Capacity Scaling of Wireless Social Networks

Cheng Wang, Lu Shao, Zhong Li, Lei Yang, *Member, IEEE, Computer Society*, Xiang-Yang Li, *Senior Member, IEEE*, and Changjun Jiang

Abstract—In this paper, we investigate capacity scaling laws of wireless social networks under the social-based session formation. We model a wireless social network as a three-layered structure, consisting of the *physical layer*, *social layer*, and *session layer*; and we introduce a cross-layer *distance&density-aware* model, called the *population-based formation model*, under which: 1) for each node v_k, the number of its friends/followers, denoted by q_k, follows a Zipf's distribution with degree clustering exponent γ; 2) q_k *anchor points* are independently chosen according to a probability distribution with density function proportional to $\left(\mathbf{E}_{k,X}\right)^{-\beta}$, where $\mathbf{E}_{k,X}$ is the expected number of nodes (population) within the distance $|v_k - X|$ to v_k, and β is the clustering exponent of friendship formation; 3) finally, q_k nodes respectively nearest to those q_k anchor points are selected as the friends of v_k. We present the general density function of social relationship distribution, with general distribution of physical layer, serving as the basis for studying general capacity of wireless social networks. As the first step of addressing this issue, for the homogeneous physical layer, we derive the *social-broadcast* capacity under both *generalized physical* and *protocol* interference models, taking into account general clustering exponents of both friendship degree and friendship formation in a 2-dimensional parameter space, i.e., $(\gamma, \beta) \in [0, \infty)^2$. Importantly, we notice that the adopted model with homogenous physical layer does *not* sufficiently reflect the advantages of the population-based formation model in terms of realistic validity and practicability. Accordingly, we introduce a random network model, called the *center-clustering random model* (CCRM) with node distribution exponent $\delta \in [0, \infty)$, highlighting the clustering and inhomogeneity property in real-life networks, and discuss how to further derive more general network capacity over 3-dimensional parameter space $(\delta, \gamma, \beta) \in [0, \infty)^3$ based on our results over $(\gamma, \beta) \in [0, \infty)^2$.

Index Terms—Scaling laws, social networks, wireless networks, network capacity

1 Introduction

Wireless networks are generally the wireless communication implementations for real-life networking applications. Then, research issues of wireless networks usually come from and aim at the challenges of wireless technology in specific applications, e.g., wireless sensor networks, wireless local area networks, and *wireless social networks*, a wireless implementation of social networks, which is the focus of this work. In social networks, the relationship/edge between users/vertices represents a specific interdependency, such as co-authorship, citationship, or friendship, and so on. Based on massive data sets of large-scale real-world online social networks, such as Myspace [1], Twitter [2], Flickr [3], LiveJournal [4], and Facebook [5], extensive studies validate respectively that these two most representative features of *complex networks*, i.e., the *small-world phenomenon* and *scale-free degree distribution*, nearly hold in online social networks. Wireless social networks can be analyzed from a layered perspective, i.e., the social network of users can be regarded as an overlay network over their physical communication network. Therefore, it is necessary for studying wireless social networks to take the property of social networks into account.

As online social networking services (SNSs) are becoming more and more popular and the adoption rates of smart wireless devices like smartphones are increasing aggressively, wireless social network applications will be recognized as the typical instances of large-scale wireless networks. Therefore, it is significant to investigate the fundamental limits of such a large-scale wireless system. As an important metric of fundamental limits, capacity scaling laws of wireless networks, i.e., the scaling of the throughput capacity in the limit when the size of network gets large, have been extensively studied in depth from both theoretical and practical perspectives, [6], [7], [8], [9], [10], [11], [12], since Gupta and Kumar [13] took the lead to study the unicast capacity for homogeneous wireless *random* and *arbitrary* networks. In wireless social networks, source-destination associations depend on their social relationships that had been shown to be inhomogeneous, however, traffic sessions are traditionally assumed to be formed independently in a uniformly random fashion in the literature, which makes the existing results *not* applicable to wireless social networks. Accordingly, this paper aims to introduce the social-based session formation model conforming with wireless social networks, and derive the corresponding capacity scaling laws.

Main challenges and our solutions. We list the following three main challenges in addressing the problem.

Challenge I: Modeling social formation based on geography. On the one hand, since the distance over which data are carried under every session is one of key factors determining

- *C. Wang, L. Shao, Z. Li, and C. Jiang are with the Department of Computer Science and Technology, Tongji University, and with the Key Laboratory of Embedded System and Service Computing, Ministry of Education, Shanghai, China. E-mail: {3chengwang, shaolu1991}@gmail.com, lizhong007@hotmail.com, cjjiang@tongji.edu.cn.*
- *L. Yang is with the Institute of Trustworthy Networks and Systems, Tsinghua University, Beijing, China. E-mail: young@tagsys.org.*
- *X.-Y. Li is with the Department of Computer Science, Illinois Institute of Technology, 10 West 31st Street, Chicago, IL 60616. E-mail: xli@cs.iit.edu.*

Manuscript received 6 Dec. 2013; revised 5 May 2014; accepted 31 May 2014. Date of publication 25 June 2014; date of current version 5 June 2015.
Recommended for acceptance by Y. Liu.
For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below.
Digital Object Identifier no. 10.1109/TPDS.2014.2333524

the throughput capacity of network, the session formation model should be necessarily relevant to the geography. On the other hand, some classical experiments showed or implied that social relationships among users strongly depend on their geographic location, [14], [15], [16]. Thus, under the assumption that a session is built only among the users with social relationships (We also call them "friends" here.), i.e., the source chooses its destination only from its friends, a geography-based social relationship formation model acts as the precondition for the session formation model. Some theoretical models have been advanced to model the social relationship formation depending on geography in social networks. Kleinberg [17] initiated a *distance-based* social model relating geographical distance and social friendship, in which the probability of befriending a particular user is inversely proportional to the positive power of the distance. In [18] and [19], the distance-based social model had been adopted in studying the unicast capacity of wireless social networks. Intuitively, such a session pattern leads to a mismatch with the characteristics of data traffic in real life social networks. Furthermore, recent experimental studies showed that the distance-based models do not really accord with the distribution of social relationships in social networks, [20]. Therefor, it is necessary to derive the more practical results on capacity scaling of wireless social networks by introducing more realistic social models.

Liben-Nowell et al. [20] stated that the social formation indeed depends on both distance and density, and introduced the *rank-based model*, where the probability of befriending a particular person is inversely proportional to the positive power of the number of closer people. To the best of our knowledge, the rank-based model is the most realistic theoretical model for social formation. But it still has shortcoming on analyzing capacity scaling laws in terms of the convenience of analysis or theoretical basis and rigor, because it directly selects nodes instead of points/positions, which leads that the distances of sessions are dependent and then causes analysis difficulty in bounding the sum of transport distances of sessions with multiple destinations, [13], [21]. Then, for addressing the capacity scaling laws of wireless networks, it should be the first step to introduce a new distance&density-aware social formation model that is suitable for capacity analysis, while keeping the advantages of rank-based model.

Our solution for challenge I. We model a wireless social network as a three-layered structure, consisting of the *physical layer*, *social layer*, and *session layer*, as illustrated in Fig. 1. On the basis of rank-based model, we present a cross-layer *distance&density-aware* social model good at the analysis of capacity scaling laws, called the *population-based formation model* $\mathbb{P}(\delta, \gamma, \beta)$, where δ, γ, and β are the clustering exponents of node distribution, friendship degree and friendship formation, respectively. Under $\mathbb{P}(\delta, \gamma, \beta)$: 1) for each node v_k, the number of its friends, denoted by q_k, follows a Zipf's distribution with friendship degree clustering exponent γ; 2) q_k *anchor points* are independently chosen according to a probability distribution with density function proportional to $\mathbf{E}_{k,X}^{-\beta}$, where $\mathbf{E}_{k,X}$ is the expected number of nodes (population) within the distance $|v_k - X|$ to v_k, and β is the clustering exponent of friendship formation; 3) finally, select q_k nodes respectively nearest to those q_k anchor points as the friends of v_k. Please refer to Section 2.2.2 for the detailed discussion on the advantages of this model.

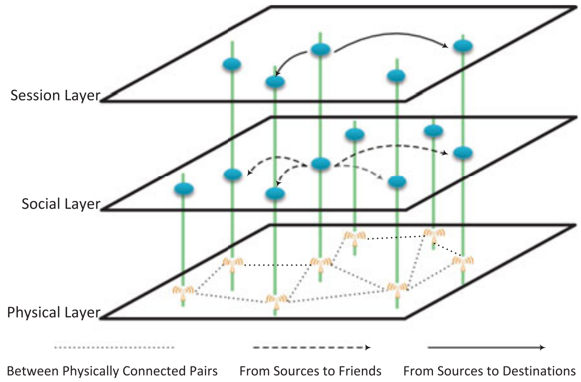


Fig. 1. Layered system model.

Challenge II: Bounding sum of distances for all sessions. The sum of transport distances for all sessions is the prerequisite for bounding the capacity scaling laws, [13], [22]. The long-tailed property of both the destination number (Zipf's Distribution) and destination distribution (Power Law Distribution) causes the transport distances of resulted sessions to be significantly inhomogeneous, which makes it more difficult to bound such a sum.

Our solution for challenge II. We present the density function of general social friendships distribution (Theorem 1) and bounds on length of euclidean spanning trees (ESTs) over nodes chosen according to this density function (Theorem 2). The results can act as the basis for addressing the capacity of wireless social networks.

Challenge III: Deriving general results on capacity scaling laws. The complete result under the system model includes the capacity for every point in the three-dimensional parameter space, i.e., $(\delta, \gamma, \beta) \in [0, \infty)^3$. The complexity of analysis is substantially increased due to the involvement of multiple clustering exponents.

