Effect of Social Groups on the Capacity of Wireless Networks

Mohsen Karimzadeh Kiskani†, Bita Azimdoost†, and Hamid R. Sadjadpour†,

***Abstract*—In this paper, we study the effects of social interactions among nodes on the capacity of wireless networks. We consider three scenarios. In the first scenario, the size of the social group for all nodes is fixed while the frequency of communication within members of a social group follows power law distribution. In the second scenario, scale free networks are studied where the size of the social group differs from node to node, and the destination in each group is selected uniformly among the members of that group. Further investigation in the second scenario reveals that traditional transport capacity definition provides misleading conclusions for such network models. We show that nodes with different social status impact the capacity differently. By separating nodes with different social status and allocating separate bandwidth to them, it is shown that majority of nodes scale in this network. In the third scenario, both the size of the social groups and the destination in each group are selected according to power law distributions. Our simulation results corroborate the analytical results. Further, we observe consistently that social interaction improves the capacity of wireless networks which implies that the Gupta-Kumar results were pessimistic for practical networks.**

***Index Terms*—Wireless Social Networks; Scaling Laws in Wireless Networks; Complex Networks**

I. Introduction

The widespread use of Internet has been a major factor in changing users' behaviors in the past decade. The extensive amount of content growth on the web has made it very difficult for users to access their desired contents without content personalization. This, along with privacy concerns have been the significant forces in moving towards social applications like Facebook, Youtube and many other social networks in which each user only connects and communicates within a restricted social group and only accesses the contents from its social group. This clearly affects the user's method and frequency of communication within each group.

On the other hand, with the emergence of advanced hardware and software technologies which enable significant processing power and storage space in mobile devices, and with their ever increasing widespread use, today's Internet is moving [1], [2] from an infrastructure-based network towards a wireless ad-hoc network. At the same time, the social behaviors and interests of users are changing wireless networks into wireless social networks which are significantly influenced by users' social behaviors. In these social networks, the users do not necessarily communicate with a central server [1], [2] and instead, based on their social interests they can communicate with nodes inside a wireless ad-hoc network. Therefore, a portion of future data communication networks can be envisioned as social wireless ad-hoc networks.

While in today's Internet because of the use of fiber optic backbone, the throughput may not be seen a big problem, the rapid increase of video steaming applications like Youtube or Netflix which account for over half of the Internet traffic in North America[1] can potentially be the bottleneck for communications over future wireless networks. A concrete example of such network can be the future 5G networks [1], [2] in which a portion of the data traffic and video streaming should be carried over wireless ad-hoc networks. Theoretical analysis of capacity for these networks becomes increasingly important.

In their seminal paper [3], Gupta and Kumar found capacity scaling laws for a dense network of n users. Many subsequent works tried to compute the throughput capacity using different assumptions. Grossglauser and Tse [4] proved that mobility increases the capacity of wireless networks, Gastpar and Vetterli [5] studied the capacity of wireless networks with relays and [6] studied the capacity of hybrid wireless networks which are formed by placing a sparse network of base stations in an ad hoc network. Kulkarn and Viswanath [7] proposed a deterministic approach to compute the throughput capacity in wireless ad-hoc networks. In all of these works and the subsequent studies, source-destination selection was completely random following a uniform distribution.

However, in social wireless ad-hoc networks the nodes are selecting their destinations in the context of social groups which means that the nodes are not communicating with random nodes outside their social groups. In many situations, the source may not have a prior knowledge about the members of social groups. Backstrom et al. [8] observed that, the probability that each node being selected as a member of social group decreases with its distance to the origin according to a power-law distribution. This paper considers distance-based communications with power law distribution and parameter α.

On the other hand, the frequency of communicating with different nodes within a social group is not the same; some nodes are contacted more frequently than others. Latane et al. [9] studied the frequency of social interactions in social networks and observed that in these networks, the probability of choosing the members of a social group is inversely proportional to distance according to a power-law distribution. We will consider a power-law distribution with parameter β

M. Kiskani†, B. Azimdoost† and H. Sadjadpour† are with the Department of Electrical Engineering, University of California, Santa Cruz. Email: {mohsen, bazimdoost, hamid}@soe.ucsc.edu.

[1]http://mashable.com/2013/11/12/internet-traffic-downstream/

for the frequency of communications within a social group.

The number of members of social groups is also a random number in actual wireless social networks. Studies on complex networks [10]–[13], which are a superset of social networks, suggest that these networks are scale-free networks meaning that they have power-law degree distributions. Therefore, we assume a power law distribution with parameter γ for the size of social groups in our derivations.

