Bond Percolation in Clustered Multilayer Networks

Yong Zhuang and Osman Yağan
Department of ECE and CyLab, Carnegie Mellon University, Pittsburgh, PA 15213 USA
{yongzhua, oyagan}@andrew.cmu.edu

***Abstract*—In today's world, individuals interact with each other in more complicated patterns than ever. Some individuals engage through online social networks (e.g., Facebook, Twitter), while some communicate only through conventional ways (e.g., face-to-face). Therefore, understanding the dynamics of information propagation calls for a multi-layer network model where an online social network is conjoined with a physical network. Here, we study information diffusion in a *clustered* multi-layer network model, where all constituent layers are random networks with high clustering. We assume that information propagates according to the SIR model and with different information transmissibility across the networks. Taking advantage of the isomorphism between bond percolation and information propagation processes, we give results for the conditions, probability, and size of information epidemics. We show that increasing the level of clustering in either one of the layers increases the epidemic threshold and decreases the final epidemic size.**

I. Introduction

The study of dynamical processes on real-world complex networks has been an active research area over the past decade. An interesting phenomenon that occurs in many such processes is the spreading of an initially localized effect throughout the whole (or, a very large part of the) network. These events are usually referred to as (information) *cascades* and can be observed in processes as diverse as the adoption of cultural fads, the diffusion of norms in social networks [4], [18], disease contagion in human and animal populations [12], [23], and the spread of failures in *interdependent* power systems [3], [22].

This work focuses on an important class of dynamical process known as the information propagation. Although well-studied in the past, the information diffusion problem has recently taken a new form by the emergence of online social networks (e.g., Facebook). In particular, due to the existence of multiple online social networks, information is now likely to spread in an unprecedented speed and scale. Although there has been a recent surge of research on multi-layer and multiplex networks (e.g., see [8], [23]), the current literature still falls short in fully quantifying this phenomenon. For instance, Yağan et al. analyzed [20], [23] the diffusion of information in a multi-layer network, but only for the cases where all constituent layers are generated by the *configuration model* [14]. However, the configuration model produces [10], [15] networks that can not accurately capture some important aspects of real-world social networks, most notably the property of *high clustering* [17], [19]. Informally known as the phenomenon that "friends of our friends" are likely to be our friends, clustering has been shown to impact significantly the dynamics of various diffusion processes [6], [16], [21].

With these in mind, we study information propagation in *clustered* multi-layer networks. Our framework consists of a physical network where information spreads amongst people through *conventional* communication media (e.g., face-to-face communication), and *overlaying* this network, there are online social networks as alternative ways for the spread (e.g., Facebook). The coupling across these networks results from nodes they have in common, i.e., individuals who participate in multiple networks simultaneously.

In this setting, we analyze the propagation of information assuming that information propagates according to the SIR epidemic model; the analogy between the spread of diseases and information has long been recognized [5]. Namely, an individual is either *susceptible* (S) meaning that she has not yet received a particular information, or *infectious* (I) meaning that she is aware of the information and is spreading it to her contacts, or *recovered* (R) meaning that she is no longer spreading the information. Let T_{ij} denote the probability that an *infectious* individual i transmits the information to a *susceptible* contact j. Throughout, we account for the fact that individuals' information spreading behaviors may differ from one network to another; e.g., one may be more active in Facebook than Twitter, or vice versa. The varying rate of information diffusion across different social networks is captured in our formulation by having the transmissibility T_{ij} depend on the network that the link $i \sim j$ belongs to.

Our main contributions are as follows. We solve analytically for the threshold, probability, and mean size of *information epidemics*, i.e., cases where information starts from a single individual and reaches a positive fraction of the population; see Section II-C for precise definitions. Our analytical approach is based on mapping the SIR propagation model to a *bond percolation* process and then utilizing a multi-type branching process to solve for the quantities of interest; the isomorphism between the SIR model and bond percolation has been established for certain cases in [7], [9]. The analytical results are validated via computer simulations. Several interesting conclusions are drawn from these results including i) clustering makes it more difficult for a single person to spread the information to the *masses*; and ii) even if the information reaches to the masses, we show that clustering decreases the total fraction of individuals informed. Due to limited space, most details are skipped here. They are given in [24].

