Modeling interest-based social networks: Superimposing Erdős–Rényi graphs over random intersection graphs

Jun Zhao
junzhao@alumni.cmu.edu

***Abstract*—There is a recent rise of interest-based social networks (e.g., Pinterest and Goodreads), which connect users by relations based on shared interests. In these networks, links between users manifest from selecting common interests from a pool of available interests. For example, two users may establish a link on Pinterest because of both liking dog photos, or on Goodreads due to reading the same novel. In this paper, we introduce a random graph model to represent an interest-based social network, in consideration of users' shared interests as well as their friend relations. More specifically, the graph model is the result of superimposing an Erdős–Rényi graph (representing friendships) over a uniform random d-intersection graph (representing common interests). We present critical conditions of the model parameters so that the network is connected. Our connectivity results are useful to understand interest-based social networks and particularly beneficial for publish-subscribe services in these networks. The formally-proved results are also confirmed via experiments.**

***Index Terms*—Social network, common interest, friendship, random graph, topology.**

I. Introduction

In typical online social networks, users are linked by symmetric friend relations [1] and can define *circles of friends* (viz., *Google+*) [3]–[5]. We view a user's circle of friends as the group of friends who share a *common interest*. A basic common interest between two friends can be represented by their selection of a number of common objects from a large pool of available objects. For example, two friends may pick the same set of photos to like from *Pinterest*'s pool, the same books to read from *Goodreads*'s pool, the same songs to listen to from *Spotify*'s pool, or the same videos to watch from *Youtube*'s pool, or the same games to play from *Playfire*. Identifying friends with common interests in a social network enables the implementation of large-scale, distributed publish-subscribe services which support dissemination of special-interest messages among the users [6], [7], [12]. Such services allow publisher nodes to post interest-specific news, recommendations, warnings, or announcements to subscriber nodes in a wide variety of applications ranging from online behavioral advertising (e.g., the message may contain an advertisement targeted to a common-interest group) to social science (e.g., the message may contain a survey request or result directed to a special-interest group).

Jun Zhao obtained his PhD from Carnegie Mellon University, Pittsburgh, PA, USA. He is now a postdoc jointly with Arizona State University, Tempe, AZ, USA, and Princeton University, Princeton, NJ, USA.

This research was supported in part by Arizona State University, by the U.S. National Science Foundation under Grant CNS-1422277, and by the U.S. Defense Threat Reduction Agency under Grant HDTRA1-13-1-0029. This research was also supported in part by CyLab and Department of Electrical & Computer Engineering at Carnegie Mellon University.

Consider a social network of n users. The common-interest relation in the network induces a graph $\mathcal{G}$, where each of the n users represents a node in $\mathcal{G}$ and two nodes are connected by an edge if and only if the users they represent are common-interest friends. The relevance of the connectivity properties of $\mathcal{G}$ in the context of large-scale, distributed publish-subscribe services can be seen as follows. Each publisher as well as each subscriber represents a node in $\mathcal{G}$. When publisher v_a posts an interest-specific message `msg`, each node v_b in v_a's circle of common-interest friends receives `msg` and posts `msg` to its own circle of common-interest friends, unless `msg` has already been posted there recently. This process continues iteratively. Obviously, the global dissemination of message `msg` can be achieved if and only if there exists a path between v_a and each subscriber among the other $(n-1)$ nodes of $\mathcal{G}$, which happens if $\mathcal{G}$ is connected, since connectivity means that any two nodes can find at least one path in between. Furthermore, even if at most $(k-1)$ users leave the network, k-connectivity of $\mathcal{G}$ assures the availability of message-dissemination paths between any two remaining nodes, since k-connectivity is defined such that the network remains connected despite the removal of any $(k-1)$ nodes [9] (removing nodes also remove their associated edges).

A possible way to construct the graph $\mathcal{G}$ on n users is as follows. Suppose that there exists an *object pool* $\mathcal{P}_n$ consisting of P_n objects and that each user picks exactly K_n distinct objects uniformly and independently from the object pool; i.e., each user has an *object ring* consisting of K_n objects (we index the parameters by n to study the scaling behavior when n gets large). Two *friends* are said to have a common-interest relation if they have at least d common objects in their object rings. The topology induced by common-interest relations, denoted by $G_d(n, K_n, P_n)$, is known in the literature as a uniform random d-intersection graph [10], [11], [13]. In order to model the friendship network, we use an Erdős-Rényi graph model as in a few prior studies [14], [16]–[18]. Although this model is simple, we will show that when it is coupled with common-interest relations, the induced analysis becomes quite involved. A future direction is to consider more complex models. Under the model of an Erdős-Rényi graph to represent the social network, any two users in the network are friends with each other with probability f_n independently from all other users. As a result, the graph $\mathcal{G}$ to model the common-interest-based subgraph of the social network becomes the intersection of an Erdős-Rényi graph $G(n, f_n)$ and a uniform random d-intersection graph $G_d(n, K_n, P_n)$, where the intersection of two graphs G_1 and G_2 defined on the same node set has the following meaning: two nodes have an edge in between in $G_1 \cap G_2$ if and only if these two nodes have an edge in G_1

TABLE I
Different constraints and their induced graph topologies.