We study the effect of social groups on the throughput capacity of social wireless ad-hoc networks considering these distributions. It is predicted that future 5G networks [1], [2] have large number of nodes, each trying to access personalized data and video over a wireless ad hoc network[2] which makes this study important in understanding the performance of such networks.

The first result on the capacity of the wireless ad-hoc networks with social consideration was reported in [14] where the authors considered the same model as in [3] but with the extra assumption that each node has a social group consisting of local neighbors and just one long-range contact as its destination. They later generalized [15] this idea to q long-range social contacts, one of which is selected uniformly at random as destination. The authors in [15] assume a very simple model in which they only consider the effect of distance on the probability of membership in social groups. They assume a constant fixed number of nodes in each social group with uniform frequency of communication within each social group. In the current paper, we improve the results in [15] by assuming social groups with different number of nodes and with non-uniform frequency of communication within each social group. Further, none of these works considers the frequency of communications inside social groups. In this paper, we consider different aspects of social networks by using parameters α, β and γ for three different power-law distributions, each of which representing a different feature of the social networks.

The rest of the paper is organized as follows. In section II, the network assumptions and the routing model are described. In section III, we assume selection of destination for each node according to a power-law distribution with parameter β. Section IV considers the number of long-range social contacts is no longer a fixed number and has a power-law distribution with parameter γ. Section V studies the impact of the power law destination selection in a power law degree distributed social network. Our analytical results are corroborated by simulation results in section VI. We conclude the paper in section VII.

II. Network Model

We study the throughput capacity of wireless social networks with n nodes randomly distributed in a square area. The scaling laws for the capacity of such networks without any social consideration was computed in [3]. In the current paper, however, nodes can only communicate with members of their social groups which are known to them a priori. We assume that the members of social groups for all nodes are selected

[2]http://flexible-radio.com/news/5g-radio-network-architecture

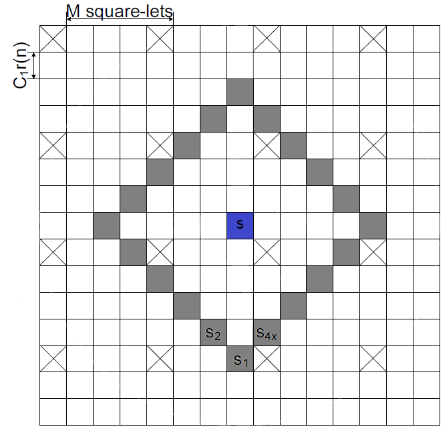

Fig. 1. All of the nodes in the network are assumed to be distributed in a unit square area which is divided into many square-lets of side length $C_1r(n)$. Dark gray cells S_i contain the nodes within the same lattice distance x from the source which is assumed to be in the center of the unit square. Cells with crosses inside are those cells who can transmit simultaneously.

in advance and does not change with time. These social considerations can result in different average hop count distance for each source-destination pair compared to [3] which will result in different throughput capacity performance. To guarantee connectivity in such dense networks, the minimum transmission range [16] is $r(n) = \Theta(\sqrt{\frac{\log n}{n}})$ which will be used in this paper. We consider *protocol model* defined in [17] for the successful communication between nodes. According to this model, if the node i is placed at the coordinates X_i, then a transmission from i to j is successful if $|X_i - X_j| < r(n)$ and for any other node k transmitting on the same frequency band, $|X_k - X_j| > (1+\Delta)r(n)$ for a fixed guard zone factor Δ.

Our routing approach is based on the fact that each node knows the location of its destination and selects the shortest path to the destination. We will use the deterministic routing strategy suggested in [7] for our network analysis. It is shown [7] that if we divide the unit square into many square cells each with a side length of $\Theta(\sqrt{\frac{\log n}{n}})$, there is at least one node in each cell almost surely. By selecting $r(n) = \Theta(\sqrt{\frac{\log n}{n}})$, the approach in [7] guarantees that the routing algorithm will converge to transport data from each source to its destination.

To avoid multiple access interference (MAI) in cells, we use [7] time division multiple access (TDMA) scheme. Assuming the side length for each cell to be equal to $C_1\sqrt{\frac{\log n}{n}}$ (see Figure 1) where C_1 is a fixed constant that assures all nodes in the adjacent cells are within the transmission range. Based on the Protocol model, we only allow simultaneous transmissions when nodes are at least M cells away from each other, with $M \geq \frac{2+\Delta}{C_1}$ (see cross sign in Fig. 1)