Our solution for challenge III. As the first work under this general model, this paper derives the capacity for *social-broadcast* sessions under the model $\mathbb{P}(\delta = 0, \gamma, \beta)$, taking into account general clustering exponents of both friendship degree and friendship formation, i.e., $(\gamma, \beta) \in [0, \infty)^2$ (Theorem 3). In addition, we probe the feasibility of studying network capacity under the social model with inhomogeneous physical layer, by extending the basic theorem (Theorem 1) for the general distribution of anchor points.

The rest of the paper is organized as follows. In Section 2, we introduce the network model. In Section 3, we present the preliminary results for the anchor points distributed in the center-clustering random model (CCRM) according to the proposed population-based formation scheme, which lay the foundation for addressing capacity scaling laws for population-based social formation model on general physical layers. Based on these results, we derive the social-broadcast capacity for homogeneous physical layer in Section 4, acting as the first step of investigating general results on the capacity of wireless social networks. Finally, we conclude the paper and discuss some topics for future research in Section 5.

2 System Model

We present a three-layer perspective for the wireless social network with social-based sessions, consisting of the *physical layer*, *social layer* and *session layer*, as in Fig. 1.

Throughout the paper, we let $\mathbf{E}[X]$ denote the mean of a random variable X.

2.1 Physical Layer Deployment

We introduce a random network model, called the *center-clustering random model*, highlighting the clustering and inhomogeneity property in real-life networks.

2.1.1 Center-Clustering Random Model

We consider the network composed of a random number of N wireless ad hoc nodes/users distributed over a square region of area $S := n$, where $\mathbf{E}[N] = n$. To avoid border effects, we consider wraparound conditions at the network edges, i.e., the network area is assumed to be the surface of a two-dimensional Torus $\mathcal{O}$. To simplify the description, we assume that the number of nodes is exactly n, and denote the set of nodes by $\mathcal{V} = \{v_k\}_{k=1}^{n}$, without changing our results in order sense, [6], [7], [11].

To emulate the clustering behavior of users distribution in wireless networks, we construct the center-clustering random model by the following procedure: First, making a center of $\mathcal{O}$ as the center point, denoted by O. Then, the center point O generates a point process of nodes whose local intensity at position X is given by $\mathbf{d}(X) = n \cdot \kappa(O, X)$, where $\kappa(O, \cdot)$ is a dispersion density function. As in the literature, we restrict ourselves to the kernel $\kappa(O, \cdot)$ that is invariant under both translation and rotation, i.e., $\kappa(O, X) = \kappa(|X - O|)$ depends only on the euclidean distance $|X - O|$ of point X from the cluster center O, [23], [24]. Moreover, we assume that $\kappa(O, \cdot)$ is a summable, non-increasing, bounded and continuous function, and $\int_{\mathcal{O}} \kappa(O, X) dX = 1$. Following a common normalizing method, the kernel can be specified by first defining a non-increasing, bounded and continuous function $g(s)$ and then normalizing it over the area $\mathcal{O}$:

$$\kappa(O, X) = \frac{g(|X - O|)}{\int_{\mathcal{O}} g(|Y - O|) dY}.$$

Then, we can conclude that a center-clustering random model is determined by three factors/parameters, i.e., the number of all nodes n, the area of deployment region S, and the critical parameter of dispersion density function $g(\cdot)$. We denote a CCRM by $\mathcal{N}(n, S; g(\cdot))$ in this paper.

Specifically, we define $g(s) := \min\{1, s^{-\delta}\}$, where $\delta \in [0, \infty)$ is the *clustering exponent of node distribution*. Note that when $\delta = 0$, the model degenerates into the homogeneous *random extended network*, [6], [25].

We introduce an approximate real-world case of CCRM based on Google+ data set. Please see details in Appendix D.2, which can be found on the Computer Society Digital Library at http://doi.ieeecomputersociety.org/10.1109/TPDS.2014.2333524.

2.1.2 Extending Generality of Physical Layer Model

In this paper, we only study the capacity scaling laws under the CCRM with extended scaling pattern, where there is only one center point. We notice that based on the CCRM, one can develop more general physical layer models in terms of *clustering patterns* and *scaling patterns*. To smoothen the mainline of our work, we have moved the details of extended discussion to Section 5.2.

2.2 Social Layer Formation

We introduce a social formation model, called *population-based social model*. We will clarify the advantages of this model later in Section 2.2.2.

2.2.1 Population-Based Social Formation Model

Let $\mathcal{D}(u, r)$ denote the disk centered at a point u with radius r in the deployment region $\mathcal{O}$, and let $N(u, r)$ denote the number of nodes contained in $\mathcal{D}(u, r)$.

For a node $v_k \in \mathcal{V}$, construct its friendship set of $q_k, q_k \geq 1$, nodes/friends, say $\mathcal{F}_k$, by the following procedure:

1. *Zipf's degree distribution of social relations.* Assume that the number of friends (or followers) of a particular node $v_k \in \mathcal{V}$, denoted by q_k, follows a Zipf's distribution [26], i.e.,

$$\Pr(q_k = l) = \left(\sum_{j=1}^{n-1} j^{-\gamma} \right)^{-1} \cdot \; l^{-\gamma}. \tag{1}$$

From Eq. (1), we can observe that the degree distribution above depends on the specific network size (the number of users n). Please see a numerical validation based on Google+ data set for the Zipf's degree distribution in Appendix D.3, available in the online supplemental material.

2. *Population-based formation of social relations.* Making the position of node v_k as the *reference point*, choose q_k *points* independently on the torus region $\mathcal{O}$ according to a probability distribution with density function:

$$f_{v_k}(X) = \Phi_k(S, \beta) \cdot (\mathbf{E}[N(v_k, |X - v_k|)] + 1)^{-\beta}, \tag{2}$$

where the random variable $X := (x, y)$ denotes the position of a selected point in the deployment region, $|X - v_k|$ denotes the Euclidean distance between point X and node v_k, $\beta \in [0, \infty)$ represents the clustering exponent of friendship formation; the coefficient $\Phi_k(S, \beta) > 0$ depends on β and S (the area of deployment region), satisfying that:

$$\Phi_k(S, \beta) \cdot \int_{\mathcal{O}} (\mathbf{E}[N(v_k, |X - v_k|)] + 1)^{-\beta} dX = 1. \tag{3}$$

3. *Nearest-principle position of friends.* Let $\mathcal{A}_k = \{p_{k_i}\}_{i=1}^{q_k}$ denote the set of these q_k points. Let v_{k_i} be the nearest node to p_{k_i}, for $1 \leq i \leq q_k$ (ties are broken randomly). Denote the set of these q_k nodes by $\mathcal{F}_k = \{v_{k_i}\}_{i=1}^{q_k}$. Please see the illustration in Fig. 2. We call point p_{k_i} the *anchor point* of v_{k_i}, and define a set $\mathcal{P}_k := \{v_k\} \cup \mathcal{A}_k$.

Throughout this paper, we use $\mathbb{P}(\delta, \gamma, \beta)$ to denote the population-based social model.

2.2.2 Advantages of Population-Based Model

After Kleinberg [17] proposed a *distance-based* social model relating geographical distance and social friendship, Liben-Nowell et al. [20] introduced the *rank-based model*, where the probability of befriending a particular person is

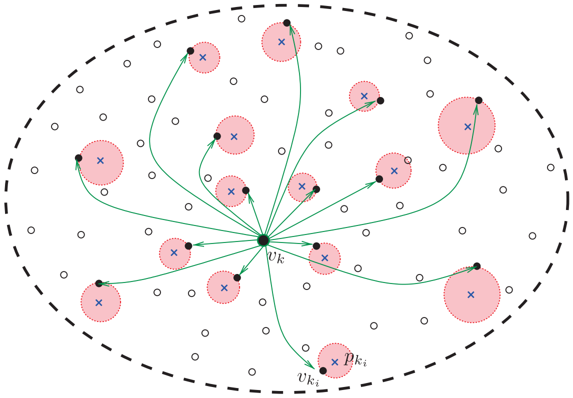


Fig. 2. Friendships of source v_k. Making the position of node v_k as the *reference point*, choose independently q_k *anchor points* on the torus region $\mathcal{O}$, denoted by p_{k_i} with $1 \leq i \leq q_k$, according to a probability density function as in Eq. (2). The q_k friends of v_k are located according to the positions of the corresponding anchor points.

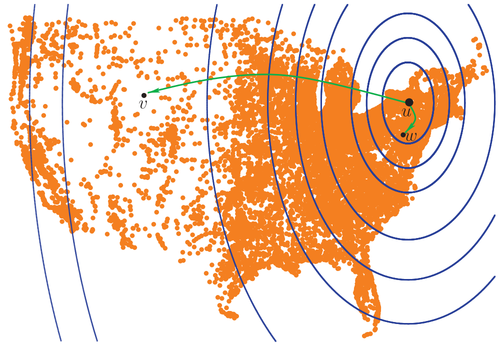


Fig. 3. Inhomogeneity of LiveJournal population [20]. A dot is shown for every distinct United States location home to at least one LiveJournal user (up to Feb. 2004). The population of each successive displayed circle (all centered on Ithaca, NY) increases by 50,000 people. The friendships of $u \rightarrow v$ and $u \rightarrow w$ are respectively *underestimated* and *overestimated* by the distance-based model.

inversely proportional to the power of the number of closer people. They validated the practicality of rank-based model by analyzing the data of an online social network, the LiveJournal online community. They pointed out that the weakness of distance-based models lies in that for a particular user, it underestimates the friendship probability of the distant nodes in the low-density region, when the geographical distribution of users is inhomogeneous in common occurrence, as illustrated in Fig. 3. In addition, we also give a validation of rank-based model based on Google+ data set in Appendix D.4, available in the online supplemental material.