Constraints	Induced graph
interest-based relations: each user selects K_n interests uniformly at random from the same pool of P_n interests; two users establish a common-interest relation if and only if they share at least d interests.	$G_d(n, K_n, P_n)$
social friendships: two users are friends of each other with probability f_n.	$G(n, f_n)$
link failures: the link between two users fails with probability $1 - g_n$, and remains with probability g_n.	$G(n, g_n)$
interest-based relations & social friendships & link failures: our studied system model in consideration of these three types of constriants.	$\mathbb{G}_d(n, K_n, P_n, f_n, g_n)$

and also have an edge in G_2. We denote the above graph $\mathcal{G}$ by $\mathbb{G}_d(n, K_n, P_n, f_n)$ to elipticity express its parameters; i.e.,

$$\mathbb{G}_d(n, K_n, P_n, f_n) := G_d(n, K_n, P_n) \cap G(n, f_n). \quad (1)$$

To consider the resilience of the interest-based social network $\mathbb{G}_d(n, K_n, P_n, f_n)$ against link failure, we consider a simple model where each link fails independently with probability $1-g_n$; i.e., under link failure, each link is preserved with probability g_n. Link failure in social networks may result from adversarial attacks [20], [21], [23], [24]. Then the graph model for the interest-based social network under link failure is obtained by further superimposing an Erdős-Rényi graph $G(n, g_n)$ over $\mathbb{G}_d(n, K_n, P_n, f_n)$. Letting $\mathbb{G}_d(n, K_n, P_n, f_n, g_n)$ denote the induced graph model, we obtain

$$\mathbb{G}_d(n, K_n, P_n, f_n, g_n) = \mathbb{G}_d(n, K_n, P_n, f_n) \cap G(n, g_n). \quad (2)$$

Substituting (1) into (2), we further have

$$\begin{aligned}&\mathbb{G}_d(n, K_n, P_n, f_n, g_n)\\ &= G_d(n, K_n, P_n) \cap G(n, f_n) \cap G(n, g_n). \quad (3)\end{aligned}$$

Table I summarizes the graph notation.

We will study connectivity behavior of graph $\mathbb{G}_d(n, K_n, P_n, f_n, g_n)$ when an arbitrary set of m nodes can fail. Under node failure, we remove the failed nodes and their associated edges from the graph. We will derive a zero–one law for connectivity in $\mathbb{G}_d(n, K_n, P_n, f_n, g_n)$ in the presence of node failure, where the zero-law (resp., one-law) shows that the remaining graph is disconnected (resp., connected) asymptotically. Our results enable us to answer the two key questions for the design of a large-scale, reliable publish-subscribe service: (1) what values should the parameters take in order to achieve connectivity between publisher and subscriber nodes in the interest-based network; and (2) how can reliable message dissemination be achieved when links and nodes are both allowed to fail. These failures could happen as a result of discretionary user action (e.g., a node may decide not to forward a particular message, or all messages, of a particular publisher); or voluntary account deletion (e.g., *Facebook* account deletions are not uncommon events [27]); or involuntary account deletion caused by adversarial attacks [15], [28], [37] (e.g., Agarwalla [28] shows that clickjacking vulnerability found in Linkedin results in involuntary account deletion).

We organize the rest of the paper as follows. We detail the analytical results as Theorem 1 in Section II. In Section III, we provide experimental results to confirm Theorem 1. Subsequently, we explain the basic ideas for proving Theorem 1 in Section V. Section IV surveys related work. Finally, we conclude the paper in Section VI.

II. The Results

We present and discuss our results in this section. The natural logarithm function is given by $\ln$. All limits are understood with $n \to \infty$. We use the standard asymptotic notation $o(\cdot), O(\cdot), \Omega(\cdot), \omega(\cdot), \Theta(\cdot), \sim$; see [8, Page 2-Footnote 1]. Throughout the paper, m and d are positive constant integers so they do not scale with n. The notation $\mathbb{P}[\mathcal{E}]$ denotes the probability that an event $\mathcal{E}$ happens.