We denote the data rate for each node as λ, the number of hops for each source-destination pair as X and its average value as $E[X]$. The total throughput by all nodes is $n\lambda$. Therefore, on average the network delivers $n\lambda E[X]$ bits in a unit of time. There are exactly $\frac{1}{(MC_1r(n))^2}$ square-lets at any time slot available for transmission and the total network bandwidth is W which is a constant value independent of n. Thus, the total number of bits that the network is capable of delivering is upper bounded by $\frac{W}{(MC_1 r(n))^2}$. Hence,

$$\lambda \leq \lambda_{\max} = \frac{W}{nE[X](MC_1 r(n))^2} = \Theta(\frac{1}{E[X]\log n}). \quad (1)$$

This result implies that the maximum throughput can be derived by computing $E[X]$. Our problem is reduced to computing the average number of hops between source-destinations pairs which depends on the social characteristics of the network. Our goal is to study the average hop count and capacity and since the geographical location of the source node only changes the distances by a constant value (not a function of n), we can do our calculations for a node at the center of the square and consider the result as the order of average hop count for all the nodes.

In order to compute the average hop count of $E[X]$, it is easy to observe that the maximum number of hops for each source-destination pair is in the order of $\frac{1}{C_1 r(n)} = \Theta(r(n)^{-1})$. Thus, we have $E[X] = \sum_{x=1}^{\frac{1}{r(n)}} x\Pr(X = x)$. The geometrical locus of cells in the unit square with the hop distance of x from the source which is located in the center is a rhombus (see figure 1 for the case of $x = 4$). The probability that the number of hops between source and destination is x equals the probability that the destination is located in one of the cells of this rhombus. Let's denote the cells on this rhombus as $s_1, s_2, ..., s_{4x}$. For the rest of this paper, we denote the social group of the source located in the center by G, the source by S, the destination by T, the node i by v_i and its distance to the source by d_i for $1 \leq i \leq n$. Using these notations, the probability that the number of hops between the source and destination X, is equal to x is given by $\Pr(X = x) = \sum_{l=1}^{4x} \sum_{v_k \in s_l} \Pr(T = v_k)$. Since each node communicates only with the members of its social group, then $\Pr(T = v_k \mid v_k \notin G) = 0$ for $1 \leq k \leq n$ and we have,

$$\Pr(T = v_k) = \Pr(T = v_k | v_k \in G)\Pr(v_k \in G). \quad (2)$$

We use *elementary symmetric polynomial* notations to simplify the presentation. If we have n variables $x_1, x_2, ..., x_n$, then the k-th degree elementary symmetric polynomial of these variables is denoted as $\sigma_k(\mathbf{X}) = \sigma_k(x_1, ..., x_n) \triangleq \sum_{1\leq i_1<i_2<..<i_k\leq n} x_{i_1}...x_{i_k}$. The elementary symmetric polynomial [18] for n variables excluding the j-th element using the vector representation $\mathbf{X}^{\bar{j}} = (x_1, ..., x_{j-1}, x_{j+1}, ..., x_n)$ is defined as $\sigma_k(\mathbf{X}^{\bar{j}}) \triangleq \sigma_k(x_1, ..., x_{j-1}, x_{j+1}, ..., x_n)$.

To compute $\Pr(v_k \in G)$, we use the results from [8], [9], [19] to assume that each node selects its social group members according to a power-law distribution versus distance with parameter α. Hence, the probability of a node j with distance d_j belonging to social group G is proportional to $d_j^{-\alpha}$. If the nodes in G are labeled as $g_1, g_2, ..., g_q$, then the probability that G consists of these nodes is

$$\Pr(G = \{v_{g_1}, ..., v_{g_q}\}) = \frac{d_{g_1}^{-\alpha}...d_{g_q}^{-\alpha}}{\sum_{1\leq i_1<...<i_q\leq n} d_{i_1}^{-\alpha}...d_{i_q}^{-\alpha}}. \quad (3)$$

Therefore, the probability of a particular node v_k being a member of G is given by

$$\begin{aligned}\Pr(v_k \in G) &= \frac{\sum_{1\leq i_1<...<i_{q-1}\leq n, i_j\neq k} d_k^{-\alpha} d_{i_1}^{-\alpha}...d_{i_{q-1}}^{-\alpha}}{\sum_{1\leq i_1<...<i_q\leq n} d_{i_1}^{-\alpha}...d_{i_q}^{-\alpha}} \\ &= \frac{d_k^{-\alpha}\sigma_{q-1}(\mathbf{d_n^{\bar{k}}})}{\sigma_q(\mathbf{d_n})}, \end{aligned} \quad (4)$$

where $\mathbf{d_n} \triangleq (d_{i_1}^{-\alpha}, ..., d_{i_n}^{-\alpha})$. Thus

$$E[X] = \sum_{x=1}^{\frac{1}{r(n)}} x \sum_{l=1}^{4x} \sum_{v_k \in s_l} \frac{d_k^{-\alpha}\sigma_{q-1}(\mathbf{d_n^{\bar{k}}})}{\sigma_q(\mathbf{d_n})} \Pr(T = v_k | v_k \in G). \quad (5)$$