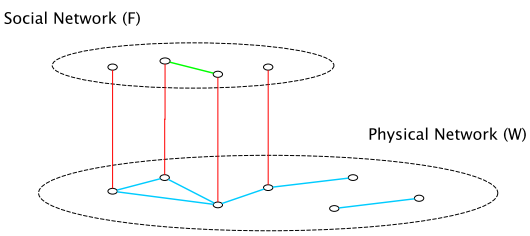


Fig. 1. Nodes in the upper circle and lower circle indicate the individuals in *social network* ($\mathbb{F}$) and *physical network* ($\mathbb{W}$) respectively. Some of the nodes connected across the two networks by a red line indicates the fact that they represent the same individual.

II. Problem Formulation

A. Random Graphs with Clustering

Our modeling framework is based on *random* networks with *clustering* as introduced independently by Miller [11] and Newman [13]. This model considers a vertex set $V = 1, 2, \dots, n$, where each vertex is independently assigned a random number of *stubs* according to a joint degree distribution $\{p_{st}\}_{s,t=0}^{\infty}$ that gives the probability that a node has s single edges and t triangles. Namely, each node will be given s stubs labeled as *single* and $2t$ stubs labeled as *triangles* with probability p_{st}, for any $s, t = 0, 1, 2, \dots$. Then, stubs that are labeled as single are *randomly* joined to form single edges that are not part of a triangle, whereas *pairs* of triangle stubs from three nodes are *randomly* matched to form triangles between the three participating nodes; of course the total degree of a node will be distributed by $p_k = \sum_{s,t:s+2t=k} p_{st}$.

The resulting level of clustering of the model described above can be quantified in a number of ways. Here we consider two widely used metrics known as the *global* clustering coefficient [13] and *local* clustering coefficient [1].

B. Multilayer Network Models with Clustering

In this paper, we consider a multilayer network where each layer is generated independently and constitutes a random graph with clustering as introduced in Section II-A. For brevity, we only consider two layers but most of the arguments can easily be extended higher number of layers. Namely, we let $\mathbb{W}$ and $\mathbb{F}$ denote the two constituent layers of networks with the possible motivation that $\mathbb{W}$ models the *physical* contact network among individuals, i.e., models face-to-face relationships, while network $\mathbb{F}$ stands for an online social network, say Facebook. In line with this terminology, we assume that the network $\mathbb{W}$ is defined on the vertices $\mathcal{N} = \{1, \dots, n\}$, while $\mathbb{F}$ contains only a *subset* of the nodes in $\mathcal{N}$ to account for the fact that not every individual participates in online social networks; see Figure 1 for an illustration of the two-layer network model we are considering. To specify this model further, we assume that each vertex in $\mathcal{N}$ participates in $\mathbb{F}$ independently with probability $\alpha \in (0, 1]$. This implies that the fraction of nodes that belong to $\mathbb{F}$ is α in the large n limit.

We assume that both $\mathbb{F}$ and $\mathbb{W}$ are random networks with clustering. In particular, we let $\{p_{st}^f, s, t = 0, 1, \dots\}$ and $\{p_{st}^w, s, t = 0, 1, \dots\}$ denote the joint distributions for single edges and triangles for $\mathbb{F}$ and $\mathbb{W}$, respectively. Then both networks are generated independently according to the algorithm described in Section II-A, and they are denoted respectively by $\mathbb{F} = \mathbb{F}(n; \alpha, p_{st}^f)$ and $\mathbb{W} = \mathbb{W}(n; p_{st}^w)$. We define the multi-layer network $\mathbb{H}$ as the *disjoint* union $\mathbb{H} = \mathbb{F} \coprod \mathbb{W}$ and represent it by $\mathbb{H}(n; \alpha, p_{st}^f, p_{st}^w)$. Here, the disjoint union operation implies that we still distinguish $\mathbb{F}$-edges from $\mathbb{W}$-edges in network $\mathbb{H}$, and this is done to accommodate the possibly different rates (or, even rules) of information propagation across the two networks.