Theorem 1 below presents a zero–one law for connectivity in graph $\mathbb{G}_d(n, K_n, P_n, f_n, g_n)$ under node failure. Note that g_n (more precisely, its complement $1 - g_n$) in $\mathbb{G}_d(n, K_n, P_n, f_n, g_n)$ already encodes link failure. The zero-law means that the probability of connectivity asymptotically converges to 0 under some conditions and the one-law means that the probability of connectivity asymptotically converges to 1 under some other conditions.

Theorem 1. *For a graph $\mathbb{G}_d(n, K_n, P_n, f_n, g_n)$ which models an interest-based social network under link failure, if there exists a sequence α_n with $\lim_{n\to\infty} \alpha_n \in [-\infty, +\infty]$ such that*

$$f_n \cdot g_n \cdot \sum_{u=d}^{K_n} \frac{\binom{K_n}{u}\binom{P_n-K_n}{K_n-u}}{\binom{P_n}{K_n}} = \frac{\ln n + m \ln\ln n + \alpha_n}{n}, \quad (4)$$

then it holds under $P_n = \Omega(n)$ and $\frac{{K_n}^2}{P_n} = o(1)$ that

$$\lim_{n\to\infty} \mathbb{P}\begin{bmatrix}\mathbb{G}_d(n, K_n, P_n, f_n, g_n) \textit{ is connected} \\ \textit{even after an arbitrary set of } m \textit{ nodes fail.}\end{bmatrix} \quad (5)$$

$$= \begin{cases} 0, & \textit{if } \lim_{n\to\infty} \alpha_n = -\infty, & (6a)\\ 1, & \textit{if } \lim_{n\to\infty} \alpha_n = \infty. & (6b)\end{cases}$$

It is straightforward to show that the left hand side of (4) equals the edge probability of $\mathbb{G}_d(n, K_n, P_n, f_n, g_n)$. To see this, for two users selecting K_n interests independently from the same pool of P_n interests, the probability that they share at least d interests equals $\sum_{u=d}^{K_n} \frac{\binom{K_n}{u}\binom{P_n-K_n}{K_n-u}}{\binom{P_n}{K_n}}$. In addition, f_n is the probability that they are friends, and g_n is the probability of a link being active.

Theorem 1 shows that a critical scaling for connectivity in graph $\mathbb{G}_d(n, K_n, P_n, f_n, g_n)$ under the failure of m nodes is that the left hand side of (4) equals $\frac{\ln n + m \ln \ln n}{n}$. When $\mathbb{G}_d(n, K_n, P_n, f_n, g_n)$ is connected even after an arbitrary set of m nodes fail, we can equivalently say that $\mathbb{G}_d(n, K_n, P_n, f_n, g_n)$ is $(m+1)$-connected (i.e., k-connected for $k = m + 1$).

We discuss the practicality of the conditions $P_n = \Omega(n)$ and $\frac{K_n^2}{P_n} = o(1)$ in Theorem 1. Both conditions are enforced here merely for technical reasons, but they hold often in realistic social network applications because it is expected [25] that the object pool size P_n will be much larger than both the number n of participating users and the number K_n of objects associated with each user.

We explain the basic ideas for proving Theorem 1 in Section V.

III. Experimental Results

We present experiments below to confirm our theoretical results of connectivity. Since the experiments are for finite networks, we suppress the subscript n in $\mathbb{G}_d(n, K_n, P_n, f_n, g_n)$ to have $\mathbb{G}_d(n, K, P, f, g)$. In the presence of node failure, we study the connectivity behavior

- when the object ring size K varies given different m in Figure 1-(a), where m denotes the number of failed nodes,
- when the object pool size P varies given different m in Figure 1-(b),
- when the link-active probability g varies given different m in Figure 1-(c), and
- when m varies given different K in Figure 1-(d).

For each data point, we generate 500 independent samples of graph $\mathbb{G}_d(n, K, P, f, g)$ with m failed nodes, record the count that the remaining graph is connected, and then divide the count by 500 to obtain the corresponding empirical probability of network connectivity. Moreover, in each subfigure, the vertical line presents the *critical* parameter for connectivity based on (4): the critical object ring size in Figure 1-(a), the critical object pool size in Figure 1-(b), and the critical link-active probability in Figure 1-(c). Namely, in Figure 1-(a), each the vertical line stands for the minimum integer K^* that satisfies $f \cdot g \cdot \sum_{u=d}^{K^*} \frac{\binom{K^*}{u}\binom{P-K^*}{K^*-u}}{\binom{P}{K^*}} \geq \frac{\ln n + m \ln \ln n}{n}$. In Figure 1-(b), each vertical line represents the maximal integer P^* that satisfies $f \cdot g \cdot \sum_{u=d}^{K} \frac{\binom{K}{u}\binom{P^*-K}{K-u}}{\binom{P^*}{K}} \geq \frac{\ln n + m \ln \ln n}{n}$. In Figure 1-(c), each vertical line shows the probability g^* that satisfies $f \cdot g^* \cdot \sum_{u=d}^{K} \frac{\binom{K}{u}\binom{P-K}{K-u}}{\binom{P}{K}} = \frac{\ln n + m \ln \ln n}{n}$, where the computed g^* for each line in Figure 1-(c) is less than 1 and hence is indeed a probability.