The rank-based model states that the friendship probability depends on both the geographic distance and node density. Following this observation, by modifying the rank-based model, we propose the distance&density-aware population-based social model. We highlight that the population-based model is more convenient and systematic for addressing the issue of capacity scaling laws. Anchor points are usefully introduced, in order to ensure the independence of length of certain euclidean spanning trees, thus makes it convenient to bound the total length, e.g., the proof of Lemma 6. However, under the rank-based model where the friendships are directly built over nodes without anchor points, the corresponding independence cannot be guaranteed, which usually brings the difficulty on the theoretical rigor. The advantages of the point-based model for the basis and rigor in analysis, compared to the node-based model, had been apparent in [13], [21]

2.3 Session Layer Construction

After the social layer is formed, social sessions can be defined according to the specific applications: For the *social-unicast/social-multicast*, the source node delivers message to one/multiple selected friend(s).

For the *social-broadcast*, the source node broadcasts message to all its friends, such as tweets in Twitter and posts in Facebook. Accordingly, we can define other session patterns based on the definitions of corresponding non-social sessions, such as *social-anycast* [27] and *social-manycast* [28].

In this work, we mainly study social-broadcast sessions.

2.4 Communication Model

We mainly adopt the *generalized physical model* [6], [21] due to its generality and practicality compared to other models like the *protocol model* and *physical model* [13]. Let $\mathcal{L}_t$ denote a *scheduling set* of links in which all links can be scheduled simultaneously in time slot t; let $\alpha > 2$ denote the power attenuation exponent; let B and P denote the bandwidth and transmitting power, respectively.

Definition 1. *Under the generalized physical model, when a scheduling set $\mathcal{L}_t$ is scheduled, the rate of a link $< u, v > \in \mathcal{L}_t$ is achieved at*

$$R_{u,v;t} = B \times \mathbf{1} \cdot \{< u, v > \in \mathcal{L}_t\} \times \log\big(1 + SINR_{u,v;t}\big),$$

where $SINR_{u,v;t} = \frac{P \cdot \ell(|u-v|)}{N_0 + \sum_{<i,j> \in \mathcal{L}_t / <u,v>} P \cdot \ell(|i-v|)}$, N_0 denotes the background noise, $|u - v|$ represents the euclidean distance between nodes u and v; $\ell(\cdot)$ denotes the power attenuation function that is assumed to depend only on the distance between the transmitter and receiver, to be specific, $\ell(|\cdot|) := |\cdot|^{-\alpha}$ for dense networks, and $\ell(|\cdot|) := \min\{1, |\cdot|^{-\alpha}\}$ for extended networks.

2.5 Network Capacity for Social Sessions

Denote a session by $\mathcal{S}_k := \{v_k\} \cup \mathcal{D}_k$, where $\mathcal{D}_k \subseteq \mathcal{F}_k = \{v_{k_i}\}_{i=1}^{q_k}$ is the set of destinations of v_k. For social-broadcast sessions, it holds that $\mathcal{D}_k = \mathcal{F}_k$.

Let $\Lambda = (\lambda_1, \lambda_2, \ldots, \lambda_n)$ denote a *rate vector* of the data rate of all sessions. A rate vector Λ is *feasible* if there is a $T < \infty$ such that in every time interval (with unit seconds) $[(t-1) \cdot T, t \cdot T]$, every source node v_k can send $T \cdot \lambda_k$ bits to all its destinations. For a rate vector, we define the *per-session throughput* as $\Lambda(n) = \min_{v_k \in \mathcal{V}} \lambda_k$.

Definition 2 (Achievable Throughput). *We say a per-session throughput $\Lambda(n)$ is* achievable *for all social sessions if there is a feasible rate vector $\Lambda = (\lambda_1, \lambda_2, \ldots, \lambda_n)$ such that $\Lambda(n) = \min_{v_k \in \mathcal{V}} \lambda_k$.*

Definition 3 (Social Capacity). *The per-session capacity for a class of random networks is of order $\Theta(\Gamma(n))$ if there are constants $0 < \underline{c} < \overline{c} < +\infty$ such that*

$$\lim_{n \rightarrow +\infty} \Pr(\Lambda(n) = \underline{c} \cdot \Gamma(n) \textit{ is achievable}) = 1,$$
$$\liminf_{n \rightarrow +\infty} \Pr(\Lambda(n) = \overline{c} \cdot \Gamma(n) \textit{ is achievable}) < 1.$$

TABLE 1
Notations for Exponents

Notation	Definition
$\delta \in [0, \infty)$	clustering exponent of node distribution
$\gamma \in [0, \infty)$	clustering exponent of friendship degree
$\beta \in [0, \infty)$	clustering exponent of friendship formation
$\alpha \in (2, \infty)$	attenuation exponent of signal transmission

To facilitate the reader, we have reported in Table 1 a collection of frequently-used system parameters.

3 Preliminary Results for Distribution of Anchor Points

In this section, we mainly present some preliminary results for the anchor points distributed according to the population-based social formation scheme in the center-clustering random model. The results can serve as the basis for addressing the capacity of wireless social networks under the general CCRM and population-based social formation model.

In the center-clustering random model, say $\mathcal{N}(n, S; g(\cdot))$, we construct a set of $q+1$ *anchor points*, denoted by $\mathcal{P} = \{X_i\}_{i=0}^{q}$, by the following procedure: 1. Select arbitrarily a point from $\mathcal{O}$ as the first one in $\mathcal{P}$, denoted by X_0. 2. Making point X_0 as the *reference point* denoted by O', select independently other q points at random according to the probability distribution with density function as described in Eq.(2) of the population-based formation model (Let $v_k := X_0$).

Let $\mathcal{A} := \{X_i\}_{i=1}^{q}$. Next, we propose two theorems, and prove them in Section 3.3.

3.1 General Distribution Density Function

For the probability density function of point distribution, we have Theorem 1.

Theorem 1. *Making the point X_0 as the reference point O', the distribution of points in $\mathcal{A} = \{X_i\}_{i=1}^{q}$ follows the probability with density function*

$$f_{X_0}(X) = \frac{\left[\int_{\mathcal{D}(X_0,\, |X-X_0|)} \mathbf{d}(Y)dY + 1\right]^{-\beta}}{\int_{\mathcal{O}}\left[\int_{\mathcal{D}(X_0,\, |Z-X_0|)} \mathbf{d}(Y)dY + 1\right]^{-\beta} dZ}, \quad (4)$$

where $\mathbf{d}(Y) = n \cdot \frac{\min\{1,\, |Y-O|^{-\delta}\}}{\int_{\mathcal{O}} \min\{1,\, |Z-O|^{-\delta}\} dZ}$.

3.2 Euclidean Minimum Spanning Tree

For the Euclidean minimum spanning trees of $\mathcal{P}$ and $\mathcal{A}$, denoted by $\mathrm{EMST}(\mathcal{P})$ and $\mathrm{EMST}(\mathcal{A})$, respectively, we have Theorem 2.

Theorem 2. *As $q \to \infty$: With probability 1, it holds that*

$$|EMST(\mathcal{A})| = \Theta\left(\sqrt{q} \cdot \int_{\mathcal{O}} \sqrt{f_{X_0}(X)} dX\right); \quad (5)$$

with high probability $1 - o(1/\hat{N})$, it holds that

$$|EMST(\mathcal{P})| : \left[|EMST(\mathcal{A})|, |EMST(\mathcal{A})| + \overline{L}\right],^{1} \quad (6)$$

where $f_{X_0}(X)$ is defined in Theorem 1, and

$$\overline{L} = \min\left\{L \,\middle|\, \int_{\mathcal{D}(X_0,L)} f_{X_0}(X)dX = \Omega\left(\min\left\{\frac{\log \hat{N}}{q}, 1\right\}\right)\right\}, \quad (7)$$

with $\hat{N} : (1, n]$ is a given parameter.

Note that the parameter $\hat{N}$ can be defined as the number of nodes with degree of order $\omega(1)$.

3.3 Proof of Theorem 1 and Theorem 2

First of all, the cost of an edge (X_i, X_j) is given by $\Psi(|X_i - X_j|) = |X_i - X_j|$, that is, the exponent σ in Lemma A.1, available in the online supplemental material, equals 1. In addition, $\Psi(x)$ is a monotonically increasing function. Let L denote the distance between the center O and reference point. Then, under the center-clustering random model $\mathcal{N}(n, S; g(\cdot))$, by Eqs. (2) and (3), the density function is specified into Eq. (4). Then, by Lemma A.1, available in the online supplemental material, we get that

$$|\mathrm{EMST}(\mathcal{A})| = \Theta\left(\sqrt{q} \cdot \int_{\mathcal{O}} \sqrt{f_{X_0}(X)} dX\right).$$

It is straightforward that $|\mathrm{EMST}(\mathcal{P})| = \Omega(|\mathrm{EMST}(\mathcal{A})|)$. On the other hand, let $\underline{L}$ denote the smallest distance from the points in $\mathcal{A}$ to point X_0. Then,

$$\left(1 - \int_{\mathcal{D}(X_0,\underline{L})} f_{X_0}(X)dX\right)^{q} = o(1).$$

That is, $\int_{\mathcal{D}(X_0,\underline{L})} f_{X_0}(X)dX = \omega(1/q)$. Thus, $\underline{L} \le \overline{L}$, where $\overline{L}$ is defined in Eq. (7), which completes the proof. Note that we deliberately relax the upper bound of $\underline{L}$ as in Eq. (7) in order to ensure Eq. (6) to hold with uniformly high probability for $\Theta(n)$ Euclidean spanning trees, [29].

4 Social-Broadcast Capacity for Homogeneous Physical Layer

In this work, we specifically reduce the complexity from three dimensions $(\delta, \gamma, \beta) \in [0, \infty)^3$ to two dimensions $(\gamma, \beta) \in [0, \infty)^2$ by letting $\delta = 0$. In this case of extremely weak clustering behavior, the physical layer degenerates into the homogeneous random network model, [6], [13], [30], where $\mathbf{d}(Y) \equiv \Theta(1)$.