With these definitions in mind, let d_{fs} and d_{ws} to denote the random variables corresponding to the number of *single edges* for a vertex in $\mathbb{F}$ and $\mathbb{W}$, respectively, while n_{ft} and n_{wt} are defined similarly for the number of triangles assigned. Then the *colored* degree $\mathbf{d}$ of a vertex is given by

$$\mathbf{d} = (d_{fs}, 2n_{ft}, d_{ws}, 2n_{wt}) \tag{1}$$

meaning that the vertex has d_{fs} *single edges* and $2n_{ft}$ *triangle edges* in network $\mathbb{F}$, and d_{ws} *single edges* and $2n_{wt}$ *triangle edges* in network $\mathbb{W}$. Then the distribution of this colored degree is given by

$$p_{\mathbf{d}} = \left(\alpha p^f_{d_{fs} n_{ft}} + (1-\alpha)\mathbf{1}[d_{fs} = 0 \ \wedge \ n_{ft} = 0]\right) p^w_{d_{ws} n_{wt}} \tag{2}$$

where the term $(1-\alpha)\mathbf{1}[d_{fs} = 0 \ \wedge \ n_{ft} = 0]$ accounts for the fact that if the node does not belong to $\mathbb{F}$ (which happens with probability $1-\alpha$), then its degree from single and triangle edges will both be zero.

C. Problems of Interest

We consider the propagation of information in $\mathbb{H}$ according to SIR model. As we explain in details in [24], the SIR propagation model is *isomorphic* to a bond percolation process [2] for the purposes of computing the threshold, probability, and expected size of epidemics. Then it can be assumed that information propagates over $\mathbb{W}$ (resp. over $\mathbb{F}$) as if all transmission probabilities were equal to T_w (resp. to T_f). The outbreak is triggered by infecting a randomly selected node and propagates in the network according to the SIR model. The final size of an outbreak is defined as the number of nodes that are *recovered* at the steady-state, and its relative final size is its final size divided by the total size n of the network. Following [7], we define a *self-limited outbreak* as an outbreak whose relative final size approaches zero, and an *epidemic* to be an outbreak whose relative final size is positive, both in the limit of large n. There is a *critical boundary* in the space of all network parameters, often defined as the epidemic *threshold*, or epidemic boundary, that separates the cases for which the probability of an epidemic is zero from those that lead to $\mathbb{P}[\text{epidemic}] > 0$ (i.e., *super-critical*) with $n \to \infty$.

With these definitions in place, this work seeks to identify i) the epidemic boundary; ii) the relative final size of epidemics in the super-critical case; and iii) the exact probability $\mathbb{P}[\text{epidemic}]$ in the super-critical regime.

III. Main Results

As described in Section II-B, the clustered multilayer network in this paper consists of four kinds of edges, *single edges* in $\mathbb{F}$, *triangle edges* in $\mathbb{F}$, *single edges* in $\mathbb{W}$, and *triangle edges* in $\mathbb{W}$; these will be denoted by $fs-$, $ft-$, $ws-$, and $wt-$edges, respectively.

We now explain our approach based on generating functions precisely. Let $H(x)$ denote the generating function for the "*finite* number of nodes that are reached and informed" in the propagation process. We will derive an expression for $H(x)$ using four other generating functions, $h_{fs}(x)$, $h_{ft}(x)$, $h_{ws}(x)$, and $h_{wt}(x)$, where $h_{fs}(x)$ stands for the "finite number of nodes reached and informed by following a randomly selected fs-edge," and $h_{ws}(x)$ defined similarly for the $ws-$edges. The definitions for $h_{ft}(x)$ and $h_{wt}(x)$ are a bit different in the sense that they correspond to the "finite number of nodes reached and informed by following a randomly selected *triangle* in $\mathbb{F}$ (resp. in $\mathbb{W}$)" for $h_{ft}(x)$ (resp. $h_{wt}(x)$). In other words, we consider the whole triangle at once, rather than focusing on its edges separately.