In each subfigure, we clearly observe the transitional behavior of connectivity, and the transition point is around the critical parameter illustrated by the vertical line. Hence, the experiments have confirmed our Theorem 1.

IV. Related Work

Graph $G_d(n, K_n, P_n)$ models the topology of an interest-based social network under full visibility, where full visibility means that any pair of nodes have edges in between so the only requirement for a link is the sharing of at least d interests.

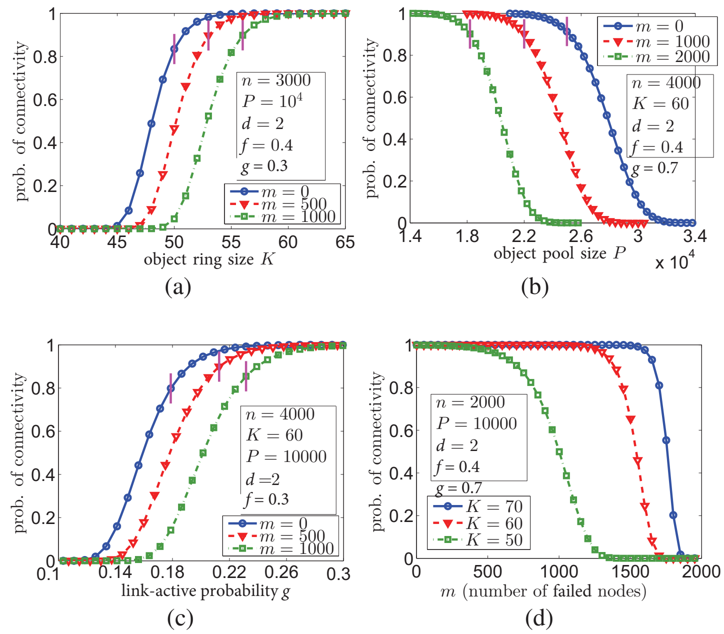

Fig. 1. We plot the connectivity probability of the network $\mathbb{G}_d(n, K, P, f, g)$ when m nodes are attacked and thus fail. Note that g (more precisely, its complement $1 - g$) in $\mathbb{G}_d(n, K, P, f, g)$ encodes link failure.

Graph $G_d(n, K_n, P_n)$ has been studied in the literature in terms of connectivity [19], [29], k-connectivity [22], [30], [31], node degree [26], [34] and k-robustness [2]. With $s(K_n, P_n, d)$ being the edge probability of $G_d(n, K_n, P_n)$, it is shown in [29] that under $c_1 n^{1/d} (\ln n)^{-1/d} \leq P_n \leq c_2 n^{1/d} (\ln n)^{2/5 - 1/d}$ for some positive constants c_1 and c_2, with α_n defined through $s(K_n, P_n, d) = \frac{\ln n + \alpha_n}{n}$, then $G_d(n, K_n, P_n)$ is disconnected with high probability if $\lim_{n\to\infty} \alpha_n = -\infty$ and connected with high probability if $\lim_{n\to\infty} \alpha_n = \infty$, where an event happens "with high probability" if its probability converges to 1 as $n \to \infty$. Bloznelis and Rybarczyk [29] extend the result to k-connectivity, which is also studied in [22]. Other properties of $G_d(n, K_n, P_n)$ are considered as well in the literature. For example, Bloznelis *et al.* [11] demonstrate that a connected component with at at least a constant fraction of n emerges with high probability when the edge probability $s(K_n, P_n, d)$ exceeds $1/n$. When $d = 1$, graph $G_1(n, K_n, P_n)$ models the topology of an interest-based network where two users only need to share one interest to form an edge. For $G_1(n, K_n, P_n)$, its connectivity has been investigated extensively [32], [33], [35], [36], [40]. In particular, Di Pietro *et al.* [36] establish that under $P_n \geq n$ and $\frac{{K_n}^2}{P_n} \sim \frac{c \ln n}{n}$ for a constant $c > 1$, $G_1(n, K_n, P_n)$ is connected with high probability, where the relation $f_n \sim g_n$ for two positive sequences f_n and g_n means $\lim_{n\to\infty}(f_n/g_n) = 1$; i.e., f_n and g_n are asymptotically equivalent. Yağan and Makowski [35] prove that under $P_n = \Omega(n)$, with α_n defined by $\frac{{K_n}^2}{P_n} = \frac{\ln n + \alpha_n}{n}$, then $G_1(n, K_n, P_n)$ is disconnected with high probability if $\lim_{n\to\infty} \alpha_n = -\infty$ and connected with high probability if $\lim_{n\to\infty} \alpha_n = \infty$.