4.1 Main Results

4.1.1 Capacity under Generalized Physical Model

Theorem 3. *Under the population-based social model $\mathbb{P}(\delta = 0, \gamma, \beta)$ and the generalized physical model (GphyM) with $\alpha > 2$, the per-session social-broadcast capacity is of order Λ, where Λ is defined in Table 2.*

From Theorem 3, there are still gaps between upper and lower bounds on social-broadcast capacity under the

1. We use the term $f(n) : [\underline{\phi}(n), \overline{\phi}(n)]$ to represent $f(n) = \Omega(\underline{\phi}(n))$ and $f(n) = O(\overline{\phi}(n))$; and use $f(n) : (\underline{\phi}(n), \overline{\phi}(n))$ to represent $f(n) = \omega(\underline{\phi}(n))$ and $f(n) = o(\overline{\phi}(n))$.

TABLE 2
Social-Broadcast Capacity under GphyM

γ	Social Capacity under GphyM – Λ:
$\gamma > 2$	$\begin{cases} \Theta((\log n)^{-\frac{\alpha}{2}}), & \beta > 2; \\ \left[(\log n)^{-\frac{\alpha+1}{2}}, (\log n)^{-\frac{\alpha}{2}}\right], & \beta = 2; \\ \Theta(1/n^{1-\frac{\beta}{2}}), & 1 < \beta < 2; \\ \Theta(\sqrt{\log n}/\sqrt{n}), & \beta = 1; \\ \Theta(1/\sqrt{n}), & 0 \le \beta < 1. \end{cases}$
$\gamma = 2$	$\begin{cases} \left[(\log n)^{-\frac{\alpha+3}{2}}, (\log n)^{-\frac{\alpha}{2}-1}\right], & \beta \ge 2; \\ \Theta(1/n^{1-\frac{\beta}{2}}), & 1 < \beta < 2; \\ \Theta(\sqrt{\log n}/\sqrt{n}), & \beta = 1; \\ \Theta(1/\sqrt{n}), & 0 \le \beta < 1. \end{cases}$
$3/2 < \gamma < 2$	$\begin{cases} \Theta((\log n)^{-\frac{\alpha}{2}}/n^{2-\gamma}), & \beta \ge 2\gamma - 2; \\ \Theta(1/n^{1-\frac{\beta}{2}}), & 1 < \beta < 2\gamma - 2; \\ \Theta(\sqrt{\log n}/\sqrt{n}), & \beta = 1; \\ \Theta(1/\sqrt{n}), & 0 \le \beta < 1. \end{cases}$
$\gamma = 3/2$	$\begin{cases} \Theta((\log n)^{-\frac{\alpha}{2}}/\sqrt{n}), & \beta \ge 1; \\ \left[\frac{(\log n)^{-\frac{\alpha+1}{2}}}{\sqrt{n}}, \frac{(\log n)^{-\frac{\alpha}{2}}}{\sqrt{n}}\right], & 0 \le \beta < 1. \end{cases}$
$1 < \gamma < 3/2$	$\Theta((\log n)^{-\frac{\alpha}{2}}/n^{2-\gamma})$
$\gamma = 1$	$\left[(\log n)^{\frac{1-\alpha}{2}}/n, (\log n)^{1-\frac{\alpha}{2}}/n\right]$
$0 \le \gamma < 1$	$\Theta((\log n)^{-\frac{\alpha}{2}}/n)$

generalized physical model in four regimes. As illustrated in Fig. 4, these four regimes are indeed lines in the two-dimensional parameter space $(\gamma, \beta) \in [0, \infty)^2$. A challenging issue is to close those gaps by presenting possibly new tighter upper and lower bounds by using some new arguments or designing new schemes.

4.1.2 Capacity under Protocol Model

We concentrate on deriving the capacity under the generalized physical model, while for completeness, we also include the results on capacity under the well-known protocol model (ProM, [13]). Based on our results on EMSTs and ESTs of social-broadcast sessions (Lemmas 6 and 9), using the analytical methods for capacity under the protocol model in [13], [22], we get that

Theorem 4. *Under the population-based social model $\mathbb{P}(\delta = 0, \gamma, \beta)$ and the ProM, the per-session social-broadcast capacity is of order Λ^{Pro} as defined in Table 3.*

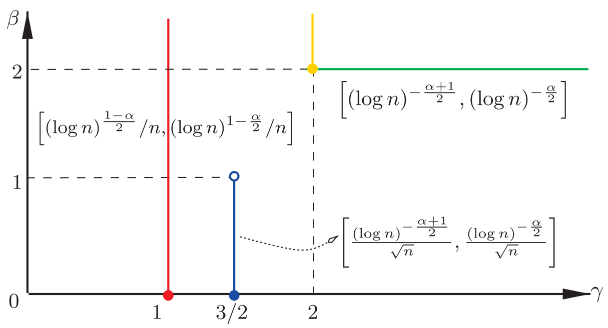


Fig. 4. Four regimes/lines where gaps exist.

TABLE 3
Social-Broadcast Capacity under ProM: Λ^{Pro}

γ	Social Capacity under ProM – Λ^{Pro}
$\gamma > 2$	$\begin{cases} \Theta(1/\log n), & \beta > 2; \\ \Theta(1/(\log n)^{\frac{3}{2}}), & \beta = 2; \\ \Theta(n^{\frac{\beta}{2}-1}/\sqrt{\log n}), & 1 < \beta < 2; \\ \Theta(1/\sqrt{n}), & \beta = 1; \\ \Theta(1/\sqrt{n \log n}), & 0 \le \beta < 1. \end{cases}$
$\gamma = 2$	$\begin{cases} \Theta(1/(\log n)^3), & \beta \ge 2; \\ \Theta(n^{\frac{\beta}{2}-1}/\sqrt{\log n}), & 1 < \beta < 2; \\ \Theta(1/\sqrt{n}), & \beta = 1; \\ \Theta(1/\sqrt{n \log n}), & 0 \le \beta < 1. \end{cases}$
$3/2 < \gamma < 2$	$\begin{cases} \Theta(n^{\gamma-2}/\log n), & \beta \ge 2\gamma - 2; \\ \Theta(n^{\frac{\beta}{2}-1}/\sqrt{\log n}), & 1 < \beta < 2\gamma - 2; \\ \Theta(1/\sqrt{n}), & \beta = 1; \\ \Theta(1/\sqrt{n \log n}), & 0 \le \beta < 1. \end{cases}$
$\gamma = 3/2$	$\begin{cases} \Theta(1/(\log n \cdot \sqrt{n})), & \beta \ge 1; \\ \Theta(1/(\log n \cdot \sqrt{n \log n})), & 0 \le \beta < 1. \end{cases}$
$1 < \gamma < 3/2$	$\Theta(1/(\log n \cdot n^{2-\gamma}))$
$0 \le \gamma \le 1$	$\Theta(1/n)$

4.1.3 Intuitions of Main Results

At first, we discuss the impacts of clustering exponents of friendship degree and friendship formation, i.e., γ and β, on the social-broadcast capacity. We provide an illustration for the protocol model in Fig. 5 according to the results of Theorem 4, and we omit the counterpart for the generalized physical model due to it's similarity of change trend. From Fig. 5, the social-broadcast capacity is monotonically nondecreasing in the range $[1/\log n, 1/n]$ for both γ and β. An intuitive explanation can be made as follows: A larger clustering exponent of friendship degree γ can limit the number of friends of each user into a smaller upper bound with high probability, then leads to a larger social-broadcast capacity (Definition 3); a larger clustering exponent of friendship formation β makes the friends more closer to each user with high

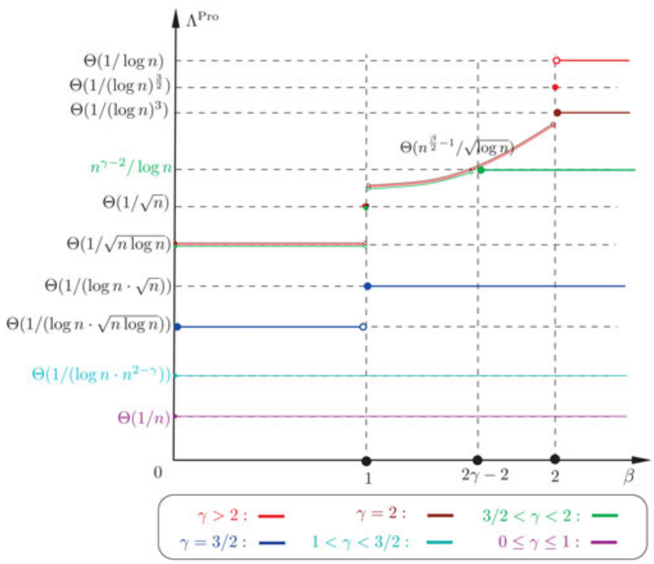


Fig. 5. The impacts of γ and β on the capacity under protocol model.

probability, then possibly reduces the total transmission distance of each social-broadcast session, finally also leads to a larger social-broadcast capacity.

Importantly, we notice that we only take account of the case that $\delta = 0$, where the population-based model degenerates to that similar to distance-based model [17], [31]. The advantages of population-based model cannot be sufficiently highlighted for such a special model, indeed. This work can act as the first step of investigating the capacity under the general population-based model. It would be a significant future work to clarify the relationships between the general clustering exponent and capacity in 3-dimensional parameter space, i.e., $(\delta, \gamma, \beta) \in [0, \infty)^3$. For such future work, we make a more detailed discussion in Section 5.1 on how to derive the capacity under more general models based on the results in this paper.