With these definitions in place, we now write $H(x)$ as

$$H(x) = x \sum_{\mathbf{d}} p_{\mathbf{d}} h_{fs}(x)^{d_{fs}} h_{ft}(x)^{n_{ft}} h_{ws}(x)^{d_{ws}} h_{wt}(x)^{n_{wt}} \quad (3)$$

where $p_{\mathbf{d}}$ denotes the colored degree distribution given by (2). The validity of (3) can be seen as follows. The term x stands for the node that is selected randomly and given the information to initiate the propagation. This node has a degree $\mathbf{d}$ with probability $p_{\mathbf{d}}$. The number of nodes reached and informed by each of its d_{fs} (resp. d_{ws}) single edges in $\mathbb{F}$ (resp. $\mathbb{W}$) has a generating function $h_{fs}(x)$ (resp. $h_{ws}(x)$). Similarly, the number of nodes informed by following each of the n_{ft} (resp. n_{wt}) triangles it participates in $\mathbb{F}$ (resp. $\mathbb{W}$) has a generating function $h_{ft}(x)$ (resp. $h_{wt}(x)$). Combining, we see from the *powers property* of generating functions [14] that the number of nodes reached and informed in this process when the initial node has degree $\mathbf{d}$ has a generating function $h_{fs}(x)^{d_{fs}} h_{ft}(x)^{n_{ft}} h_{ws}(x)^{d_{ws}} h_{wt}(x)^{n_{wt}}$. Averaging over all possible degrees $\mathbf{d}$ of the initial node, we get (3).

For (3) to be useful, we shall derive expressions for the generating functions $h_{fs}(x)$, $h_{ft}(x)$, $h_{ws}(x)$, $h_{wt}(x)$, the epidemic threshold, and final epidemic size in following sections.

A. Information Propagation via Single Edges in Network $\mathbb{F}$

We start by deriving recursive equations for $h_{fs}(x)$ and $h_{ws}(x)$. For $h_{fs}(x)$, we pick one of the single edges in $\mathbb{F}$ uniformly at random and assume that it is connected at one end a node who is in the *infected* state. Then, we compute the generating function for the number of nodes informed by following the other end of the edge. In what follows, we only derive $h_{fs}(x)$ because of the similar manner of $h_{ws}(x)$.

Similar to [23], we obtain the following expression for the generating function $h_{fs}(x)$:

$$h_{fs}(x) = (1 - T_f) + \quad (4)$$

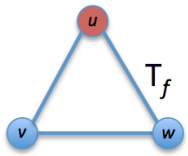


Fig. 2. The top vertex u is infected, and information is transferred through an edge only if it is occupied (happens with probability T_f in network $\mathbb{F}$).

$$T_f x \sum_{\mathbf{d}} \frac{d_{fs} p_{\mathbf{d}}}{\langle d_{fs} \rangle} h_{fs}(x)^{d_{fs}-1} h_{ft}(x)^{n_{ft}} h_{ws}(x)^{d_{ws}} h_{wt}(x)^{n_{wt}}.$$

We now explain each term appearing at (4) in turn. First of all, the term $(1 - T_f)x^0$ to $h_{fs}(x)$ means that the probability of the underlying random variable (encoded by the generating function $h_{fs}(x)$) being zero is incremented by $1 - T_f$. On the other hand, if the selected edge is occupied, which happens with probability T_f, then the node at the other end of the edge will be informed. This means that the number of informed edges in this process will be one *plus* all the nodes that are then informed by the node at the other end of the selected edge. Adding one to a random variable is equivalent to multiplying its generating function by x, whence we get the term $T_f x$.

The summation term at (4) stands for the number of nodes informed by the aforementioned *end node* of the randomly selected edge. More specifically, as it is already known to have at least one fs-edge, the probability that a node has degree $\mathbf{d}$ is $\frac{d_{fs} p_{\mathbf{d}}}{\langle d_{fs} \rangle}$. Then, the number of people it will inform is generated by $h_{fs}(x)^{d_{fs}-1} h_{ft}(x)^{n_{ft}} h_{ws}(x)^{d_{ws}} h_{wt}(x)^{d_{wt}}$, with the minus one term on d_{fs} accounting to the fact that one of its single edges in $\mathbb{F}$ has carried the information to this node and has already been taken into account. Averaging over all possible $\mathbf{d}$, we get (4).