Erdős and Rényi [38] introduce the random graph model $G(n, p_n)$ defined on a node set with size n such that an edge between any two nodes exists with probability p_n *independently* of all other edges. In a few prior studies [14], [16]–[18], graph

$G(n, p_n)$ is used to model the topology of an online social network. For graph $G(n, p_n)$, Erdős and Rényi derive a zero–one law for connectivity in [38] and extend the result to k-connectivity in [39]. Their k-connectivity result [39] is that with α_n defined through $p_n = \frac{\ln n + (k-1)\ln\ln n + \alpha_n}{n}$, then $G(n, p_n)$ is not k-connected with high probability if $\lim_{n\to\infty} \alpha_n = -\infty$ and k-connected with high probability if $\lim_{n\to\infty} \alpha_n = \infty$.

Interest-based social networks have been studied in the literature [25], [41], [42], but existing studies often lack formal analyses (in particular for connectivity under node and link failures). In this paper, we model an interest-based social network by superimposing the common-interest relations over a social network, and then formally analyze its connectivity when links and nodes are allowed to fail.

V. Ideas for Proving Theorem 1

Below we present the ideas to prove Theorem 1. We will write $\mathbb{G}_d(n, K_n, P_n, f_n, g_n)$ as $\mathbb{G}_q$ for notation brevity.

A. Confining $|\alpha_n|$ as $o(\ln n)$ in Theorem 1

To prove Theorem 1, we will show that an extra condition $|\alpha_n| = o(\ln n)$ can be introduced. Specifically, in the full version [43], we use the idea of graph coupling [44], [45] to show

$$\text{Theorem 1 under } |\alpha_n| = o(\ln n) \implies \text{Theorem 1.} \quad (7)$$

B. Connectivity versus minimum degree

The probability in (5) equals $\mathbb{P}[\,\mathbb{G}_q \text{ is } (m+1)\text{-connected.}\,]$. Clearly, if a graph G is $(m+1)$-connected, then the minimum (node) degree of G is at least $m+1$ [46]. Thus,

$$\begin{aligned}&\mathbb{P}[\,\mathbb{G}_q \text{ is } (m+1)\text{-connected.}\,]\\ &\leq \mathbb{P}[\,\text{Minimum degree of } \mathbb{G}_q \text{ is at least } m+1.\,]\end{aligned} \quad (8)$$

and

$$\begin{aligned}&\mathbb{P}[\,\mathbb{G}_q \text{ is } (m+1)\text{-connected.}\,]\\ &= \mathbb{P}[\,\text{Minimum degree of } \mathbb{G}_q \text{ is at least } m+1.\,]\\ &\quad - \mathbb{P}\left[\begin{array}{l}\text{Minimum degree of } \mathbb{G}_q \text{ is at least } m+1,\\ \text{but is not } (m+1)\text{-connected.}\end{array}\right].\end{aligned} \quad (9)$$

Given (7) (8) and (9), we will complete proving Theorem 1 once we establish Lemmas 1 and 2 below.

Lemma 1. *For a graph $\mathbb{G}_q$ under $P_n = \Omega(n)$, $\frac{{K_n}^2}{P_n} = o(1)$, Equation (4) and $|\alpha_n| = o(\ln n)$, then*

$$\lim_{n\to\infty} \mathbb{P}\left[\begin{array}{l}\text{Minimum degree of } \mathbb{G}_q\\ \text{is at least } m+1.\end{array}\right] = \begin{cases}0, & \text{if } \lim_{n\to\infty} \alpha_n = -\infty, \quad (10a)\\ 1, & \text{if } \lim_{n\to\infty} \alpha_n = \infty. \quad (10b)\end{cases}$$

Lemma 1 presents a zero–one law for the property of minimum degree being at least $m+1$ in $\mathbb{G}_q$ via (10a) and (10b). In the next subsection, we explain the idea of proving (10a) and (10b) by the method of moments.

Lemma 2. *For a graph $\mathbb{G}_q$ under $P_n = \Omega(n)$, $\frac{{K_n}^2}{P_n} = o(1)$, Equation (4) and $|\alpha_n| = o(\ln n)$, then*

$$\lim_{n\to\infty} \mathbb{P}\left[\begin{array}{l}\text{Minimum degree of } \mathbb{G}_q \text{ is at least } m+1,\\ \text{but is not } (m+1)\text{-connected.}\end{array}\right] = 0. \quad (11)$$

Lemma 2 is established in the full version [43].