4.1.4 Differences from Existing Work

As stated above, the distance-based model can be regarded as a special case of the proposed population-based model. In this work, we mainly give the complete results on capacity scaling laws for this special case. Azimdoost et al. [18], [19] had also introduced the distance-based social model into the study of wireless social networks. Naturally, it is necessary to declare the differences between [18], [19] and our results, and highlight the advantages of our work compared with those existing works. We can summarize the differences from three aspects: The first is the different generality of the studied session patterns. Both [18] and [19] focused on the simple *unicast* capacity by choosing randomly one destination node for each session from the social group (the set of friends) of the source node. They did not give sufficient consideration to the scale-free feature of social relationship distribution and the diversity of session patterns in social applications. While, to address this problem, we investigate the capacity for the social-broadcast sessions where the number of destinations in each session is assumed to follow a Zipf's distribution [26]. The second is the different practicality of the adopted communication models. As two representative communication models, the protocol model and generalized physical model have been widely used in addressing the issue of capacity scaling laws. The former is convenient analytically, while the latter can capture the nature of wireless channels better, then can derive more practical results on network capacity. Both [18] and [19] only investigated the simple protocol model. In our paper, we mainly aim to present more practical results by adopting the generalized physical model, while for completeness, we also include the results under the protocol model. The third is the different scalability of the proposed methods and results. Although only the results on the special case are completely derived in this paper, some proposed analytical methods and results can serve as the basis of addressing network capacity under the general population-based social model with inhomogeneous physical layer. However, the results and methods in [18] and [19] cannot be extended into ones in line with reality for the general physical layer due to the incurable defect of the distance-based model.

4.2 System Setting

4.2.1 Degree Distribution of Social Relationships

Considering the degree distribution, by Eq. (1), we get that

$$\Pr(q_k = l) = \begin{cases} \Theta(l^{-\gamma}), & \gamma > 1; \\ \Theta\left(\frac{1}{\log n} \cdot l^{-1}\right), & \gamma = 1; \\ \Theta(n^{\gamma-1} \cdot l^{-\gamma}), & 0 \leq \gamma < 1. \end{cases} \tag{8}$$

4.2.2 Distribution of Anchor Points

For each session $\mathcal{S}_k$ initiated by the source v_k, we can get the distribution of anchor points directly using Theorem 1,

Lemma 1. *When the clustering exponent $\delta = 0$, for a session $\mathcal{S}_k$ under the population-based social model $\mathbb{P}(\delta = 0, \gamma, \beta)$, the anchor points of the friends of source v_k follows the distribution of density function:*

$$f_{v_k}(X) = \begin{cases} \Theta((|X - v_k|^2 + 1)^{-\beta}), & \beta > 1; \\ \Theta(\frac{1}{\log n} \cdot (|X - v_k|^2 + 1)^{-1}), & \beta = 1; \\ \Theta(n^{\beta-1} \cdot (|X - v_k|^2 + 1)^{-\beta}), & 0 \leq \beta < 1. \end{cases}$$

By using Lemma 1, we can get the following result.

Lemma 2. *For a social-broadcast session $\mathcal{S}_k$ under the model $\mathbb{P}(\delta = 0, \gamma, \beta)$, it holds that:*

$$\mathbf{E}[|X - v_k|] = \begin{cases} \Theta(1), & \beta > 3/2; 1 \\ \Theta(\log n), & \beta = 3/2; \\ \Theta\left(n^{\frac{3}{2}-\beta}\right), & 1 < \beta < 3/2; \\ \Theta(\sqrt{n}/\log n), & \beta = 1; \\ \Theta(\sqrt{n}), & 0 \leq \beta < 1. \end{cases} \tag{9}$$

4.2.3 Social-Broadcast Sessions

Under the population-based social model, we denote a social-broadcast session by a set $\mathcal{S}_k := \{v_k\} \cup \mathcal{F}_k$, where v_k denotes the source node and each element in $\mathcal{F}_k = \{v_{k_i}\}_{i=1}^{q_k}$, say v_{k_i}, is the nearest *node* to the corresponding *anchor point* p_{k_i} in $\mathcal{A}_k = \{p_{k_i}\}_{i=1}^{q_k}$. Please see the illustration in Fig. 2. Recall that $\mathcal{P}_k = \{v_k\} \cup \mathcal{A}_k$, we get the following Lemma 3 for spanning trees over $\mathcal{S}_k$.

Lemma 3. *For a social-broadcast session $\mathcal{S}_k$ with $q_k = \omega(1)$ under the social model $\mathbb{P}(\delta = 0, \gamma, \beta)$, with probability 1, it holds that $|EMST(\mathcal{A}_k)| = \Theta(L_{\mathcal{P}}(\beta, q_k))$, and then $|EMST(\mathcal{P}_k)| = \Omega(L_{\mathcal{P}}(\beta, q_k))$, where*

$$L_{\mathcal{P}}(\beta, q_k) = \begin{cases} \Theta\left(\sqrt{q_k}\right), & \beta > 2; \\ \Theta\left(\sqrt{q_k} \cdot \log n\right), & \beta = 2; \\ \Theta\left(\sqrt{q_k} \cdot n^{1-\frac{\beta}{2}}\right), & 1 < \beta < 2; \\ \Theta\left(\sqrt{q_k} \cdot \sqrt{\frac{n}{\log n}}\right), & \beta = 1; \\ \Theta\left(\sqrt{q_k} \cdot \sqrt{n}\right), & 0 \leq \beta < 1. \end{cases} \tag{10}$$

Proof. From Theorem 2, it follows that with probability 1, $|\mathrm{EMST}(\mathcal{A}_k)| = \Theta(L_{\mathcal{P}}(\beta, q_k))$ for $q_k = \omega(1)$, where $L_{\mathcal{P}}(\beta, q_k)$ is defined in Eq. (10). Combining with the fact that $|\mathrm{EMST}(\mathcal{P}_k)| \geq |\mathrm{EMST}(\mathcal{A}_k)|$, we get the lemma. □

TABLE 4
$H(\gamma,\beta)$ in Bounding $\sum$ EMST and $\sum$ EST

γ	$H(\gamma,\beta)$
$\gamma > 2]$	$\begin{cases} \Theta(n), & \beta > 2; \\ \Theta(n\cdot\log n), & \beta = 2; \\ \Theta(n^{2-\frac{\beta}{2}}), & 1 < \beta < 2; \\ \Theta(n^{3/2}/\sqrt{\log n}), & \beta = 1; \\ \Theta(n^{3/2}), & 0 \le \beta < 1. \end{cases}$
$\gamma = 2$	$\begin{cases} \Theta(n\cdot\log n), & \beta \ge 2; \\ \Theta(n^{2-\frac{\beta}{2}}), & 1 < \beta < 2; \\ \Theta(n^{3/2}/\sqrt{\log n}), & \beta = 1; \\ \Theta(n^{3/2}), & 0 \le \beta < 1. \end{cases}$
$3/2 < \gamma < 2$	$\begin{cases} \Theta(n^{3-\gamma}), & \beta \ge 2\gamma-2; \\ \Theta(n^{2-\frac{\beta}{2}}), & 1 < \beta < 2\gamma-2; \\ \Theta(n^{3/2}/\sqrt{\log n}), & \beta = 1; \\ \Theta(n^{3/2}), & 0 \le \beta < 1. \end{cases}$
$\gamma = 3/2$	$\begin{cases} \Theta(n^{3/2}), & \beta > 1; \\ \Theta(n^{3/2}\cdot\sqrt{\log n}), & \beta = 1; \\ \Theta(n^{3/2}\cdot\log n), & 0 \le \beta < 1. \end{cases}$
$1 < \gamma < 3/2$	$\Theta(n^{3-\gamma})$
$\gamma = 1$	$\Theta(n^2/\log n)$
$0 \le \gamma < 1$	$\Theta(n^2)$

4.3 Upper Bounds on Social-Broadcast Capacity

4.3.1 Technical Preparations for Upper Bounds

Before computing the upper bounds, we introduce a notion called *lattice view*.

Definition 4 (Lattice View). *Partition a square deployment region $\mathcal{O}(S) = [0, \sqrt{S}]^2$ into $\lceil \sqrt{S}/c \rceil^2$ cells of side length $c : [\sqrt{S/n}, \sqrt{S})$, we call the produced lattice graph* lattice view, *and denote it by $\mathbb{V}(\sqrt{S}, c)$.*

Based on a given lattice view $\mathbb{V}(\sqrt{S}, c)$, we can get the following lemma for arbitrary routing trees for a session $\mathcal{S}_k$. By [25, Lemma 2], we have

Lemma 4. *Given a social-broadcast session $\mathcal{S}_k$, let $\mathcal{T}_k$ be a routing tree for $\mathcal{S}_k$, and let $N(\mathcal{T}_k, \sqrt{S}, c)$ denote the number of cells used by $\mathcal{T}_k$ in $\mathbb{V}(\sqrt{S}, c)$, then when $q_k = O(S/c^2)$, it holds that $N(\mathcal{T}_k, \sqrt{S}, c) = \Omega\big(\frac{1}{c}\cdot|EMST(\mathcal{S}_k)|\big)$.*

In $\mathbb{V}(\sqrt{S}, c)$, a cell is called an *island* if it contains $\Theta(\frac{nc^2}{S})$ nodes and all its eight neighbor cells are empty.

Lemma 5 ([25]). *There exists w.h.p. an island in the lattice view $\mathbb{V}(\sqrt{S}, c)$, if $c \le \frac{1}{2}\cdot\sqrt{\frac{(1-\epsilon)\cdot S\cdot\log n}{2n}}$, where $\epsilon \in (0,1)$ is constant.*

Since the sum of length of Euclidean minimum spanning trees for all n sessions, i.e., $\sum_{k=1}^{n}|\mathrm{EMST}(\mathcal{S}_k)|$, plays a key role in the analysis of network capacity, we need to give the lower bounds on $\sum_{k=1}^{n}|\mathrm{EMST}(\mathcal{S}_k)|$. Note that the bounds depend on those on $\sum_{k=1}^{n}|\mathrm{EMST}(\mathcal{P}_k)|$, which will be provided in Lemma B.1, available in the online supplemental material.