III. Distance-based Frequency of Communication

This section focuses on the case when the social connections have been formed according to the power law distribution described in section II with parameter α, all the nodes have a fixed number of long range social contacts q, and select their destinations inside the social groups based on distance, according to another power law distribution. This last assumption is based on a highly cited paper [9] on frequency of communication inside social groups. We assume that within the social group G, the source selects its destination according to a power law distribution with parameter β. By defining $\mathbf{d_q} = (d_{g_1}^{-\beta}, ..., d_{g_q}^{-\beta})$, we have

$$\Pr(T = v_k \mid v_k \in G) = \frac{d_k^{-\beta}}{\sum_{j=1}^{q} d_j^{-\beta}} = \frac{d_k^{-\beta}}{\sigma_1(\mathbf{d_q})}, \quad (6)$$

which reduces equation (5) to

$$E[X] = \sum_{x=1}^{\frac{1}{r(n)}} x \sum_{l=1}^{4x} \sum_{v_k \in s_l} \frac{d_k^{-\alpha-\beta}\sigma_{q-1}(\mathbf{d_n^{\bar{k}}})}{\sigma_1(\mathbf{d_q})\sigma_q(\mathbf{d_n})}. \quad (7)$$

Next, we compute the value of $E[X]$ based on the size of social group, q.

Theorem 1. *When $q = \Theta(n)$, the average hop count is*

$$E[X] \equiv \begin{cases} \Theta(r(n)^{-1}), & 0 \leq \beta \leq 2 \\ \Theta(r(n)^{\beta-3}), & 2 \leq \beta \leq 3 \\ \Theta(1). & 3 \leq \beta \end{cases} \quad (8)$$

Proof: The proof is in appendix. ■

By replacing $r(n)$ with its minimum value, the maximum achievable throughput is given by

$$\lambda_{max} = \begin{cases} \Theta(\frac{1}{\sqrt{n\log n}}), & 0 \leq \beta < 2 \\ \Theta(\frac{1}{\log n}\sqrt{\frac{\log n}{n}}^{3-\beta}), & 2 \leq \beta \leq 3 \\ \Theta(\frac{1}{\log n}). & 3 < \beta \end{cases} \quad (9)$$

This result demonstrates that for $q = \Theta(n)$ and when the destination is selected based on distance, then the throughput capacity is independent of α. Further, we can achieve highest possible capacity ($\lambda_{max} = \Theta(\frac{1}{\log n})$) even for small values of α when $\beta > 3$. Based on this observation, it can be concluded that for this case, selecting destination based on distance is the dominant factor. This result can be justified by observing that since the total number of social contacts is proportional to n, then selecting them based on a power law distribution (with parameter α) does not make much difference since most of the nodes belong to all social groups. As we will see in the next theorem, the effect of α will appear as the nodes become more selective in choosing the members of their social groups.

Theorem 2. *When $q = \Theta(1)$, the average hop count is*

$$E[X] \equiv \begin{cases} \Theta(r(n)^{\beta-1}), & 0 \leq \beta \leq 1, 0 \leq \alpha \leq 2 \\ \Theta(r(n)^{\alpha+\beta-3}), & 0 \leq \alpha+\beta \leq 3, 2 \leq \alpha \\ \Theta(1). & \textit{Otherwise} \end{cases} \quad (10)$$

Proof: The proof is in [20] and for minimum value of $r(n)$, $\lambda_{\max}$ is given by

$$\lambda_{\max} \equiv \begin{cases} \Theta(\frac{1}{\log n}\sqrt{\frac{\log n}{n}}^{1-\beta}), & 0 \leq \beta \leq 1, 0 \leq \alpha \leq 2 \\ \Theta(\frac{1}{\log n}\sqrt{\frac{\log n}{n}}^{3-\alpha-\beta}), & 0 \leq \alpha+\beta \leq 3, 2 \leq \alpha \\ \Theta(\frac{1}{\log n}). & \text{Otherwise} \end{cases}$$

■

The results indicate that when both α and β are small, then social characteristics of the network has little effect on the throughput capacity (first capacity region). By increasing the value of α beyond 2, social characteristics start influencing and increasing the throughput capacity while the effect of communication network decreases (second capacity region). When we move beyond these values, social characteristics become dominant factor and the communication network does not have any effect on the capacity of the network. In this capacity region, average hop count is proportional to $\Theta(1)$ which is the direct result of strong social aspects of the network.