C. Method of moments to prove results (10a) and (10b)

We prove (10a) and (10b) by the method of moments [47, Page 55] applied to the number of nodes with certain degree in $\mathbb{G}_q$.

We let the n nodes of $\mathbb{G}_q$ be $\{v_1, v_2, \ldots, v_n\}$. With indicator variables $\chi_{n,\ell,i}$ for $i = 1, \ldots, n$ defined by

$$\begin{aligned}\chi_{n,\ell,i} &= \mathbf{1}\,[\text{Node } v_i \text{ is isolated in } \mathbb{G}_q.]\\ &= \begin{cases}1, & \text{if node } v_i \text{ in } \mathbb{G}_q \text{ has degree } \ell,\\ 0, & \text{otherwise,}\end{cases}\end{aligned}$$

then $I_{n,\ell}$ denoting the number of nodes with degree ℓ equals

$$I_{n,\ell} := \sum_{i=1}^{n} \chi_{n,\ell,i}.$$

The minimum degree of $\mathbb{G}_q$ is at least $m+1$ if and only if $I_{n,\ell} = 0$ for $\ell = 0, 1, \ldots, m$.

The method of first moment [47, Equation (3.10), Page 55] relies on the well-known bound

$$1 - \mathbb{E}[I_{n,\ell}] \leq \mathbb{P}[I_{n,\ell} = 0], \quad (12)$$

where $\mathbb{E}[\cdot]$ in this paper stands for the expectation of a random variable, while the method of second moment [47, Remark 3.1, Page 55] has its starting point in the inequality

$$\mathbb{P}[I_{n,\ell} = 0] \leq 1 - \frac{\mathbb{E}[I_{n,\ell}]^2}{\mathbb{E}\left[{I_{n,\ell}}^2\right]}. \quad (13)$$

Noting that the random variables $\chi_{n,\ell,1}, \ldots, \chi_{n,\ell,n}$ are exchangeable due to node symmetry, we find

$$\mathbb{E}[I_{n,\ell}] = n\mathbb{E}[\chi_{n,\ell,1}] \quad (14)$$

and

$$\begin{aligned}\mathbb{E}\left[{I_{n,\ell}}^2\right] &= n\mathbb{E}\left[{\chi_{n,\ell,1}}^2\right] + n(n-1)\mathbb{E}[\chi_{n,\ell,1}\chi_{n,\ell,2}]\\ &= n\mathbb{E}[\chi_{n,\ell,1}] + n(n-1)\mathbb{E}[\chi_{n,\ell,1}\chi_{n,\ell,2}],\end{aligned} \quad (15)$$

where the last step uses $\mathbb{E}\left[{\chi_{n,\ell,1}}^2\right] = \mathbb{E}[\chi_{n,\ell,1}]$ as $\chi_{n,\ell,1}$ is a binary random variable. It then follows from (14) and (15) that

$$\frac{\mathbb{E}\left[{I_{n,\ell}}^2\right]}{\mathbb{E}[I_{n,\ell}]^2} = \frac{1}{n\mathbb{E}[\chi_{n,\ell,1}]} + \frac{n-1}{n} \cdot \frac{\mathbb{E}[\chi_{n,\ell,1}\chi_{n,\ell,2}]}{(\mathbb{E}[\chi_{n,\ell,1}])^2}. \quad (16)$$

The desired one-law (10b) means $\lim_{n\to\infty} \mathbb{P}[I_{n,\ell} = 0] = 1$ for $\ell = 0, 1, \ldots, m$ under $\lim_{n\to\infty} \alpha_n = \infty$. For $\ell = 0, 1, \ldots, m$, from (12) and (14), $\lim_{n\to\infty} \mathbb{P}[I_{n,\ell} = 0] = 1$ will be proved once we show $\lim_{n\to\infty} \left(n\mathbb{E}[\chi_{n,\ell,1}]\right) = 0$. The desired zero-law (10a) follows from a stronger result $\lim_{n\to\infty} \mathbb{P}[I_{n,m} = 0] = 0$ under $\lim_{n\to\infty} \alpha_n = -\infty$. From (13) and (15), we will obtain $\lim_{n\to\infty} \mathbb{P}[I_{n,m} = 0] = 0$ once deriving $\lim_{n\to\infty} \left(n\mathbb{E}[\chi_{n,m,1}]\right) = \infty$ and $\frac{\mathbb{E}[\chi_{n,m,1}\chi_{n,m,2}]}{(\mathbb{E}[\chi_{n,m,1}])^2} \leq 1 + o(1)$. The remaining details are given in the full version [43].

VI. Conclusion

In this paper, we present a zero–one law for connectivity in an interest-based social network against node and link failures. The network is modeled by composing a uniform d-intersection graph with two Erdős–Rényi graphs, where the uniform d-intersection graph characterizes common-interest relations and the Erdős–Rényi graphs capture friendships and link failures. Experimental results are shown to be in agreement with our theoretical findings.