Lemma 6. *For all social-broadcast sessions $\mathcal{S}_k$, $k = 1, 2, \ldots, n$, under the social model $\mathbb{P}(\delta = 0, \gamma, \beta)$, with high probability, $\sum_{k=1}^{n}|EMST(\mathcal{S}_k)| = \Omega(H(\gamma,\beta))$, where $H(\gamma,\beta)$ is described in Table 4.*

Proof. Please see the proof in Appendix B.1, available in the online supplemental material. □

4.3.2 Upper Bounds on Social-Broadcast Capacity

We will derive the upper bounds by combining the bounds based on the lattice views $\mathbb{V}(\sqrt{n}, \sqrt{2})$ and $\mathbb{V}(\sqrt{n}, \frac{1}{3}\sqrt{\log n})$.

Upper Bound from Lattice View $\mathbb{V}(\sqrt{n}, \sqrt{2})$. Based on Lemma 6, we can get the following lemma.

Lemma 7. *Under the social model $\mathbb{P}(\delta = 0, \gamma, \beta)$ and the generalized physical model with $\alpha > 2$, the per-session social-broadcast capacity is of order $O(n/H(\gamma,\beta))$, where $H(\gamma,\beta)$ is defined in Table 4.*

Proof. Please see the proof in Appendix B.2, available in the online supplemental material. □

Upper bound from lattice view $\mathbb{V}(\sqrt{n}, \frac{1}{3}\sqrt{\log n})$. From Lemma 5, for $\epsilon = \frac{1}{9}$, there is an island in the lattice view $\mathbb{V}(\sqrt{n}, \frac{1}{3}\sqrt{\log n})$. Thus, we get the following lemma.

Lemma 8. *Under the social model $\mathbb{P}(\delta = 0, \gamma, \beta)$ and the generalized physical model with $\alpha > 2$, the per-session social-broadcast capacity is of order*

$$\Lambda = \begin{cases} O\left((\log n)^{-\frac{\alpha}{2}}\right), & \gamma > 2; \\ O\left((\log n)^{-\frac{\alpha}{2}-1}\right), & \gamma = 2; \\ O\left(n^{\gamma-2}\cdot(\log n)^{-\frac{\alpha}{2}}\right), & 1 < \gamma < 2; \\ O\left((\log n)^{1-\frac{\alpha}{2}}/n\right), & \gamma = 1; \\ O\left((\log n)^{-\frac{\alpha}{2}}/n\right), & 0 \le \gamma < 1. \end{cases}$$

Proof. Please see the proof in Appendix B.3, available in the online supplemental material. □

Combination of upper bounds. Combining Lemmas 7 and 8, we obtain the upper bounds in Theorem 3 by performing some simple algebraic manipulations.

4.4 Lower Bounds on Social-Broadcast Capacity

n this section, we present the constructive lower bounds for the social-broadcast capacity by devising two social-broadcast strategies having their own merits respectively. Both strategies depend on the euclidean spanning trees of sessions.

4.4.1 Euclidean Spanning Trees

Recall that we denote each social-broadcast session by $\mathcal{S}_k = \{v_k\} \cup \mathcal{F}_k$, where v_k is the source node and $\mathcal{F}_k$ is the set of friends (or followers) of v_k. $\mathcal{A}_k = \{p_{k_i}\}_{i=1}^{q_k}$ denotes the set of these q_k anchor points of the nodes in $\mathcal{F}_k = \{v_{k_i}\}_{i=1}^{q_k}$.

Construction of Euclidean spanning trees. For each session $\mathcal{S}_k = \{v_k\} \cup \mathcal{F}_k$, we build an EST, denoted as $\mathrm{EST}(\mathcal{S}_k)$, by the following method (*Anchors-EMST-Based Greedy* (AEBG) Algorithm):

Step 1. Construct a Euclidean minimum spanning tree (EMST) based on $\mathcal{A}_k$, denoted as $\mathrm{EMST}(\mathcal{A}_k)$, using some classic greedy algorithms like Prim algorithm.

TABLE 5
Achievable Throughput under $\mathbb{B}_{\mathrm{p\&h}}$: $\underline{\Lambda}^{\mathbb{B}_{\mathrm{p\&h}}}$

γ	$\underline{\Lambda}^{\mathbb{B}_{p\&h}}$
$\gamma < 2$	$\begin{cases} \Omega((\log n)^{-\frac{\alpha+1}{2}}), & \beta \geq 2; \\ \Omega(1/n^{1-\frac{\beta}{2}}), & 1 < \beta < 2; \\ \Omega(\sqrt{\log n}/\sqrt{n}), & \beta = 1; \\ \Omega(1/\sqrt{n}), & 0 \leq \beta < 1. \end{cases}$
$\gamma = 2$	$\begin{cases} \Omega((\log n)^{-\frac{\alpha+3}{2}}), & \beta \geq 2; \\ \Omega(1/n^{1-\frac{\beta}{2}}), & 1 < \beta < 2; \\ \Omega(\sqrt{\log n}/\sqrt{n}), & \beta = 1; \\ \Omega(1/\sqrt{n}), & 0 \leq \beta < 1. \end{cases}$
$3/2 < \gamma < 2$	$\begin{cases} \Omega(\frac{(\log n)^{-\frac{\alpha+1}{2}}}{n^{2-\gamma}}), & \beta \geq 2\gamma - 2; \\ \Omega(1/n^{1-\frac{\beta}{2}}), & 1 < \beta < 2\gamma - 2; \\ \Omega(\sqrt{\log n}/\sqrt{n}), & \beta = 1; \\ \Omega(1/\sqrt{n}), & 0 \leq \beta < 1. \end{cases}$
$1 < \gamma \leq 3/2$	$\Omega((\log n)^{-\frac{\alpha+1}{2}}/n^{2-\gamma})$
$\gamma = 1$	$\Omega((\log n)^{\frac{1-\alpha}{2}}/n)$
$0 \leq \gamma < 1$	$\Omega((\log n)^{-\frac{\alpha}{2}}/n)$

TABLE 6
Achievable Throughput under $\mathbb{B}_{\mathrm{p}}$: $\underline{\Lambda}^{\mathbb{B}_{\mathrm{p}}}$

γ	$\underline{\Lambda}^{\mathbb{B}_p}$
$\gamma > 2$	$\begin{cases} \Omega((\log n)^{-\frac{\alpha}{2}}), & \beta > 2; \\ \Omega((\log n)^{-\frac{\alpha+1}{2}}), & \beta = 2; \\ \Omega((\log n)^{\frac{1-\alpha}{2}}/n^{1-\frac{\beta}{2}}), & 1 < \beta < 2; \\ \Omega((\log n)^{\frac{2-\alpha}{2}}/\sqrt{n}), & \beta = 1; \\ \Omega((\log n)^{\frac{1-\alpha}{2}}/\sqrt{n}), & 0 \leq \beta < 1. \end{cases}$
$\gamma = 2$	$\begin{cases} \Omega((\log n)^{-\frac{\alpha}{2}-2}), & \beta \geq 2; \\ \Omega((\log n)^{\frac{1-\alpha}{2}}/n^{1-\frac{\beta}{2}}), & 1 < \beta < 2; \\ \Omega((\log n)^{\frac{2-\alpha}{2}}/\sqrt{n}), & \beta = 1; \\ \Omega((\log n)^{\frac{1-\alpha}{2}}/\sqrt{n}), & 0 \leq \beta < 1. \end{cases}$
$3/2 < \gamma < 2$	$\begin{cases} \Omega((\log n)^{-\frac{\alpha}{2}}/n^{2-\gamma}), & \beta \geq 2\gamma - 2; \\ \Omega((\log n)^{\frac{1-\alpha}{2}}/n^{1-\frac{\beta}{2}}), & 1 < \beta < 2\gamma - 2; \\ \Omega((\log n)^{\frac{2-\alpha}{2}}/\sqrt{n}), & \beta = 1; \\ \Omega((\log n)^{\frac{1-\alpha}{2}}/\sqrt{n}), & 0 \leq \beta < 1. \end{cases}$
$\gamma = 3/2$	$\begin{cases} \Omega((\log n)^{-\frac{\alpha}{2}}/\sqrt{n}), & \beta \geq 1; \\ \Omega((\log n)^{-\frac{\alpha+1}{2}}/\sqrt{n}), & 0 \leq \beta < 1. \end{cases}$
$1 \leq \gamma < 3/2$	$\Omega((\log n)^{-\frac{\alpha}{2}}/n^{2-\gamma})$
$0 \leq \gamma < 1$	$\Omega((\log n)^{-\frac{\alpha}{2}}/n)$

Step 2. Connect the pairs v_{k_i} and v_{k_j} if and only if the link $p_{k_i}p_{k_j} \in \mathrm{EMST}(\mathcal{A}_k)$. Then, one can obtain an EST of $\mathcal{F}_k$.

Step 3. Connect the source v_k to its nearest node in $\mathcal{F}_k$ to get the final EST of $\mathcal{S}_k$, i.e., $\mathrm{EST}(\mathcal{S}_k)$.

Bounds on $\sum_{k=1}^{n} |\mathrm{EST}(\mathcal{S}_k)|$. We give the upper bounds on $\sum_{k=1}^{n} |\mathrm{EST}(\mathcal{S}_k)|$.