Theorem 3. *The average hop count when $q = \Theta(f(n))$, $\lim_{n\to\infty} q = \infty$, and $\lim_{n\to\infty} \frac{f(n)}{n} = 0$ is equal to*

$$E[X] \equiv \begin{cases} \Theta\left(f(n)r(n)^{\beta-1}\right), & 1 \leq \beta \leq 3, f(n) = \Omega(r(n)^{\beta-1}) \\ \Theta\left(f(n)r(n)^{2}\right), & 3 \leq \beta, f(n) = \Omega(r(n)^{2}) \\ \Theta(1). & \textit{Otherwise} \end{cases}$$

Proof: The proof is in the appendix and the achievable throughput is derived as

$$\lambda_{\max} \equiv \begin{cases} \Theta(\frac{1}{f(n)\log n}\sqrt{\frac{\log n}{n}}^{1-\beta}), & 1 \leq \beta \leq 3, \\ & f(n) = \Omega(\sqrt{\frac{\log n}{n}}^{\beta-1}) \\ \Theta(\frac{1}{f(n)\log n}\frac{n}{\log n}), & 3 \leq \beta, \\ & f(n) = \Omega(\frac{\log n}{n}) \\ \Theta(\frac{1}{\log n}). & \text{Otherwise} \end{cases}$$

■

Theorem 3 provides insight on the behavior of throughput capacity as a function of the number of social contacts for each node. This theorem explains how different social characteristics of the network that are represented by two parameters of β and α (α in these equations is indirectly reflected in $f(n)$) influence the throughput capacity for this general case.

From all these theorems, it can be concluded that in general when the social characteristics of the network become a dominant factor, then the throughput capacity of the network improves. On the other hand, when the wireless communication characteristics of the network is dominant, the throughput capacity will decrease up to the point that in the extreme case, it will be the same as Gupta-Kumar result (first capacity region in theorem 1 and in theorem 2 when $\beta = 0$).

Theorem 4. *These capacity results are achievable. In other words, no cell is a bottleneck and the traffic passing through each cell can be routed through.*

Proof: Proof is in the appendix. ■

IV. RANDOM SIZE OF SOCIAL GROUPS FOR EACH NODE

In this section, we again assume that the social network has been built according to the power law distribution with parameter α, but the size of each social group is a random variable denoted by Q. There are many studies [10], [11], [13] demonstrated that this random variable also follows power law distribution. Furthermore, the destination for each node among members of its social group is selected using uniform distribution ($\beta = 0$).

Based on these assumptions, we have

$$\Pr(T = v_k | v_k \in G, Q = q) = \frac{1}{q}. \quad (11)$$

Let's define $\mathbf{b} = (1^{-\gamma}, 2^{-\gamma}, ..., n^{-\gamma})$, then the probability distribution of Q will have the form

$$\Pr(Q = q) = \frac{q^{-\gamma}}{\sum_{b=1}^{n} b^{-\gamma}} = \frac{q^{-\gamma}}{\sigma_1(\mathbf{b})}, \quad (12)$$

where γ is the power law exponent of the distribution for Q. From (5) we arrive at

$$E[X|Q=q] \equiv \sum_{x=1}^{\frac{1}{r(n)}} x \sum_{l=1}^{4x} \sum_{v_k \in s_l} \frac{d_k^{-\alpha}\sigma_{q-1}(\mathbf{d}_\mathbf{n}^{\bar{\mathbf{k}}})}{q\sigma_q(\mathbf{d_n})}, \quad (13)$$

and

$$\begin{aligned} E[X] &= \sum_{q=1}^{n} \Pr(\mathrm{Q=q}) E[X|Q=q] \\ &\equiv \sum_{x=1}^{\frac{1}{r(n)}} x \sum_{l=1}^{4x} \sum_{v_k \in s_l} \sum_{q=1}^{n} \frac{q^{-\gamma} d_k^{-\alpha}\sigma_{q-1}(\mathbf{d}_\mathbf{n}^{\bar{\mathbf{k}}})}{q\sigma_1(\mathbf{b})\sigma_q(\mathbf{d_n})}. \quad (14) \end{aligned}$$

Theorem 5. *In this case, $E[X]$ has the order of $\Theta(r(n)^{-1})$.*

Proof: The proof is in [21]. Using minimum transmission range, then $\lambda_{\max} = \Theta(\frac{1}{\sqrt{n \log n}})$. During the proof process, we have expanded the sum into two terms E_1 and E_2, where assuming that q_0 is a large constant number, E_2 denotes the case when nodes have $q \leq q_0$ social contacts and E_1 is for the case with nodes having $q \geq q_0 + 1$ social contacts. Then it is shown that E_1 is the dominant factor in the summation and its order is equal to $\frac{1}{r(n)}$. ■