REFERENCES

[1] R. Kumar, J. Novak, and A. Tomkins, "Structure and evolution of online social networks," in *Link mining: models, algorithms, and applications*, pp. 337–357, Springer, 2010.

[2] J. Zhao, "Threshold functions in random s-intersection graphs," in *Allerton Conference on Communication, Control, and Computing (Allerton)*, pp. 1358–1365, 2015.

[3] X. Yang, H. Steck, and Y. Liu, "Circle-based recommendation in online social networks," in *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1267–1275, 2012.

[4] X. Qian, H. Feng, G. Zhao, and T. Mei, "Personalized recommendation combining user interest and social circle," *IEEE Transactions on Knowledge and Data Engineering*, vol. 26, no. 7, pp. 1763–1777, 2014.

[5] J. J. McAuley and J. Leskovec, "Learning to discover social circles in ego networks," in *Conference on Neural Information Processing Systems (NIPS)*, vol. 2012, pp. 548–56, 2012.

[6] D. Kato, K. Elkhiyaoui, K. Kunieda, K. Yamada, and P. Michiardi, "A scalable interest-oriented peer-to-peer pub/sub network," *Peer-to-Peer Networking and Applications*, vol. 4, no. 2, pp. 165–177, 2011.

[7] N. Kayastha, D. Niyato, P. Wang, and E. Hossain, "Applications, architectures, and protocol design issues for mobile social networks: A survey," *Proceedings of the IEEE*, vol. 99, no. 12, pp. 2130–2158, 2011.

[8] J. Zhao, O. Yağan, and V. Gligor, "k-Connectivity in random key graphs with unreliable links," *IEEE Transactions on Information Theory*, vol. 61, no. 7, pp. 3810–3836, July 2015. Online: http://arxiv.org/abs/1206.1531

[9] B. Bollobás, *Random graphs*. Cambridge Studies in Advanced Mathematics, 2001.

[10] M. Bloznelis, "Degree and clustering coefficient in sparse random intersection graphs," *The Annals of Applied Probability*, vol. 23, no. 3, pp. 1254–1289, 2013.

[11] M. Bloznelis, J. Jaworski, and K. Rybarczyk, "Component evolution in a secure wireless sensor network," *Networks*, vol. 53, pp. 19–26, January 2009.

[12] J. Zhao, "Initial spreaders in online social networks," in *Allerton Conference on Communication, Control, and Computing (Allerton)*, 2016.

[13] M. Bradonjić, A. Hagberg, N. W. Hengartner, and A. G. Percus, "Component evolution in general random intersection graphs," in *International Workshop on Algorithms and Models for the Web-Graph (WAW)*, pp. 36–49, Springer, 2010.

[14] M. E. J. Newman, D. J. Watts, and S. H. Strogatz, "Random graph models of social networks," *Proceedings of the National Academy of Sciences of the United States of America (PNAS)*, vol. 99, pp. 2566–2572, Feb. 2002.

[15] J. Zhao, "On the resilience to node capture attacks of secure wireless sensor networks," in *Allerton Conference on Communication, Control, and Computing (Allerton)*, pp. 887–893, 2015.

[16] S. R. Etesami and T. Başar, "Complexity of equilibrium in competitive diffusion games on social networks," *Automatica*, vol. 68, pp. 100–110, 2016.

[17] R. Korolov, J. Peabody, A. Lavoie, S. Das, M. Magdon-Ismail, and W. Wallace, "Actions are louder than words in social media," in *IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, pp. 292–297, 2015.

[18] C. D. Brummitt, S. Chatterjee, P. S. Dey, D. Sivakoff, *et al.*, "Jigsaw percolation: What social networks can collaboratively solve a puzzle?," *The Annals of Applied Probability*, vol. 25, no. 4, pp. 2013–2038, 2015.

[19] J. Zhao, "Parameter control in predistribution schemes of cryptographic keys," in *IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pp. 863–867, 2015.

[20] N. Shrivastava, A. Majumder, and R. Rastogi, "Mining (social) network graphs to detect random link attacks," in *IEEE International Conference on Data Engineering*, pp. 486–495, IEEE, 2008.

[21] R.-H. Li, J. X. Yu, X. Huang, H. Cheng, and Z. Shang, "Measuring robustness of complex networks under mvc attack," in *ACM International Conference on Information and Knowledge Management*, pp. 1512–1516, ACM, 2012.

[22] J. Zhao, O. Yağan, and V. Gligor, "On k-connectivity and minimum vertex degree in random s-intersection graphs," in *ACM-SIAM Meeting on Analytic Algorithmics and Combinatorics (ANALCO)*, 2015.

