size social network may not always yield a larger size of information epidemic in this coupled social-physical network. Specifically, we show that given the fixed degree distribution, the growing size of the social network could essentially reduce the *coupling strength* between two networks. Under certain conditions, due to the reduction in the network coupling, the size of information epidemic could decrease with the growing size of the social network while fixing its degree distribution. This is in stark contrast to the information diffusion behavior in a single network.

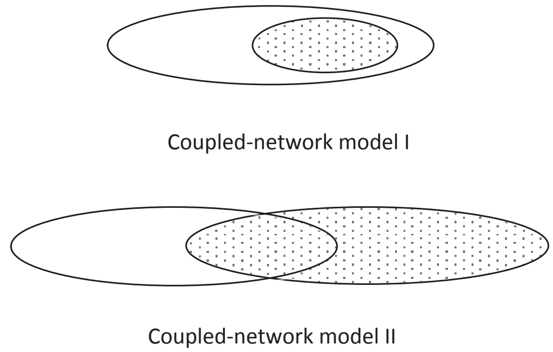

Fig. 1 An illustration of two models. The blank ellipse and the dotted ellipse stand for the physical network and the social network, respectively. In coupled-network model II, the overlapping fraction between two networks represents the collection of online users who are also in the physical network, while the other fraction in the social network represents the collection of online users who are outside the physical network.

In related work, it is assumed [7,8] that the message propagates at the same speed along different social relationships. Clearly, this assumption is not appropriate for the diffusion of real-time information, where propagation speeds play a key role. Very recent work [3] considered online connections and face-to-face connections for general information diffusion, but did not study the impact of the clique structure on information diffusion. To the best of our knowledge, this paper is the first attempt on the diffusion of real-time information while considering the clique structure in social networks. We believe that our work will offer initial steps towards understanding the diffusion behaviors of real-time information in a coupled social-physical network.

2 Coupled-Network Model I

2.1 Illustration of Model Structure

Fig. 2 illustrates the structure of model I. We consider an overlaying social-physical network $\mathbb{H}$ that consists of a physical network $\mathbb{W}$ and a social network $\mathbb{F}$. The collection of the nodes $\mathscr{N}_W = \{1,2,...,N\}$ in the physical network $\mathbb{W}$ stand

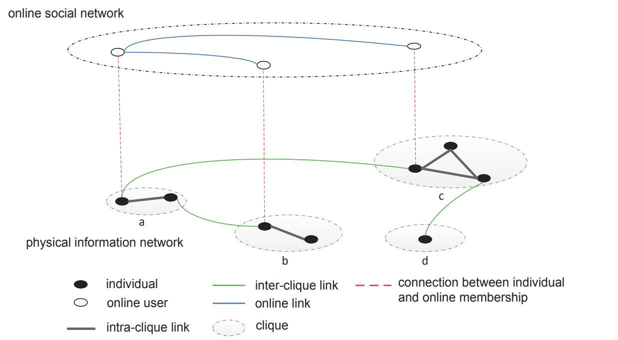

Fig. 2 Structure of coupled-network model I

for the human beings in the real world. Meanwhile, each node in $\mathbb{W}$ is also a member of the social network $\mathbb{F}$ with probability α, and the collection of nodes in $\mathbb{F}$, denoted by $\mathscr{N}_F$, stand for their online memberships[1]. We also refer to the nodes in $\mathbb{W}$ and $\mathbb{F}$ as "individuals" and "online users," respectively.

Cliques in the physical information network. Based on empirical studies in [2,6], we assume that the physical network has N nodes which are gathered into many cliques with different sizes. Each clique represents a group of people with frequent face-to-face interactions, e.g., family in a house or colleagues in an office. It is assumed that 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 cliques as follows: at step $t=1$, we randomly choose n nodes from the collection $\mathscr{N}_W$ 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 $\mathscr{N}_W$ by $\mathscr{N}_1$. At each step t, we repeat the above procedure to create a new clique from the collection $\mathscr{N}_{t-1}$[2], and assume that we can finally generate N_c cliques in $\mathbb{W}$[3]. It follows that $N = N_c \sum_n n\mu_n^w$. Generally speaking, the existence of large size cliques indicates that many individuals are close to each other.

As we elaborate in the following, the links connecting the nodes in $\mathbb{W}$ stand for traditional face-to-face connections, while the links in $\mathbb{F}$ represent online connections.

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" in social networks [9], in the sense that the nodes in a clique are fully connected to each other.

Type-1 (inter-clique) links in $\mathbb{W}$. We assume that a face-to-face interaction is still possible to happen 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 social network $\mathbb{F}$ represent the online users. As in [3], 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 person.

Online users associated with a clique. To avoid confusions, we say *"size-n clique with m online members"* when referring to the case that a clique contains n individuals and only m of them participate in the social network $\mathbb{F}$. Specifically, for the collection of size-n cliques with m online members, $m \le n \le D$, we assume that their fractional size in the whole collection of cliques is μ_{nm}. It is easy to see that

$$\mu_{nm} = \mu_n^w \binom{n}{m} \alpha^m (1-\alpha)^{n-m} \quad \text{and} \quad \mu_n^w = \sum_{m=1}^{n} \mu_{nm}. \qquad (1)$$

2.2 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 with fast propagation speed. With this insight, we assign each link with a *transmissibility* as in [3,8], i.e., the probability that the message can successfully pass through.