At first, this theorem implies that by introducing some social characteristics in the network, the capacity of wireless networks becomes similar to that of Gupta-Kumar which is counter-intuitive. However by carefully reviewing the result, we observe that when nodes have large social contact size, they require significant network resources for communications while with small social contact size, they require much less network resources to transport packets from sources to destinations. Such significant disparity in capacity behavior among nodes suggests that the conventional definition of transport capacity for wireless communication networks is not appropriate for scale free wireless social networks. In order to demonstrate this unique behavior of these composite networks, we divide the nodes into two groups based on their social status, i.e., popularities. Furthermore, we divide the bandwidth W into two equal parts and allow communication for each group of nodes within its allocated bandwidth. Note that this will not change the order of the throughput capacity for each group. Clearly, in order to preserve the connectivity in the network, we still allow nodes in different social status to relay messages for the other group of nodes. This approach may not be a practical technique, but it will shed some light on the behavior of composite networks. Future investigation is required to come up with practical communication techniques for composite networks.

Lemma 1. *Let q_0 be a large constant number. For small social degree distribution exponent ($0 < \gamma < 1$), the number of nodes with more than q_0 social contacts ($N_{>q_0}$) is $\Theta(n)$ and the number of nodes with less than q_0 social contacts ($N_{\leq q_0}$) is $\Theta(n^\gamma)$. Furthermore, for large social degree distribution exponent ($\gamma > 2$), this ratio of $\frac{N_{>q_0}}{N_{\leq q_0}}$ is $q_0^{1-\gamma}\Theta(1)$ which is a very small number for sufficiently large q_0.*

Proof: This is lemma 2 in [21] and due to page limitations, we omit the proof here. ■

In other words, for large γ, the number of nodes involving in E_1 is much less than the nodes which generate the E_2 part of the total average number of hops. it is shown in [21] that for large values of γ and α, E_2 is much larger than E_1. The following remark describes the network resource usages of $E1$ and E_2 for large values of α and γ.

Remark 1. *In highly concentrated social networks (large α) with large social degree distribution exponent (large γ) , a very small group of nodes ($N_{>q_0}$) use the majority of the resources (due to the large average number of hops traveled by each packet to reach the destination), while a large group of nodes ($N_{\leq q_0}$) use a small portion of the resources.*

This remark implies that conventional definition of transport capacity may not be appropriate for scale-free networks. In these networks, transportation of a single packet requires different amount of network resources in terms of relaying and average number of hops to reach destination. Based on this observation, it makes sense that we divide the nodes into two categories. One group of nodes are less popular and their social group size is small, i.e., $N_{\leq q_0}$ and the other group of nodes are those nodes that are more popular with higher social status with many social contacts, i.e., $N_{>q_0}$. We divide the available bandwidth W into two equal parts and allow communication for each group inside their own bandwidth. By doing so, there is more fairness in each group in terms of utilizing the network resources for transmission of packets to destinations which will ultimately allow us to better understand the performance of the network. Note that by dividing the available bandwidth into two, the order of the throughput capacity will not change for each group. Clearly, in order to preserve the connectivity in the network, we still allow nodes in different social status to relay messages for the other group of nodes. For example if $q_0 = 100$ and $\gamma = 2.5$, then it is easy to show that 99.9% of nodes can scale while only 0.1% of nodes with larger than 100 social contacts will not scale. It is easy to demonstrate that the maximum data rate for sources with $N_{>q_0}$ (E_1) is similar to that of Gupta-Kumar. We use the results of Lemma 1 for sources in the second category, i.e., $N_{\leq q_0}$, to compute the throughput capacity.

$$\begin{aligned} \lambda_{max_{\leq q_0}} &= \Theta(\frac{W/2}{E_2 \log n}) \\ &= \begin{cases} \Theta(\frac{1}{\sqrt{n \log n}}) & 0 < \alpha < 2 \\ \Theta(\frac{1}{\sqrt{n^{3-\alpha} \log^{\alpha-1} n}}) & 2 < \alpha < 3 \\ \Theta(\frac{1}{\log n}) & 3 < \alpha \end{cases} \end{aligned} \tag{15}$$

These two capacity results prove the following theorem.