[23] Y. E. Sagduyu and Y. Shi, "Navigating a mobile social network," *IEEE Wireless Communications*, vol. 22, no. 5, pp. 122–128, 2015.

[24] M. Conti, A. Hasani, and B. Crispo, "Virtual private social networks and a facebook implementation," *ACM Transactions on the Web (TWEB)*, vol. 7, no. 3, p. 14, 2013.

[25] J. Zhang, Y. Liu, and Y. Chen, "Social learning in networks of friends versus strangers," *Marketing Science*, vol. 34, no. 4, pp. 573–589, 2015.

[26] J. Zhao, "Topological properties of wireless sensor networks under the q-composite key predistribution scheme with unreliable links," *IEEE/ACM Transactions on Networking*, 2017.

[27] http://www.pcworld.com/article/196861/study_60_percent_of_facebook_users_mulling_to_quit.html.

[28] http://seclists.org/fulldisclosure/2011/Oct/359.

[29] M. Bloznelis and K. Rybarczyk, "k-Connectivity of uniform s-intersection graphs," *Discrete Mathematics*, vol. 333, no. 0, pp. 94–100, 2014.

[30] J. Zhao, "Designing secure networks under q-composite key predistribution with unreliable links," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2017.

[31] M. Bloznelis and T. Łuczak, "Perfect matchings in random intersection graphs," *Acta Mathematica Hungarica*, vol. 138, no. 1-2, pp. 15–33, 2013.

[32] K. Rybarczyk, "Diameter, connectivity and phase transition of the uniform random intersection graph," *Discrete Mathematics*, vol. 311, 2011.

[33] S. Blackburn and S. Gerke, "Connectivity of the uniform random intersection graph," *Discrete Mathematics*, vol. 309, no. 16, August 2009.

[34] J. Zhao, O. Yağan, and V. Gligor, "On topological properties of wireless sensor networks under the q-composite key predistribution scheme with on/off channels," in *IEEE International Symposium on Information Theory (ISIT)*, 2014.

[35] O. Yağan and A. M. Makowski, "Zero–one laws for connectivity in random key graphs," *IEEE Transactions on Information Theory*, vol. 58, pp. 2983–2999, May 2012.

[36] R. Di Pietro, L. V. Mancini, A. Mei, A. Panconesi, and J. Radhakrishnan, "Redoubtable sensor networks," *ACM Transactions on Information and System Security (TISSEC)*, vol. 11, no. 3, pp. 13:1–13:22, 2008.

[37] J. Zhao, "On resilience and connectivity of secure wireless sensor networks under node capture attacks," *IEEE Transactions on Information Forensics and Security*, vol. 12, no. 3, pp. 557–571, 2017.

[38] P. Erdős and A. Rényi, "On random graphs, I," *Publicationes Mathematicae (Debrecen)*, vol. 6, pp. 290–297, 1959.

[39] P. Erdős and A. Rényi, "On the strength of connectedness of random graphs," *Acta Math. Acad. Sci. Hungar*, pp. 261–267, 1961.

[40] J. Zhao, "A comprehensive guideline for choosing parameters in the Eschenauer–Gligor key predistribution," in *Allerton Conference on Communication, Control, and Computing (Allerton)*, 2016.

[41] R. Bosagh Zadeh, A. Goel, K. Munagala, and A. Sharma, "On the precision of social and information networks," in *Proceedings of the first ACM Conference on Online Social Networks*, pp. 63–74, ACM, 2013.

[42] Q. Fang, J. Sang, C. Xu, and Y. Rui, "Topic-sensitive influencer mining in interest-based social media networks via hypergraph learning," *IEEE Transactions on Multimedia*, vol. 16, no. 3, pp. 796–812, 2014.

[43] J. Zhao, "Modeling interest-based social networks: Superimposing Erdos–Renyi graphs over random intersection graphs," 2017. Full version of this paper, available online at https://sites.google.com/site/workofzhao/ICASSP17-social.pdf

[44] K. Rybarczyk, "Sharp threshold functions for the random intersection graph via a coupling method," *The Electronic Journal of Combinatorics*, vol. 18, pp. 36–47, 2011.

[45] K. Rybarczyk, "The coupling method for inhomogeneous random intersection graphs," *ArXiv e-prints*, Jan. 2013. Available online at http://arxiv.org/abs/1301.0466

[46] P. Gupta and P. R. Kumar, "Critical power for asymptotic connectivity in wireless networks," in *Proc. IEEE CDC*, pp. 547–566, 1998.

[47] S. Janson, T. Łuczak, and A. Ruciński, *Random graphs*. Wiley-Interscience Series on Discrete Mathematics and Optimization, 2000.