For ease of exposition, 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, 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.

2.3 Information Cascade

We give a brief description of the information diffusion process in the following. All individuals in network $\mathbb{W}$, i.e., the

[1] Throughout, we use "nodes in $\mathbb{W}$" and "nodes in $\mathscr{N}_W$" interchangeably. So it is the same with the social network $\mathbb{F}$ and $\mathscr{N}_F$.

[2] 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 would be negligible if the number of cliques is large enough.

[3] Throughout, 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 $\mathbb{H}$.

collection of nodes in $\mathcal{N}_W$, are potential information recipients. 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 nodes in other cliques through occupied type-1 and type-2 links. This process could continue iteratively in this manner and may eventually lead to an information epidemic; i.e., a non-zero fraction of individuals may receive the information in the limit $N \to \infty$ [3].

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 largest connected components of $\mathbb{H}$, which contains only the *occupied* type-1 and type-2 links [3, 8, 7] of $\mathbb{H}$. Thus, the information diffusion process considered here is equivalent to a heterogeneous bond-percolation process over $\mathbb{H}$; 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 behaviors of $\mathbb{H}$. A key observation is that the system $\mathbb{H}$ exhibits a *phase transition* behavior at a *critical threshold* [3]. Specifically, a *giant connected component* G_H that covers a non-trivial fraction of $\mathbb{H}$ is likely to appear above the critical threshold, meaning that *information epidemics* are possible. Below that critical threshold, all the connected components in $\mathbb{H}$ are small indicating that the influenced individuals fraction tends to zero in the large network 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 size of information epidemic:

- S_c: The fractional size of the influenced cliques in $\mathbb{W}$. Specifically, S_c is the ratio of the number of the cliques contained in G_H to the total number of cliques in $\mathbb{W}$.
- S_n: The fractional size of the influenced individuals in $\mathbb{W}$. Specifically, S_n is the ratio of the number of the individuals contained in G_H to the total number of nodes in $\mathcal{N}_W$.

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

3 Abstract Mapping Graph: a Clique Level Approach

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

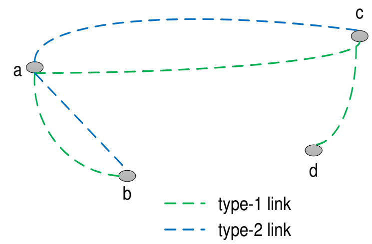

Fig. 3 Abstract mapping graph $\mathbb{E}$. Nodes {a,b,c,d} in this graph corresponds to the cliques {a,b,c,d} of $\mathbb{H}$ in Fig. 2. We assign type-1 and type-2 links in $\mathbb{E}$ according to the same types of links connecting cliques in Fig. 2.

- What is the critical threshold of $\mathbb{H}$? 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 nodes fraction and cliques fraction 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 $\mathbb{H}$. Due to the clique structure in our system model, the techniques employed in existing works [3, 7, 8] cannot be directly applied here. To tackle this challenge, we develop an abstract mapping random graph $\mathbb{E}$ that exhibits the same phase transition behavior as $\mathbb{H}$. Then, we characterize the phase transition behaviors in the graph $\mathbb{E}$ by capitalizing on the recent results in *inhomogeneous random graph* [10, 11].

We first construct a graph $\mathbb{E}$ which is an abstract mapping of the graph $\mathbb{H}$. Specifically, each node in the graph $\mathbb{E}$ represents a cluster in the graph $\mathbb{H}$ and each link in the graph $\mathbb{E}$ corresponds to a link connecting two clusters in the graph $\mathbb{H}$. 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 abstract mapping graph in Fig. 3 that corresponds to the original model I in Fig. 2. 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 graph $\mathbb{E}$ is expected to exhibit the same phase transition behavior as the original model $\mathbb{H}$ since both graphs have the similar graph interconnection pattern. In particular, the fractional size of the giant component G_E in the 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 fraction 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 $d = [d^w\ d^f]$ 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 $\{P_{nk}^w, k=0,1,2,...; n=1,2,...,D\}$. 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 $\{P_{mk}^f, k=0,1,2,...; m=0,1,...,D\}$. It is clear to see that an arbitrary node in $\mathbb{E}$ has link degree $[i\ j]$ with probability

$$p(i,j) = \sum_{n=1}^{D}\sum_{m=0}^{n} \mu_{nm} P_{ni}^w P_{mj}^f \quad i,j \in N. \tag{2}$$

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 $\mathrm{E}[(d_f)^2] = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} p(i,j)j^2$.

4 Analytical Solutions

In this section, we analyze information diffusion process by characterizing the phase transition behaviors in the 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 $0 \le T_f \le 1$.