Theorem 6. *Assume that the social connectivity between nodes is highly concentrated ($\alpha > 2$) with large social degree distribution exponent ($\gamma > 2$). Let's divide the total bandwidth (W) into two distinct parts, $W/2$ each; one part to be used to transfer the information generated from the highly connected source nodes ($G_{>q_0}$) and the other part to be used for communication by the source nodes with small social group size ($G_{\leq q_0}$) where q_0 is a constant value independent of n. The maximum data rate for the first group ($G_{>q_0}$) is $\lambda_{max} = \Theta(\frac{1}{\sqrt{n \log n}}),\ for\ 2 < \alpha$. The maximum data rate for the second group ($G_{\leq q_0}$) is given in equation (15).*

This theorem shows that nodes with different social status, i.e., different number of social contacts, have different effect on throughput capacity. It can be observed that the limiting factor in scaling the capacity is the existence of few nodes with high social status that consume majority of the network resources in terms of relaying requirements. More specifically, it was shown that the nodes that limit the capacity consist of a small portion of the network under the condition that the social groups are geographically highly concentrated ($\alpha > 2$) and the degree distribution exponent is large ($\gamma > 2$). Figures 2(a) and (b) demonstrate data rates for these two groups of nodes, when $\gamma > 2$.

Theorem 7. *The obtained capacity results in theorem 6 are achievable. In other words, no cell is a bottleneck and the traffic passing through each cell can be routed through.*

Proof: Proof is in the appendix. ■

V. POWER-LAW DESTINATION SELECTION WITH RANDOM NUMBER OF SOCIAL CONTACTS

In this section we study the impact of the combination of all three power law distributions on the network performance; the social network formation with parameter α for selecting the long range contacts and parameter γ for the number of long range contacts, and the communication among the members of the social group with parameter β. We can modify the analysis

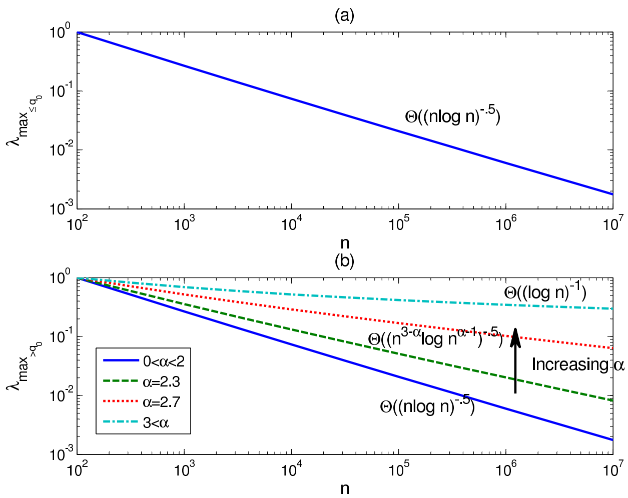

Fig. 2. Maximum achievable data rate order for (a) highly connected source nodes (Group $G_{>q_0}$) with large degree distribution exponent ($\gamma > 2$), (b) nodes with small social group (Group $G_{\leq q_0}$).

in section IV to get the results. Using equations (5), (6), (13) and (14) we have

$$\begin{aligned} E[X|Q=q] &\equiv \sum_{x=1}^{\frac{1}{r(n)}} x \sum_{l=1}^{4x} \sum_{v_k \in s_l} \Pr(T = v_k) \\ &\equiv \sum_{x=1}^{\frac{1}{r(n)}} x \sum_{l=1}^{4x} \sum_{v_k \in s_l} \frac{d_k^{-\alpha-\beta} \sigma_{q-1}(\mathbf{d_n})}{\sigma_1(\mathbf{d_q}) \sigma_q(\mathbf{d_n^{\bar{k}}})} \quad (16) \end{aligned}$$

and

$$\begin{aligned} E[X] &= \sum_{q=1}^{n} \Pr(\mathrm{Q=q}) E[X|Q=q] \\ &\equiv \sum_{x=1}^{\frac{1}{r(n)}} x \sum_{l=1}^{4x} \sum_{v_k \in s_l} \sum_{q=1}^{n} \frac{q^{-\gamma} d_k^{-\alpha-\beta} \sigma_{q-1}(\mathbf{d_n^{\bar{k}}})}{\sigma_1(\mathbf{b}) \sigma_1(\mathbf{d_q}) \sigma_q(\mathbf{d_n})} \quad (17) \end{aligned}$$

Theorem 8. *The average hop count in this case has the order*

$$E[X] \equiv \begin{cases} \Theta\left(r(n)^{-1+\beta}\right), & 0 \leq \beta \leq 1 \\ \Theta\left(1\right), & 1 \leq \beta \end{cases} \quad (18)$$

Proof: The proof is in the appendix. ∎

Corollary 1. *The maximum throughput capacity is*

$$\lambda_{\max} \equiv \begin{cases} \Theta\left(\dfrac{1}{\sqrt{n^{1-\beta}\log^{1+\beta} n}}\right), & 0 \leq \beta \leq 1 \\ \Theta\left(\dfrac{1}{\log n}\right), & 1 \leq \beta \end{cases} \quad (19)$$