Lemma 9. *For all social-broadcast sessions $\mathcal{S}_k$, $k = 1, 2, \ldots, n$, under the social model $\mathbb{P}(\delta = 0, \gamma, \beta)$, using the aforementioned AEBG algorithm, with high probability, it holds that $\sum_{k=1}^{n} |\mathrm{EST}(\mathcal{S}_k)| = O(H(\gamma, \beta))$, where $H(\gamma, \beta)$ is described in Table 4.*

Proof. Please see the proof in Appendix B.4, available in the online supplemental material. □

Combining Lemmas 6 and 9, we get that

Theorem 5. *For all social-broadcast sessions $\mathcal{S}_k$, under the model $\mathbb{P}(\delta = 0, \gamma, \beta)$, it holds that with high probability, $\sum_{k=1}^{n} |EMST(\mathcal{S}_k)| = \Theta(H(\gamma, \beta))$, where $H(\gamma, \beta)$ is described in Table 4.*

4.4.2 Social-Broadcast Schemes

Since the physical layer with $\delta = 0$ is a *random extended network*, the percolation-based routing backbone [6] still applies to the social model $\mathbb{P}(\delta = 0, \gamma, \beta)$. We introduce two routing backbones called *highways* [6] and *parallel arterial roads* [30] to design the social-broadcast schemes. For the sake of completeness, we include concisely the construction procedures of these backbone systems in Appendix C.1, available in the online supplemental material. We propose two types of social-broadcast schemes: The former's hierarchical routing backbone consists of *highways* and *parallel arterial roads*, denoted by $\mathbb{B}_{\mathrm{p\&h}}$; the latter is only based on *parallel arterial roads*, denoted by $\mathbb{B}_{\mathrm{p}}$. Due to the relatively insignificant novelty, we move the complete descriptions of these schemes into Appendix C.2, available in the online supplemental material.

Next, we give a theorem to present the achievable social-broadcast throughputs under $\mathbb{B}_{\mathrm{p\&h}}$ and $\mathbb{B}_{\mathrm{p}}$.

Theorem 6. *Under the social model $\mathbb{P}(\delta = 0, \gamma, \beta)$ and generalized physical model with $\alpha > 2$, using $EST(\mathcal{S}_k)$ derived by AEBG algorithm as the input of schemes $\mathbb{B}_{\mathrm{p\&h}}$ or $\mathbb{B}_{\mathrm{p}}$, then it holds that with high probability:*

▷ Under scheme $\mathbb{B}_{\mathrm{p\&h}}$, the achievable throughput, denoted by $\underline{\Lambda}^{\mathbb{B}_{\mathrm{p\&h}}}$, is described in Table 5.

▷ Under scheme $\mathbb{B}_{\mathrm{p}}$, the achievable throughput, denoted by $\underline{\Lambda}^{\mathbb{B}_{\mathrm{p}}}$, is described in Table 6.

Proof. Please see the proof in Appendix C.3, available in the online supplemental material. □

Combining the throughputs under schemes $\mathbb{B}_{\mathrm{p\&h}}$ and $\mathbb{B}_{\mathrm{p}}$ in Theorem 6, we get the lower bounds in Theorem 3.

5 Conclusion and Future Work

In this paper, we mainly address capacity scaling laws of wireless social networks under the social-based session formation scheme. A three-layered model is proposed for abstracting wireless social networks. As one of main contributions, a cross-layer and distance&density-aware social model is proposed, which captures the formation characteristics of real-world social networks better, and specializes in the analysis of capacity scaling laws. We derive the social-broadcast capacity, taking into account the general clustering exponents of both friendship degree and friendship formation. Moreover, we present the density function of general social friendships distribution that will be the basis for investigating the capacity of general wireless social networks. This work can act as the first step of investigating the capacity under the proposed population-based model.

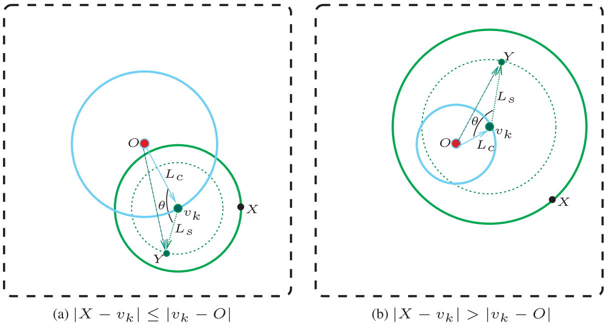


Fig. 6. Illustrations of positions of the center point O, the reference point v_k, the targeted point X, and any interior point Y in $\mathcal{D}(v_k, |X - v_k|)$.

Next, we make a discussion on how to derive the capacity under more general models based on the results of this work.

5.1 Extending to General Physical Clustering Case

The advantages of population-based model cannot be sufficiently highlighted for the case that $\delta = 0$, indeed. It would be a significant future work to investigate the relationships between the general clustering exponent and capacity in 3-dimensional parameter space, i.e., $(\delta, \gamma, \beta) \in [0, \infty)^3$.

Here, we probe the feasibility of studying network capacity under the social model with inhomogeneous physical layer, by extending the proposed basic theorem (Theorem 1) for the general distribution of anchor points. The key factor is to determine $\mathbf{d}(Y)$. Furthermore, since $\mathbf{d}(Y) = \frac{n \cdot \min\{1, |Y-O|^{-\delta}\}}{\int_{\mathcal{O}} \min\{1, |Z-O|^{-\delta}\} dZ}$, then it should be the first key step to deal with $|Y - O|$ based on the "known" *distance deviating from the source* $L_s = |Y - v_k|$ and *distance deviating from the center* $L_c = |v_k - O|$. As illustrated in Fig. 6, it follows that

$$|Y - O| = \sqrt{L_c^2 + L_s^2 - 2L_c \cdot L_s \cdot \cos\theta},$$

which can serve as a basis for the further study on the general model.

5.2 Extending to General Physical Layer

The proposed center-clustering random model can be developed to more general physical layer models in terms of *clustering patterns* and *scaling patterns*.

5.2.1 Multiple Clustering Centers

In the CCRM, there is only one center. Although the results under this model also apply to the model with physical layer model having a finite number of centers due to the characteristic of capacity scaling issues, more realistic and general clustering behaviors in real-life networks cannot be fully embodied by this simple model. To model the clustering behavior of node distribution with $\omega(1)$ centers in real-life applications of wireless networks, the shotnoise Cox process (SNCP, [23], [32]) can be introduced, where M center points generate their respective point processes, and the conditional local intensity at X is determined by the superposition of the individual processes, i.e., $\mathbf{d}(X) = \sum_j \rho_j \cdot \kappa(c_j, X)$, Then, the node process forms an inhomogeneous Poisson point process.

5.2.2 General Scaling Models

In the research of network capacity scaling laws, there are two typical models in terms of scaling patterns of network: *dense scaling model* and *extended scaling model* [6], [30], [33]. They have different engineering implications related to the classical notions of *interference-limitedness* and *coverage-limitedness* [33]: The former is only interference-limited; while, the latter is both interference-limited and coverage-limited. In the CCRM adopted in this paper, the area of deployment region is $S := n$, which implies that it is of *extended scaling pattern*. Then, it is necessary to extend the CCRM to one with general scaling pattern by setting a general area of deployment region. Such a work will directly involve defining the dispersion density function $g(\cdot)$ and power attenuation function $\ell(\cdot)$ which might have a significant impact on the network capacity. Unlike the setting for extended scaling models, these two functions for dense scaling models are usually defined as: $g(s) := s^{-\delta}$ [23] and $\ell(s) := s^{-\alpha}$ [6], respectively.

5.3 Extending to General Underlying Networks

For addressing the performance of data transmission in underlying communication networks for social networking services, besides the architecture of specific SNSs, there is another key factor to consider. It is the architecture of underlying communication network, e.g., wireless, wired, or hybrid networking architectures. This has a significant impact on implementing routing for specific data dissemination sessions, and plays a key role in limiting network capacity. In this paper, we focus on the scenario where social networking applications operate on the underlying wireless ad hoc networks, and aim to investigate the capacity scaling laws of a large-scale ad hoc network when it undertakes the data transmission of social applications. However, we notice that the real-world underlying network for social networking applications is most unlikely a pure wireless ad hoc network. A realistic and effective underlying network might be a hybrid network consisting of the Internet and different types of wireless networks, including static wireless and mobile networks. Therefore, it will be a very challenging but significant future work to extend the study by considering the diversity of real-world underlying networks, based on the preliminary results on wireless ad hoc networks in this work. For example, the next feasible and necessary extension could be addressing the fundamental limits of mobile social ad hoc networks where the underlying communication networks can be modeled as mobile ad hoc networks.

ACKNOWLEDGMENTS

The research of authors was partially supported by the National Natural Science Foundation of China (NSFC) under grant No. 61202383, the National Basic Research Program of China (973 Program) under grant No. 2010CB328101, Shanghai Rising-Star Program under grant No. 14QA1403700, the Program for New Century Excellent Talents in University (NCET) under grant No. NCET-12-0414, the Natural Science Foundation of Shanghai under grant No. 12ZR1451200, the Integrated Project for Major Research Plan of the National Natural Science

Foundation of China under grant No. 91218301, and the Research Fund for the Doctoral Program of Higher Education of China (RFDP) under grant No. 20120072120075.

REFERENCES

[1] Y. Ahn, S. Han, H. Kwak, S. Moon, and H. Jeong, "Analysis of topological characteristics of huge online social networking services," in *Proc. 16th ACM Int. Conf. World Wide Web*, 2007, pp. 835–844.

[2] H. Kwak, C. Lee, H. Park, and S. Moon, "What is twitter, a social network or a news media?" in *Proc. 19th ACM Int. Conf. World Wide Web*, 2010, pp. 591–600.

[3] A. Mislove, H. Koppula, K. Gummadi, P. Druschel, and B. Bhattacharjee, "Growth of the flickr social network," in *Proc. 1st ACM Workshop Online Social Netw.*, 2008, pp. 25–30.