In what follows, we characterize the phase transition behavior of the giant component in $\mathbb{E}$ by capitalizing on the theory of inhomogeneous random graphs [10, 11, 12]. 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 [3, 10, 12], we have the following result.

Lemma 4.1 *Let*

$$\sigma = \frac{1}{2}\left(a_{11} + a_{22} + \sqrt{(a_{11} - a_{22})^2 + 4a_{12}a_{21}}\right) \tag{3}$$

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 8.1. As we discussed in Section 2.3, 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:

$$h_1 = \frac{1}{\mathrm{E}[d_w]}\sum_{n=1}^{D}\sum_{m=0}^{n} \mu_{nm} \mathrm{E}[K_n^w h_1^{K_n^w - 1}]\mathrm{E}[h_2^{K_m^f}], \tag{4}$$

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

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*

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

The fractional size of influenced individuals in $\mathbb{W}$ (equivalently, the influenced nodes fraction in $\mathscr{N}_{\mathbb{W}}$) is given by

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

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

The proof of Lemma 4.2 is relegated to Appendix 8.1. For any given set of parameters, Lemma 4.2 reveals the individuals fraction and cliques fraction that are likely to receive an information that is started from an arbitrary individual.

We next generalize Lemma 4.1 and Lemma 4.2 to the case $0 \le T_w \le 1$ and $0 \le T_f \le 1$. We first break down the first/second moments of d_w and d_f from the condition (3) in Lemma 4.1 into the linear combinations of the first/second moments of k^w and k^f as follows:

$$\mathrm{E}[d_w] = \sum_{n=1}^{D} \mu_n^w n \mathrm{E}[k^w] \quad \mathrm{E}[d_f] = \sum_{m=1}^{D} \mu_m^f m \mathrm{E}[k^f], \tag{8}$$

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

$$\mathrm{E}[(d_w)^2] = \sum_{n=1}^{D} \mu_n^w \left(n\mathrm{E}[(k^w)^2] + (n^2 - n)(\mathrm{E}[k^w])^2\right), \tag{10}$$

$$\mathrm{E}[(d_f)^2] = \sum_{m=1}^{D} \mu_m^f \left(m\mathrm{E}[(k^f)^2] + (m^2 - m)\left(\mathrm{E}[k^f]\right)^2\right). \tag{11}$$

Similarly, $\mathrm{E}[h_1^{K_n^w}]$, $\mathrm{E}[K_n^w h_1^{K_n^w - 1}]$, $\mathrm{E}[h_2^{K_m^f}]$ and $\mathrm{E}[K_m^f h_2^{K_m^f - 1}]$ in (4)-(7) can boil down to the integrals with respect to the distributions of k^w and k^f by utilizing the following transformations:

$$\mathrm{E}[h_1^{K_n^w}] = (\mathrm{E}[h_1^{k^w}])^n \quad \mathrm{E}[h_2^{K_m^f}] = (\mathrm{E}[h_2^{k^f}])^m, \tag{12}$$

$$\mathrm{E}[K_n^w h_1^{K_n^w - 1}] = n\left(\mathrm{E}[h_1^{k^w}]\right)^{n-1}\mathrm{E}[k^w h_1^{k^w - 1}], \tag{13}$$

$$\mathrm{E}[K_m^f h_2^{K_m^f - 1}] = m\left(\mathrm{E}[h_2^{k^f}]\right)^{m-1}\mathrm{E}[k^f h_2^{k^f - 1}]. \tag{14}$$

In this way, the calculations in (3)-(7) can be simplified by utilizing (8)-(14).

As aforementioned in Section 2.2, for the case with $0 \leq T_w \leq 1$ and $0 \leq T_f \leq 1$, the original degree distributions k^w and k^f should be replaced by the degree distributions of occupied links. Specifically, we maintain the *occupied* links in the graph $\mathbb{E}$ by deleting each type-1 and type-2 link 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 $\{\tilde{p}^w_k, k=0,1,...\}$ and $\{\tilde{p}^f_k, k=0,1,...\}$. According to [8], the generating functions corresponding to $\tilde{k}^w$ and $\tilde{k}^f$ can be given by

$$\tilde{g}(x) = g\left(1+T_w(x-1)\right) \quad \tilde{q}(x) = q\left(1+T_f(x-1)\right). \tag{15}$$

From (8)-(14), 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 (8)-(14) by those associated with $\tilde{k}^w$ and $\tilde{k}^f$, respectively. To this end, by using the generating functions (15), we find

$$\mathrm{E}[\tilde{k}^w] = T_w \mathrm{E}[k^w],$$

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

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 (8)-(11).

In order to compute the giant component size, we only need to replace the corresponding terms in (12)-(14) 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 (15), we have

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

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

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 (12)-(14) into (4)-(7).