B. Information Propagation via Triangles in Network $\mathbb{F}$

We now derive $h_{ft}(x)$, i.e., the generating function for the number of nodes informed by following a random triangle in $\mathbb{F}$; similar arguments hold for $h_{wt}(x)$. We demonstrate this situation in Figure 2, where the top vertex u is *infected*, and we are interested in computing the generating function for the number of nodes that will be informed *by* nodes v *and* w. By conditioning on the state of the three edges forming this triangle, we compute the probabilities for neither, one, or both of v and w being informed, respectively. We have

$$\begin{cases} \mathbb{P}[\text{none of } v \text{ and } w \text{ are informed}] & = (1 - T_f)^2 \\ \mathbb{P}[\text{one of } v \text{ and } w \text{ are informed}] & = 2T_f(1 - T_f)^2 \\ \mathbb{P}[\text{both of } v \text{ and } w \text{ are informed}] & = 2T_f^2(1 - T_f) + T_f^2. \end{cases}$$

Then, the generating function $h_{ft}(x)$ can be obtained in the same vein as $h_{fs}(x)$ by conditioning on the three events discussed above. We get

$$\begin{aligned} &h_{ft}(x) \quad (5) \\ &= (1 - T_f)^2 + \left(2T_f (1 - T_f)^2\right) x \sum_{\mathbf{d}} \left(\frac{n_{ft} p_{\mathbf{d}}}{\langle n_{ft} \rangle} h_{fs}(x)^{d_{fs}} \right. \\ &\left. \cdot\, h_{ft}(x)^{n_{ft}-1} h_{ws}(x)^{d_{ws}} h_{wt}(x)^{n_{wt}} \right) + \left(2T_f^2(1 - T_f) + T_f^2\right) \end{aligned}$$

$$\cdot \left(x \sum_{\mathbf{d}} \frac{n_{ft} p_{\mathbf{d}}}{\langle n_{ft} \rangle} h_{fs}(x)^{d_{fs}} h_{ft}(x)^{n_{ft}-1} h_{ws}(x)^{d_{ws}} h_{wt}(x)^{n_{wt}} \right)^2 .$$

C. Computing the Final Epidemic Size

To compute the final epidemic size, we take advantage of the "conservation of probability" property of generating functions, i.e., the fact that $H(1) = 1$ when the number of nodes reached and informed is always *finite*. If on the other hand $H(1) < 1$, we understand that there is a positive probability $1 - H(1)$ to lead to an *infinite* component of informed nodes. In this case, $1 - H(1)$ stands for the relative final size of epidemics.

With these in mind, the threshold and size of information epidemics can be obtained by seeking a fixed point of the derived recursive equations $h_{fs}(x)$, $h_{ft}(x)$, $h_{ws}(x)$, and $h_{wt}(x)$ at the point $x = 1$. For convenience, define $h_1 := h_{fs}(1)$, $h_2 := h_{ft}(1)$, $h_3 := h_{ws}(1)$, and $h_4 := h_{wt}(1)$. Then, for a given fixed point solution h_1, h_2, h_3, h_4, we have

$$H(1) = \sum_{\mathbf{d}} p_{\mathbf{d}} h_1^{d_{fs}} h_2^{n_{ft}} h_3^{d_{ws}} h_4^{n_{wt}}. \tag{6}$$