[4] A. Mislove, M. Marcon, K. Gummadi, P. Druschel, and B. Bhattacharjee, "Measurement and analysis of online social networks," in *Proc. 7th ACM SIGCOMM Conf. Internet Meas.*, 2007, pp. 29–42.

[5] B. Viswanath, A. Mislove, M. Cha, and K. Gummadi, "On the evolution of user interaction in facebook," in *Proc. 2nd ACM Workshop Online Social Netw.*, 2009, pp. 37–42.

[6] M. Franceschetti, O. Dousse, D. Tse, and P. Thiran, "Closing the gap in the capacity of wireless networks via percolation theory," *IEEE Trans. Inf. Theory*, vol. 53, no. 3, pp. 1009–1018, Mar. 2007.

[7] Q. Peng, X. Wang, and H. Tang, "Heterogeneity increases multicast capacity in clustered network," in *Proc. IEEE Conf. Comput. Commun.*, 2011, pp. 703–711.

[8] L. Ying, S. Yang, and R. Srikant, "Optimal delay–Throughput tradeoffs in mobile ad hoc networks," *IEEE Trans. Inf. Theory*, vol. 54, no. 9, pp. 4119–4143, Sep. 2008.

[9] P. Li, C. Zhang, and M. Y. Fang, "Capacity and delay of hybrid wireless broadband access networks," *IEEE J. Sel. Areas Commun.*, vol. 27, no. 2, pp. 117–125, Feb. 2009.

[10] P. Li, C. Zhang, and M. Y. Fang, "The capacity of wireless ad hoc networks using directional antennas," *IEEE Trans. Mobile Comput.*, vol. 10, no. 10, pp. 1374–1387, Oct. 2011.

[11] X. Wang, L. Fu, and C. Hu, "Multicast performance with hierarchical cooperation," *IEEE/ACM Trans. Netw.*, vol. 20, no. 3, pp. 917–930, Jun. 2012.

[12] N. Lu, T. H. Luan, M. Wang, X. Shen, and F. Bai, "Capacity and delay analysis for social-proximity urban vehicular networks," in *Proc. IEEE Conf. Comput. Commun.*, 2012, pp. 1476–1484.

[13] P. Gupta and P. R. Kumar, "The capacity of wireless networks," *IEEE Trans. Inf. Theory*, vol. 46, no. 2, pp. 388–404, Mar. 2000.

[14] S. Milgram, "The small world problem," *Psychol. Today*, vol. 2, no. 1, pp. 60–67, 1967.

[15] P. Killworth and H. Bernard, "The reversal small-world experiment," *Social Netw.*, vol. 1, no. 2, pp. 159–192, 1979.

[16] P. Dodds, R. Muhamad, and D. Watts, "An experimental study of search in global social networks," *Science*, vol. 301, no. 5634, pp. 827–829, 2003.

[17] J. Kleinberg, "Navigation in a small world," *Nature*, vol. 406, no. 6798, p. 845, 2000.

[18] B. Azimdoost, H. Sadjadpour, and J. Garcia-Luna-Aceves, "Capacity of composite networks: Combining social and wireless ad hoc networks," in *Proc. IEEE Wireless Commun. Netw. Conf.*, 2011, pp. 464–468.

[19] B. Azimdoost and H. Sadjadpour, "Capacity of scale free wireless networks," in *Proc. IEEE Global Commun. Conf.*, 2012, pp. 2379–2384.

[20] D. Liben-Nowell, J. Novak, R. Kumar, P. Raghavan, and A. Tomkins, "Geographic routing in social networks," *Proc. Nat. Acad. Sci. USA*, vol. 102, no. 33, pp. 11623–11628, 2005.

[21] S. Li, Y. Liu, and X.-Y. Li, "Capacity of large scale wireless networks under Gaussian channel model," in *Proc. 14th ACM Int. Conf. Mobile Comput. Netw.*, 2008, pp. 140–151.

[22] X.-Y. Li, S.-J. Tang, and O. Frieder, "Multicast capacity for large scale wireless ad hoc networks," in *Proc. 13th Annu. ACM Int. Conf. Mobile Comput. Netw.*, 2007, pp. 266–277.

[23] G. Alfano, M. Garetto, and E. Leonardi, "Capacity scaling of wireless networks with inhomogeneous node density: Upper bounds," *IEEE J. Sel. Areas Commun.*, vol. 27, no. 7, pp. 1147–1157, Sep. 2009.

[24] G. Alfano, M. Garetto, E. Leonardi, and V. Martina, "Capacity scaling of wireless networks with inhomogeneous node density: Lower bounds," *IEEE/ACM Trans. Netw.*, vol. 18, no. 5, pp. 1624–1636, Oct. 2010.

[25] C. Wang, C. Jiang, X.-Y. Li, S. Tang, Y. He, X. Mao, and Y. Liu, "Scaling laws of multicast capacity for power-constrained wireless networks under gaussian channel model," *IEEE Trans. Comput.*, vol. 61, no. 5, pp. 713–725, May 2012.

[26] C. Manning and H. Schütze, *Foundations of Statistical Natural Language Processing*, Cambridge, MA, USA: MIT Press, 1999.

[27] V. Lenders, M. May, and B. Plattner, "Density-based anycast: A robust routing strategy for wireless ad hoc networks," *IEEE/ACM Trans. Netw.*, vol. 16, no. 4, pp. 852–863, Aug. 2008.

[28] C. Carter, S. Yi, P. Ratanchandani, and R. Kravets, "Manycast: Exploring the space between anycast and multicast in ad hoc networks," in *Proc. 9th ACM Annu. Int. Conf. Mobile Comput. Netw.*, 2003, pp. 273–285.

[29] M. Penrose, "A strong law for the longest edge of the minimal spanning tree," *Ann Probab.*, vol. 27, no. 1, pp. 246–260, 1999.

[30] C. Wang, C. Jiang, X.-Y. Li, S. Tang, and P. Yang, "General capacity scaling of wireless networks," in *Proc. IEEE Conf. Comput. Commun.*, 2011, pp. 712–720.

[31] B. Latané, J. Liu, A. Nowak, M. Bonevento, and L. Zheng, "Distance matters: Physical space and social impact," *Personality Social Psychol. Bull.*, vol. 21, no. 8, pp. 795–805, 1995.

[32] J. Møller, "Shot noise cox processes," *Adv. Appl. Probab.*, vol. 35, no. 3, pp. 614–640, 2003.

[33] A. ÖzgÜr, O. LÉvÊque, and D. Tse, "Hierarchical cooperation achieves optimal capacity scaling in ad hoc networks," *IEEE Trans. Inf. Theory*, vol. 53, no. 10, pp. 3549–3572, Oct. 2007.

[34] J. Steele, "Growth rates of euclidean minimal spanning trees with power weighted edges," *Ann. Probab.*, vol. 16, pp. 1767–1787, 1988.

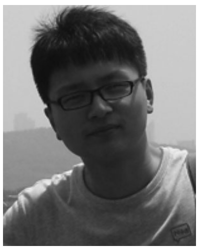
[35] D. Williams, *Probability with Martingales*, Cambridge, U.K.: Cambridge Univ. Press, 1991.

[36] N. Z. Gong and W. Xu, "Reciprocal versus Parasocial Relationships in Online Social Networks," Springer Social Network Analysis and Mining, vol. 4, no. 1, pp. 184–197, 2014.

[37] N. Z. Gong, A. Talwalkar, L. Mackey, L. Huang, E. C. R. Shin, E. Stefanov, E. R. Shi, and D. Song, "Joint link prediction and attribute inference using a social-attribute network," *ACM Trans. Intell. Syst. Technol.*, vol. 5, no. 2, pp. 1–27, Apr. 2014.

[38] N. Z. Gong, W. Xu, L. Huang, P. Mittal, E. Stefanov, V. Sekar, and D. Song, "Evolution of social-attribute networks: Measurements, modeling, and implications using google+," in *Proc. ACM Conf. Internet Meas. Conf.*, 2012, pp. 131–144.

Cheng Wang received the PhD degree from the Department of Computer Science at Tongji University in 2011. He is currently a research professor of computer science at Tongji University. His research interests include wireless networking, mobile social networks, and cloud computing.



Lu Shao received the bachelor's degree of science at the College of Science from China University of Petroleum in June, 2013. He is currently working toward the master degree in the Department of Computer Science at Tongji University in Shanghai, China. His research interests include wireless networks and distributed computing.

Zhong Li received the BS and MS degrees from the Department of Computer Science and Technology from Shandong Normal University, in 2007 and 2010, respectively. She is currently working toward the PhD degree from the Department of Computer Science at Tongji University in Shanghai, China. Her research interests include wireless communication, social network analysis, and distributed computing.

Lei Yang received the BS degree in the School of Software and the PhD degree in computer science and technology from Xian Jiaotong University, China, in 2004 and 2014, respectively. He is currently with the School of Software in Tsinghua University. His research interests include RFID, ambient backscatter, pervasive computing, smart home, and communication security. He is a member of the IEEE and the IEEE Computer Society.

Xiang-Yang Li (M'99-SM'08) received the bachelor's degree from the Department of Computer Science, Tsinghua University in 1995. He received the MS and PhD degrees from the Department of Computer Science, University of Illinois at Urbana-Champaign, in 2000 and 2001, respectively. He is currently a professor of computer science at the Illinois Institute of Technology. He received the China National Science Foundation (NSF) Outstanding Overseas Young Researcher (B). His research interests include mobile computing, cyber physical systems, wireless networks, security and privacy, and algorithms.

Changjun Jiang received the PhD degree from the Institute of Automation, Chinese Academy of Sciences, Beijing, China, in 1995. He is currently a professor with the Department of Computer Science and Engineering, Tongji University, Shanghai. His research interests include concurrent theory, Petri net, and formal verification of software, wireless networks, concurrency processing, and intelligent transportation systems.