5 Coupled-Network Model II

In practical scenarios, the social networks, e.g., Facebook and Twitter, enable the message to reach remote online users from other cities or countries. With this insight, we consider a more realistic model II, where the social network also has online users outside the physical network. Specifically, we assume that the physical network has the same clique structure as in model I. A fraction α of the total N nodes in the physical network are also online users in social network $\mathbb{F}$. For convenience, we denote the collection of these online users from the physical network as $\mathcal{N}_{F_1}$ and hence $|\mathcal{N}_{F_1}| = \alpha N$. In contrast to model I, we assume that the nodes in $\mathcal{N}_{F_1}$ only occupy a fraction β of the total online users, $\beta \in (0,1]$, since the social network also has online users who do not belong to physical network $\mathbb{W}$. Therefore, the size of social network $\mathbb{F}$ turns out to be $|\mathcal{N}_F| = \alpha N/\beta$. Moreover, we use $\mathcal{N}_{F_2}$ to denote the collection of the online users outside the physical network and hence $|\mathcal{N}_{F_2}| = \alpha N(1-\beta)/\beta$. Clearly, when $\beta = 1$, model II reduces to model I as a special case.

By doing so, the collection of the potential information recipients extends from $\mathcal{N}_W$ to $\mathcal{N}_W \cup \mathcal{N}_{F_2}$. We are particularly interested to see the size of information epidemic among the overall population, equivalently, the number of the influenced nodes in $\mathcal{N}_W \cup \mathcal{N}_{F_2}$. We tackle this problem by transforming the coupled networks into an abstract mapping graph defined in Section 3. Specifically, we assume that each node in $\mathcal{N}_{F_2}$ has a virtual counterpart node in $\mathbb{W}$, which has no type-1 links. Each pair of these two nodes can be treated as a single virtual node in the abstract mapping graph $\mathbb{E}$ and this node only has type-2 links with degree distribution $\{p^f_k, k=0,1,...\}$. By definition of the abstract mapping graph in Section 3, the fractional size of such virtual nodes can be given by $\nu^w_0 = \frac{|\mathcal{N}_{F_2}|}{N_c+|\mathcal{N}_{F_2}|}$. Furthermore, we use ν^w_n to denote the fractional size of the virtual nodes in $\mathbb{E}$ which correspond to size-n cliques and it follows that $\nu^w_n = \frac{N_c}{N_c+|\mathcal{N}_{F_2}|}\mu^w_n$. We also use ν^w_{nm} to denote the fractional size of the virtual nodes which correspond to the size-n cliques with m online users, following the similar definitions as (1).

The abstract mapping graph $\mathbb{E}$ is an inhomogeneous random graph with two types of links. As in Section 3, an arbitrary node in $\mathbb{E}$ has link degree $[i\ j]$ with probability

$$p(i,j) = \sum_{n=1}^{D}\sum_{m=0}^{n} \nu_{nm} P^w_{ni} P^f_{mj} + \nu^w_0 P^w_{0i} P^f_{1j} \quad i,j \in N, \tag{16}$$

where P_{ni} and P_{mj} follow the same definitions in (2) and

$$P^w_{0i} = \begin{cases} 1 & \text{if i} = 0, \\ 0 & \text{if i} > 0. \end{cases}$$

Based on the abstract mapping graph, the phase transition behaviors in model II can be characterized in the same way as in model I, only with a different degree distribution in $\mathbb{E}$. Therefore, we can still use Lemma 4.1 to characterize the existence condition of the giant component in $\mathbb{E}$, which indicates the outbreak of the information epidemic.

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

$$h_1 = \frac{1}{\mathrm{E}[d_w]}\sum_{n=1}^{D}\sum_{m=0}^{n} \nu_{nm} \mathrm{E}[K^w_n h_1^{K^w_n-1}]\mathrm{E}[h_2^{K^f_m}], \tag{17}$$

$$h_2 = \frac{1}{\mathrm{E}[d_f]}\left(\sum_{n=1}^{D}\sum_{m=0}^{n} \nu_{nm} \mathrm{E}[h_1^{K^w_n}]\mathrm{E}[K^f_m h_2^{K^f_m-1}] + \nu^w_0 \mathrm{E}[K^f_m h_2^{K^f_m-1}]\right). \tag{18}$$

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

Lemma 5.3 *The fractional size of the giant component in $\mathbb{E}$ is given by*

$$S_c = \sum_{n=1}^{D}\sum_{m=0}^{n} v_{nm} n \left(1 - \mathrm{E}[h_1^{K_n^w}]\mathrm{E}[h_2^{K_m^f}]\right) + v_0^w \left(1 - \mathrm{E}[h_2^{K_m^f}]\right). \tag{19}$$

The fractional size of influenced nodes in $\mathcal{N}_W \cup \mathcal{N}_{F_2}$ is given by

$$S_n = \frac{1}{C}\left(\sum_{n=1}^{D}\sum_{m=0}^{n} n v_{nm}\left(1 - \mathrm{E}[h_1^{K_n^w}]\mathrm{E}[h_2^{K_m^f}]\right) + v_0\left(1 - \mathrm{E}[h_2^{K_m^f}]\right)\right) \tag{20}$$

with the normalization term $C = \sum_{n=1}^{D} n v_n + v_0$.