It is clear that the recursions, $h_{fs}(x)$, $h_{ft}(x)$, $h_{ws}(x)$, and $h_{wt}(x)$, exhibit a trivial fixed point $h_1 = h_2 = h_3 = h_4 = 1$, which leads to $H(1) = 1$, meaning that all informed components have *finite* size. However, the solution $h_1 = h_2 = h_3 = h_4 = 1$ is physical only when it is a *stable* fixed point. We check the stability of this solution via *linearization* of the four recursions around $x = 1$. Namely, the solution will be stable if the largest eigenvalue in absolute value of the corresponding Jacobian matrix $\mathbf{J}$ is larger than one. Otherwise, another fixed point with $h_1, h_2, h_3, h_4 < 1$ will be the stable solution leading to $H(1) < 1$. The corresponding deficit $1 - H(1)$ is then attributed to the fact that an *infinite* informed component exists. Collecting, we have i) the threshold of information epidemics is given by $\sigma(\mathbf{J}) = 1$, where $\sigma(\mathbf{J})$ is the *spectral radius* of the Jacobian matrix; and ii) the mean epidemic size can be computed by first finding the pointwise smallest solution of the four recursions, and then reporting the result into (6) to get the mean size of epidemics, $1 - H(1)$; see [24] for more details.

IV. Numerical Results and Discussion

A. Networks with Doubly Poisson Distributions

Consider the case where both p^f_{st} and p^w_{st} are doubly Poisson; i.e., the number of *single edges* and *triangles* in both networks are independent and they all follow a Poisson distribution. Namely, we set

$$p^f_{st} = e^{-\mu_{f,1}} \frac{(\mu_{f,1})^s}{s!} e^{-\mu_{f,2}} \frac{(\mu_{f,2})^t}{t!}, \quad \text{s, t = 1, 2, \ldots,} \tag{7}$$

and

$$p^w_{st} = e^{-\mu_{w,1}} \frac{(\mu_{w,1})^s}{s!} e^{-\mu_{w,2}} \frac{(\mu_{w,2})^t}{t!}, \quad \text{s, t = 1, 2, \ldots,} \tag{8}$$

where s and t are the number of *single edges* and *triangles* in the corresponding networks while $\mu_{f,1}$ and $\mu_{f,2}$ (resp. $\mu_{w,1}$ and $\mu_{w,2}$) are the mean number of them respectively in $\mathbb{F}$ (resp. in $\mathbb{W}$).

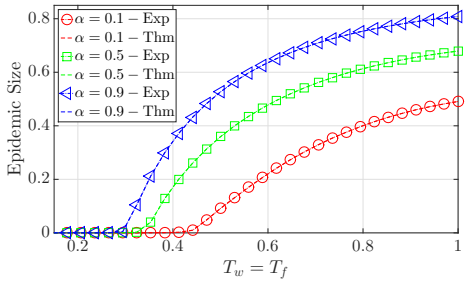


Fig. 3. Simulation for doubly Poisson degree distributions.

Under this setting, the mean epidemic size as well as the epidemic threshold can be computed from the analytical results presented in Section III-C. To check the validity of our analysis for finite-sized networks, we have also conducted an extensive numerical study. In particular, we consider $n = 5 \times 10^5$ nodes in the population and three different values $\alpha = 0.1, 0.5, 0.9$ for the size of network $\mathbb{F}$. We let $\mu_{f,1} = \mu_{f,2} = \lambda_{fs} = \lambda_{ft} = 0.5$ and similarly $\mu_{w,1} = \mu_{w,2} = \lambda_{ws} = \lambda_{wt} = 0.5$. For various information transmissibility parameters $T_w = T_f$ we generate 100 independent realizations of the multi-layer network $\mathbb{H}$ and compute the size of the largest connected component in each case. The results are then averaged over 100 experiments to obtain the *empirical* size of epidemics.

The results are depicted in Figure 3, where the curves stand for the theoretical results in Section III-C, while the markers stand for the empirical results obtained from simulation experiments. We see that there is a perfect agreement between the analytical and experimental results confirming the validity of our results even when n is finite. We also see that as α increases, the critical threshold is reduced and the epidemic size is enlarged. This is an intuitive consequence given that the network becomes *denser* with increasing α.