Lemma 5.3 can be proved in the same way as in Lemma 4.2. Note that in model II the collection of information recipients becomes $\mathcal{N}_W \cup \mathcal{N}_{F_2}$. Therefore, the total number of the influenced nodes turns out to be $(N + |\mathcal{N}_{F_2}|) S_n$. Furthermore, Lemma 5.3 can be generalized to the case with $0 \le T_w \le 1$ and $0 \le T_f \le 1$ along the same line as in model I.

6 Numerical Results

In Section 4 and 5 we analyzed the critical threshold and the size of information epidemic in both model I and model II. To get a more concrete sense of the analytical results, we study the diffusion behavior of real-time information via numerical examples. Particularly, we focus on two main features in our models, i.e., the clique structure and the network coupling, and their impacts on information diffusion.

6.1 Information Epidemics and Clique Structure

We first investigate how the clique structure could impact information diffusion. Particularly, we compare four scenarios in model I, each with different clique size distribution as illustrated in Table 1.

For the sake of fair comparison, the total number of nodes in the physical network is fixed at 12000 in each scenario. From scenario 1 to scenario 4, we can see that the average clique size increases from 1 to 2, 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 the social network follows a power-law distribution with exponential cutoff, i.e., $p_0^f = 0$ and

$$p_k^f = \frac{1}{C} k^{-\gamma} e^{-\frac{k}{\Gamma}}, \ \ k = 1, 2, \ldots, \tag{21}$$

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

We compare the sizes of information epidemic in terms of the influenced nodes fraction in $\mathbb{W}$. For each scenario,

Table 1 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

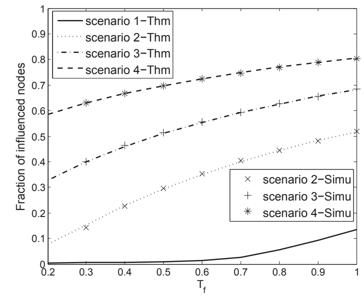

Fig. 4 The influenced nodes fraction in $\mathcal{N}_W$ versus T_f. The curves stand for analytical results obtained by (7) and the marked points stand for the simulation results. The analytical results are in good agreement with the simulations. For comparison, we also plot the fractional size of influenced cliques in scenario 1 where the each clique has only one node.

we plot the fractional size of the nodes that will receive the information versus T_f in Figure 4. 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. We set $T_w = 0.3$, $\lambda = 2$, $\alpha = 0.3$, $\gamma = 3$ and $\Gamma = 10$. It is easy to check that the analytical results are in good agreement with the simulations. Obviously, the information is much easier to propagate when larger size cliques exist. For instance, when $T_f = 1$, as the average clique size increases from 1 (scenario 1) to 2 (scenario 4), the fractional size of individuals that receive the message grows sharply from 14% to 80%. 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.

6.2 Information Epidemics and Network Coupling

We next investigate how the coupling between the social and physical networks could facilitate the information diffusion. As illustrated in Fig. 1, we say the social and physical networks are coupled in the sense that a fraction of nodes are in both networks. Generally speaking, the *coupling strength* between two networks depends on the fractional size of the overlapping part in Fig. 1 (determined by α and β) and the number of links therein [13]. Clearly, the strong network coupling enables the information propagation in one network more likely to trigger further propagation in the other network and hence facilitates the diffusion process. To get a more concrete sense, we study the following two cases in model II.

In the first case, we change the network coupling strength by choosing different α, i.e., the fractional size of the nodes in the physical network which are also online users, while fixing the total number of individuals. Specifically, the size of the physical network is fixed at $N = 12000$. We select the clique size distribution in scenario 2 in Table 1 with $\lambda = 2$. It is also assumed that the social network has a fixed power-law degree distribution (21) with $\gamma = 3$ and $\Gamma = 10$. In Fig. 5, we plot the influenced nodes fraction versus α. Meanwhile, we let $\beta = \frac{\alpha}{\alpha+1}$ so that the number of online users outside the physical network and the total number of individuals are fixed at 12000 and 24000, respectively. The curves in Fig. 5 show that as the coupling strength between two networks increases with the growing α, the influenced nodes fraction could increase monotonically.

In the second case, we change both the network coupling strength and the total number of individuals by increasing the size of the social network while fixing its degree distribution. Specifically, we fix the size of the physical network at $N = 12000$ and let $\alpha = 0.2$. It follows that the number of online users who are also in the physical network is fixed at $|\mathscr{N}_{F_1}| = \alpha N$. We select the clique size distribution in scenario 2 in Table 1 with $\lambda = 2$. It is also assumed that the social network has a fixed power-law degree distribution. We increase the size of the social network, i.e., $|\mathscr{N}_F| = \alpha N / \beta$ by decreasing β from 1 to 0. At the same time, the number of online users outside the physical network and the total number of individuals $|\mathscr{N}_W \cup \mathscr{N}_{F_2}|$ would increase as well. In what follows, we evaluate the size of the information epidemic in terms of the number of the influenced nodes in $\mathscr{N}_W \cup \mathscr{N}_{F_2}$.

Note that in a single social network, the influenced nodes fraction depends on the degree distribution [7,8]. Therefore the number of the influenced nodes would increase monotonically with the growing network size while fixing its degree distribution. On the contrary, it is more intricate in a coupled social-physical network. One key observation is that the growing size of the social network with the fixed degree distribution could yield two opposite effects on information diffusion. Clearly, the information could spread to more recipients as the network size grows. On the other hand, since the degree distribution is fixed and online users randomly connect to each other [8], there would be fewer links in the overlapping part in Fig. 1 (online users who are also in the physical network) as the non-overlapping fraction in the social network increases (online users outside the physical network). This essentially amounts to reducing the coupling strength between the two networks. Simply put, the growing size of the social network increases the number of total individuals on one hand, but on the other hand reduces the network coupling strength which makes the information

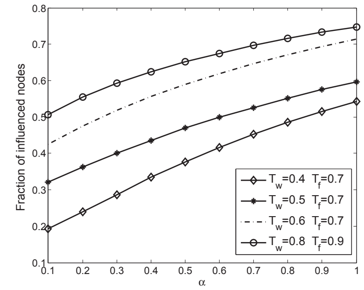

Fig. 5 The influenced nodes fraction in $\mathscr{N}_W \cup \mathscr{N}_{F_2}$ versus α. The curves stand for analytical results obtained by (20).

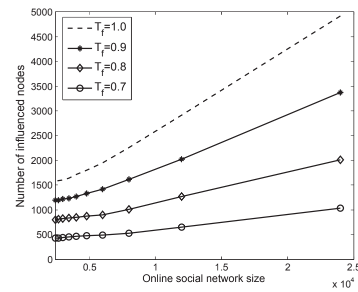

Fig. 6 The number of the influenced nodes in $\mathscr{N}_W \cup \mathscr{N}_{F_2}$ versus the size of social network $|\mathscr{N}_F|$ with $T_w = 0.1$, $\gamma = 2.6$ and $\Gamma = 10$. The size of social network increases from 2.4×10^3 to 2.4×10^4 as β decreases from 1 to 0.1.

more difficult to propagate between two networks. Clearly, the size of information epidemic may either increase or decrease depending on which effect (increase in total number of individuals or reduction in the network coupling strength) is dominant. In what follows, we study the overall impact of these two conflicting effects on a case-by-case basis.

As illustrated in Figs. 6-8, as the size of the social network grows, the number of influenced nodes exhibits different behaviors under different values of transmissibility T_w, i.e., the probability the message can successfully pass through a type-1 (face-to-face) link. For the case with low and high T_w ($T_w = 0.1$ and $T_w = 0.8$), the number of the influenced nodes increases with the growing size of the social network. In contrast, for the case with median $T_w = 0.3$, the number of the influenced nodes decreases as the social network size grows.

We believe that such diverse behaviors can be attributed to the following reasons. For low transmissibility ($T_w = 0.1$ in Fig. 6), the information is difficult to spread through type-1 link, indicating that the propagation in the social network is less likely to trigger further propagation in the physical network. While for high transmissibility ($T_w = 0.8$ in Fig. 8), most of the nodes in the physical network could already receive the information through type-1 link and hence the online contacts are not necessary. Therefore, in both cases, the network coupling does not contribute much in facilitating

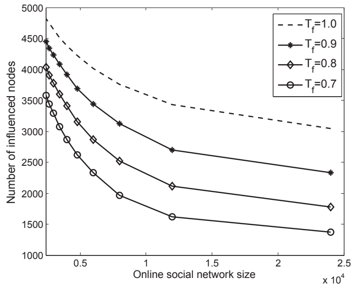

Fig. 7 The number of the influenced nodes in $\mathscr{N}_W \cup \mathscr{N}_{F_2}$ versus the size of social network $|\mathscr{N}_F|$ with $T_w = 0.3$, $\gamma = 3$ and $\Gamma = 10$. The size of social network increases from 2.4×10^3 to 2.4×10^4 as β decreases from 1 to 0.1.

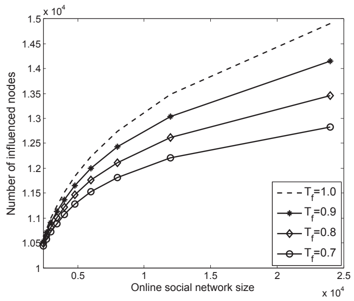

Fig. 8 The number of the influenced nodes in $\mathscr{N}_W \cup \mathscr{N}_{F_2}$ versus the size of social network $|\mathscr{N}_F|$ with $T_w = 0.8$, $\gamma = 3$ and $\Gamma = 10$. The size of social network increases from 2.4×10^3 to 2.4×10^4 as β decreases from 1 to 0.1.

the information diffusion. In other words, the impact of reducing the coupling strength would be trivial. As the size of the social network grows, the increase in the total number of individuals becomes the dominant impact which makes the total number of the influenced nodes keep growing up. On the contrary, for the median transmissibility $T_w = 0.3$, only a limited fraction of nodes can receive the information purely through type-1 link (in contrast to the case with $T_w = 0.8$), indicating that in this case the network coupling would have a great potential to enhance the information diffusion by triggering the propagation between two networks. In other words, the reduction in the coupling strength could result in a substantial negative impact on information diffusion and makes the total number of the influenced nodes decrease as illustrated in Fig. 7.