B. Impact of clustering on information epidemics

An important goal of this work is to understand how clustering affects the dynamics of information propagation in multi-layer networks. We control the level of clustering in network $\mathbb{W}$ while keeping its mean total degree fixed. More precisely, we use Poisson distributions for the number of single edges and triangles in both networks with parameters given in Table I. Put differently, network $\mathbb{F}$ has a fixed clustering coefficient while with $c \in [0, 4]$ the clustering of $\mathbb{W}$ varies between the two extremes: i) when $c = 4$, $\mathbb{W}$ will have no *single*-edges and consist only of *triangles* resulting with a clustering coefficient close to one; and ii) with $c = 0$, there will be no triangles in $\mathbb{W}$ and hence its clustering coefficient will be close to zero. Thus, with increasing c, the clustering coefficient of $\mathbb{W}$ increases, which in turn increases clustering in the multilayer network $\mathbb{H}$; see Table II for specific clustering coefficients corresponding to several c values considered.

In Figure 4(a), for each parameter pair (c, α), the curves separates the region where information epidemics can take place (north-east of the curves) from the region where they can not (south-west of the curves). We see that with the same T_f, clustering increases the minimum T_w needed for information epidemics to be possible. In other words, we see again that

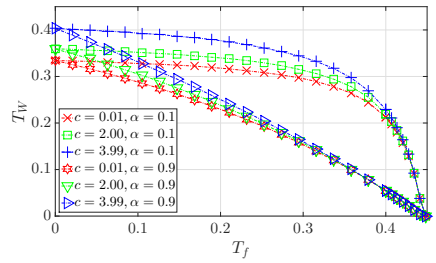
	Network $\mathbb{F}$	Network $\mathbb{W}$
Distribution of *single*-edges	Poi($2\lambda_{\mathbb{F}}$)	2 Poi($\frac{4-c}{2}\lambda_{\mathbb{W}}$)
Distribution of *triangles*	Poi($\lambda_{\mathbb{F}}$)	Poi($\frac{c}{2}\lambda_{\mathbb{W}}$)

TABLE I
PARAMETERS OF THE DOUBLY POISSON DISTRIBUTION. FIGURE 4(A) IS WITH $\lambda_{\mathbb{F}} = \lambda_{\mathbb{W}} = 0.5$, FIGURE 4(B) IS WITH $\lambda_{\mathbb{F}} = 0.36$ AND $\lambda_{\mathbb{W}} = 0.5$.

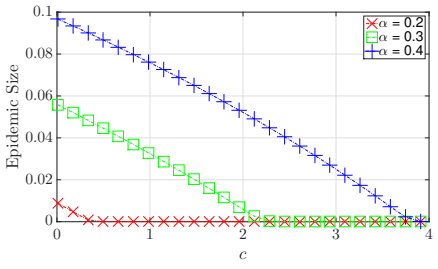
α	Clust. Coef.	$c = 0.01$	$c = 2.00$	$c = 3.99$
0.1	Global	0.005	0.095	0.185
	Local	0.006	0.230	0.453
0.9	Global	0.023	0.075	0.126
	Local	0.044	0.152	0.260

TABLE II
STATISTICS OF THE NETWORK $\mathbb{H}$ IN FIGURE 4(A).

clustering increases the threshold of epidemics. Next, we look at the effect of clustering on the relative final size of information epidemics for specific percolation probabilities. From Figure 4(b), we see that the epidemic size decreases as the clustering coefficient increases, again confirming that high clustering reduces the epidemic size.



(a)



(b)

Fig. 4. a) Comparison of the epidemic *boundary* under several cases; the north and east of each curve specifies the region of (T_f, T_w) values for which epidemics are possible, while the south and west part of each curve stands for the region where epidemics can *not* take place. b) Illustration of how clustering affects the size of epidemics when $T_f = T_w = 0.3$.

V. CONCLUSION

We analyze the propagation of information in clustered multilayer networks. We solve analytically for the threshold, probability, and mean size of information epidemics, and confirm our findings via extensive computer simulations. We show that clustering increases the epidemic threshold and decreases the final epidemic size in multi-layer networks. There are many open problems one might consider for future work. For instance, one can consider networks that exhibit clustering not only through triangles, but also through larger cliques. Extending the work to the case of influence propagation (i.e., complex contagions) would also be interesting.