7 Conclusion

In this study, we explored the diffusion of real-time information in a coupled social-physical networks. We developed a model that consists of an online social network and a physical information network with clique structure. One interesting finding is that a larger size online social network, with the same degree distribution, may not necessarily yield a larger size of information epidemic. In fact, under certain conditions, the size of information epidemic could even decrease with the growing size of the online social network. This is in stark contrast to that in a single network. We believe that our studies will offer initial steps towards understanding the diffusion behaviors of real-time information.

8 Appendix

8.1 Proofs of Lemma 4.1 and Lemma 4.2

In [11, 12] Söderberg studied the phase transition behaviors of inhomogeneous random graphs where nodes are connected by different types of links. Such graphs are also called "colored degree-driven random graphs" in the sense that different types of links correspond to different colors. In a graph with r-types of links, the link degree of an arbitrary node can be represented by a r-dimension vector $d = [d^1 \ \cdots \ d^r]$, where d^j stands for the number of type-j links incident on that node. In our study, the abstract mapping graph $\mathbb{E}$ has two types of links and the degree distribution of an arbitrary node is denoted by $p(i,j) = \mathrm{P}[d_w = i, d_f = j]$. Also, the generating function of degree distribution can be defined by $H(x_1,x_2) = \sum_i \sum_j p(i,j) x_1^i x_2^j$. Clearly, the multivariable combinatorial moments can be achieved by partial differentiation at $x_1 = 1$ and $x_2 = 1$, i.e.,

$$\mathrm{E}[d_w] = \partial_1 H(x_1,x_2)|_{x_1=x_2=1},$$
$$\mathrm{E}[d_f] = \partial_2 H(x_1,x_2)|_{x_1=x_2=1},$$
$$\mathrm{E}[d_w d_f] = \partial_1\partial_2 H(x_1,x_2)|_{x_1=x_2=1},$$
$$\mathrm{E}[(d_w)^2] = \partial_1^2 H(x_1,x_2)|_{x_1=x_2=1},$$
$$\mathrm{E}[(d_f)^2] = \partial_2^2 H(x_1,x_2)|_{x_1=x_2=1}.$$

Let $\{a_k\}$ denote the size distribution of the largest connected component that can be reached from an arbitrary node in $\mathbb{E}$, whose generating function is defined by $g(z) = \sum_k a_k z^k$. Furthermore, we define a two-dimension vector $\mathbf{h}(z) = [h_1(z), h_2(z)]$, where $h_i(z)$ stands for the generating function of the size distribution of the component connected by type-i links. According to the existing results in [3, 11, 12], we have that

$$g(z) = z \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} p(i,j) h_1(z)^i h_2(z)^j = zH(\mathbf{h}(z)), \tag{22}$$

where $\mathbf{h}(z)$ satisfies the following recursive equations:

$$h_1(z) = \frac{z}{\mathrm{E}[d_w]} \partial_1 H(h(z)), \tag{23}$$

$$h_2(z) = \frac{z}{\mathrm{E}[d_f]} \partial_2 H(h(z)). \tag{24}$$

The emergence of the giant component in $\mathbb{E}$ can be checked by examining the stability of the recursive equations (23)-(24) at the point $h_1 = h_1(1) = 1$ and $h_2 = h_2(1) = 1$. Along the same line as in [11,12], we define a 2×2 Jacobian matrix $\mathbf{J}$, i.e.,

$$\mathbf{J}=\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix},$$

where

$$a_{11} = \frac{1}{\mathrm{E}[d_w]}\partial_1^2 H(h(z))|_{h_1=h_2=1} = \mathrm{E}[(d_w)^2 - d_w]\Big/\mathrm{E}[d_w],$$
$$a_{12} = \frac{1}{\mathrm{E}[d_w]}\partial_1\partial_2 H(h(z))|_{h_1=h_2=1} = \mathrm{E}[d_w d_f]/\mathrm{E}[d_w],$$
$$a_{21} = \frac{1}{\mathrm{E}[d_f]}\partial_1\partial_2 H(h(z))|_{h_1=h_2=1} = \mathrm{E}[d_w d_f]/\mathrm{E}[d_f],$$
$$a_{22} = \frac{1}{\mathrm{E}[d_f]}\partial_2^2 H(h(z))|_{h_1=h_2=1} = \mathrm{E}[(d_f)^2 - d_f]\Big/\mathrm{E}[d_f].$$

The spectral radius of $\mathbf{J}$ is given by

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

By [3,10,12], if $\sigma > 1$, with high probability there exist a giant component in the graph $\mathbb{E}$; otherwise, a giant component is very less likely to exist in $\mathbb{E}$. Therefore, the condition (3) in Lemma 4.1 is achieved. Furthermore, the fractional size S_c equals $1-g(1)$ [3]. By (22), we have that

$$\begin{aligned} S_c &= 1-g(1) = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} p(i,j)\left(1-h_1^i h_2^j\right) \\ &= \sum_{n=1}^{D}\sum_{m=0}^{n}\mu_{nm}\left(1-\mathrm{E}[h_1^{K_n^w}]\mathrm{E}[h_2^{K_m^f}]\right). \end{aligned} \tag{25}$$