ACKNOWLEDGMENT

This research was supported in part by National Science Foundation through grant CCF #1422165, and in part by the Department of Electrical and Computer Engineering at Carnegie Mellon University.

REFERENCES

[1] A. Barrat and M. Weigt. On the properties of small-world network models. *The European Physical Journal B-Condensed Matter and Complex Systems*, 13(3):547–560, 2000.
[2] S. R. Broadbent and J. M. Hammersley. Percolation processes. *Math. Proc. of the Cambridge Philosophical Society*, 53:629–641, 1957.
[3] C. D. Brummitt, R. M. D'Souza, and E. Leicht. Suppressing cascades of load in interdependent networks. *Proceedings of the National Academy of Sciences*, 109(12):E680–E689, 2012.
[4] P. Dodds and D. J. Watts. Universal behavior in a generalized model of contagion. *Phys. Rev. Lett.*, 92:218701–, 2004.
[5] D. Gruhl, R. Guha, D. Liben-Nowell, and A. Tomkins. Information diffusion through blogspace. In *Proceedings of the 13th international conference on World Wide Web*, pages 491–501. ACM, 2004.
[6] A. Hackett, S. Melnik, and J. P. Gleeson. Cascades on a class of clustered random networks. *Physical Review E*, 83(5):056107, 2011.
[7] E. Kenah and J. Robins. Second look at the spread of epidemics on networks. *Phys. Rev. E*, 76:036113, Sep 2007.
[8] M. Kivelä, A. Arenas, M. Barthelemy, J. P. Gleeson, Y. Moreno, and M. A. Porter. Multilayer networks. *Journal of Complex Networks*, 2(3):203–271, 2014.
[9] J. Miller. Epidemic size and probability in populations with heterogeneous infectivity and susceptibility. *Phys. Rev. E*, 76:010101, Jul 2007.
[10] J. C. Miller. Percolation and epidemics in random clustered networks. *Physical Review E*, 80(2):020901, 2009.
[11] M. Molloy and B. Reed. A critical point for random graphs with a given degree sequence. *Random structures & algorithms*, 6(2-3):161–180, 1995.
[12] J. D. Murray. *Math. Bio.* Springer, New York (NY), 3 edition, 2002.
[13] M. E. Newman. The structure and function of complex networks. *SIAM review*, 45(2):167–256, 2003.
[14] M. E. Newman, S. H. Strogatz, and D. J. Watts. Random graphs with arbitrary degree distributions and their applications. *Physical review E*, 64(2):026118, 2001.
[15] M. E. J. Newman. Random graphs with clustering. *Physical review letters*, 103(5):058701, 2009.
[16] D. Qian, O. Yagan, L. Yang, and J. Zhang. Diffusion of real-time information in social-physical networks. In *Global Communications Conference (GLOBECOM), 2012 IEEE*, pages 2072–2077. IEEE, 2012.
[17] M. A. Serrano and M. Boguñá. Clustering in complex networks. i. general formalism. *Phys. Rev. E*, 74:056114, Nov 2006.
[18] D. J. Watts. A simple model of global cascades on random networks. *Proceedings of the National Academy of Sciences*, 99:5766–5771, 2002.
[19] D. J. Watts and S. H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393(6684):440–442, 1998.
[20] O. Yağan, D. Qian, J. Zhang, and D. Cochran. Information diffusion in overlaying social-physical networks. In *Information Sciences and Systems (CISS), 2012 46th Annual Conference on*, pages 1–6, 2012.
[21] O. Yağan and V. Gligor. Analysis of complex contagions in random multiplex networks. *Physical Review E*, 86(3):036103, 2012.
[22] O. Yağan, D. Qian, J. Zhang, and D. Cochran. Optimal Allocation of Interconnecting Links in Cyber-Physical Systems: Interdependence, Cascading Failures and Robustness. *IEEE Trans. on Parallel and Distributed Systems*, 23(9):1708–1720, 2012.
[23] 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.
[24] Y. Zhuang and O. Yağan. Information propagation in clustered multilayer networks. *arXiv preprint arXiv:1509.03909*, 2015.