In view of (23) and (24), we have that

$$\begin{aligned} h_1 &= \frac{1}{\mathrm{E}[d_w]}\sum_{i=0}^{\infty}\sum_{j=0}^{\infty} p(i,j)\, i h_1^{i-1} h_2^j \\ &= \frac{1}{\mathrm{E}[d_w]}\sum_{n=1}^{D}\sum_{m=0}^{n}\mu_{nm}\mathrm{E}[K_n^w h_1^{K_n^w-1}]\mathrm{E}[h_2^{K_m^f}], \end{aligned}$$

$$\begin{aligned} h_2 &= \frac{1}{\mathrm{E}[d_f]}\sum_{i=0}^{\infty}\sum_{j=0}^{\infty} p(i,j)\, j h_1^{i} h_2^{j-1} \\ &= \frac{1}{\mathrm{E}[d_f]}\sum_{n=1}^{D}\sum_{m=0}^{n}\mu_{nm}\mathrm{E}[h_1^{K_n^w}]\mathrm{E}[K_m^f h_2^{K_m^f-1}]. \end{aligned}$$

Furthermore, (25) can be rewritten in the following form:

$$S_c = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\sum_{n=1}^{D}\sum_{m=0}^{n}\mu_{nm}P_{ni}^{w}P_{mj}^{f}(1-h_1^i h_2^j).$$

Clearly, the term in parentheses gives the probability that a node with colored degree $[d_w = i, d_f = j]$ belongs to the giant component. In other words, the term in parentheses is the expected number of cliques added to the giant cluster by a degree $[d_w = i, d_f = j]$ clique. Hence, summing over all such i, j's we get an expression for the *expected* size of the giant component (in terms of the number of cliques).

In order to compute S_n, we can modify the above expression such that the term $n(1-h_1^i h_2^j)$ gives the expected number of nodes to be included in the giant component by a degree $[d_w = i, d_f = j]$ clique. In other words, with probability $(1-h_1^i h_2^j)$ the clique under consideration will belong to the giant component G_H and will bring n nodes to the actual size S_n. This yields

$$\begin{aligned} \bar{S}_n &= \sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\sum_{n=1}^{D}\sum_{m=0}^{n}\mu_{nm}P_{ni}^{w}P_{mj}^{f} n\left(1-h_1^i h_2^j\right) \\ &= \sum_{n=1}^{D}\sum_{m=0}^{n} n\mu_{nm}\left(1-\mathrm{E}[h_1^{K_n^w}]\mathrm{E}[h_2^{K_m^f}]\right). \end{aligned}$$

We next have

$$S_n = \frac{1}{C}\bar{S}_n, \quad C = \sum_{n=1}^{D} n\mu_n,$$

where the normalized term C makes $S_n = 1$ at $h_1 = h_2 = 0$. Therefore, the conclusions (6) and (7) in Lemma 4.2 have been obtained.

References

1. D. Qian, O. Yagan, L. Yang, and J. Zhang. Diffusion of real-time information in social-physical networks. *IEEE GLOBECOM*, 2012.
2. K. Zhao, J. Stehle, G. Bianconi, and A. Barrat. Social network dynamics of face-to-face interactions. *Phys. Rev. E*, 83(5):056109, 2011.
3. O. Yağan, D. Qian, J. Zhang, and D. Cochran. Conjoining speeds up information diffusion in overlaying social-physical networks. *IEEE Journal on Selected Areas in Communications*, 31(6):1038-1048, 2013.
4. O. Yağan and V. Gligor. Analysis of complex contagions in random multiplex networks. *Phys. Rev. E*, 86:036103, 2012.
5. J. Yang and J. Leskovec. Modeling information diffusion in implicit networks. In *Proceedings of the 10th IEEE International Conference on Data Mining*, 2010.
6. J. Stehlé, A. Barrat, and G. Bianconi. Dynamical and bursty interactions in social networks. *Phys. Rev. E*, 81(3):035101, 2010.
7. 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.
8. M. E. J. Newman. Spread of epidemic disease on networks. *Phys. Rev. E*, 66(1), 2002.
9. M.E.J. Newman and M. Girvan. Finding and evaluating community structure in networks. *Physical review E*, 69(2):026113, 2004.
10. B. Söderberg. General formalism for inhomogeneous random graphs. *Phys. Rev. E*, 66:066121, 2002.
11. B. Söderberg. Random graphs with hidden color. *Phys. Rev. E*, 68:015102(R), 2003.
12. B. Söderberg. Random graph models with hidden color. *Acta Phys. Pol. B*, 34:5085–5102, 2003.
13. R. Parshani, S.V. Buldyrev, and S. Havlin. Interdependent networks: Reducing the coupling strength leads to a change from a first to second order percolation transition. *Physical review letters*, 105(48701), 2010.