VI. Numerical Simulations

The results in this paper, which are obtained through mathematical proofs are expressed in terms of scaling laws. In order to validate our theoretical results with simulations, we need to use very large values for n. However, using very large values for n is not practical due to the non-polynomial number of computations. For instance, in section III, we have $\binom{n}{q}$ different possibilities to choose q members of the social group from the total number of nodes n. Each one of these choices has an associated probability expressed in equation (3). This means that for any numerical simulation, we need to compute the associated probabilities. Now, if $q = \Theta(f(n))$ then we should compute these probabilities for at least $\binom{n}{q} = \binom{n}{f(n)} \geq \left(\frac{n}{f(n)}\right)^{f(n)}$ different choices. This value grows faster than exponential for many choices of $f(n)$. Therefore, conducting a comprehensive numerical analysis for the theoretical results in this paper is almost impossible except for special cases of $q = \Theta(1)$ and $q = \Theta(n)$ that we have been able to simulate our results and compare them against the theoretical results. Figure 3 shows the average hop count between theory and simulation. The results clearly demonstrate that our theoretical derivations are very close to simulation results as the number of nodes in the network increases. For the case of $\beta = 3.5$, we only show the simulation results which is consistent with theory, i.e., $E[X] = \Theta(1)$.

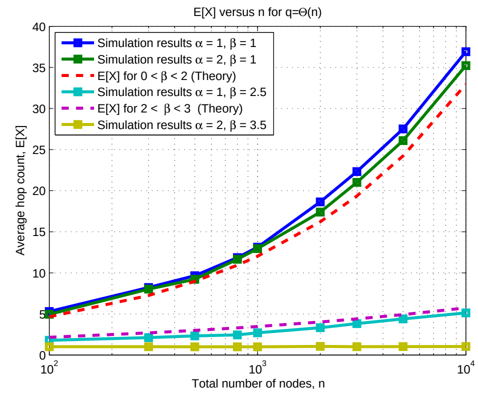

Fig. 3. Average number of hops versus the number of nodes when $q = \Theta(n)$.

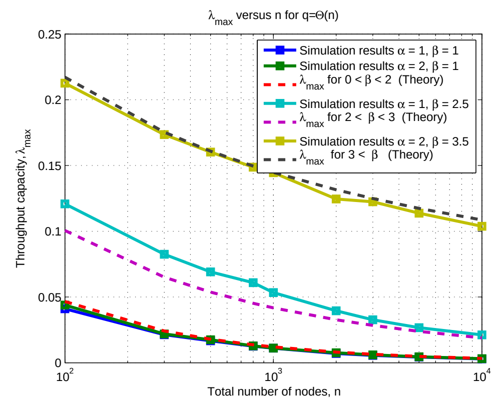

Fig. 4. Maximum throughput capacity versus the number of nodes when $q = \Theta(n)$.

Figure 4 demonstrates the maximum throughput as a function of n when $q = \Theta(n)$. We can see from this figure, that for different values of α and β, the simulation results are very close to theoretical results which verifies the accuracy of the analytic work.

Figures 5 and 6 compare the simulation results with theory for the case of $q = \Theta(1)$. Both the analytical results for the average number of hops and throughput capacity for different values of α and β are close to simulation results. From all these results, we can conclude that the analytical results accurately predict the behavior of the network when social characteristics are considered.

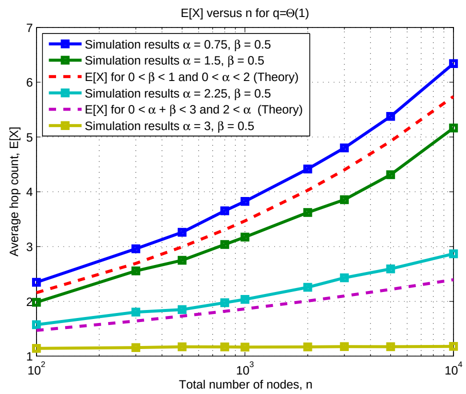

Fig. 5. Average number of hops versus the number of nodes when $q = \Theta(1)$.

VII. Conclusion

In this paper, a comprehensive study of the effects of social communication which include non-uniform frequency of communication and variable size of social groups, on the capacity of wireless ad hoc networks has been investigated. We have shown that the traditional concept of capacity introduced by Gupta-Kumar may not be appropriate for these composite networks. Instead, if we divide the nodes based on their social status, we actually observe a completely different behavior in the network. We believe that based on our results, a new definition of capacity for wireless networks with social behaviors should be proposed which takes into account the social characteristics of network.

It is worth to emphasize that the effects of social group evolution is not considered in our network model and a more comprehensive work, should consider such effects in the study of wireless networks with social considerations. For future work, proper protocols for these wireless social networks can be studied, different resource allocations based on social status can be also investigated to name a few.

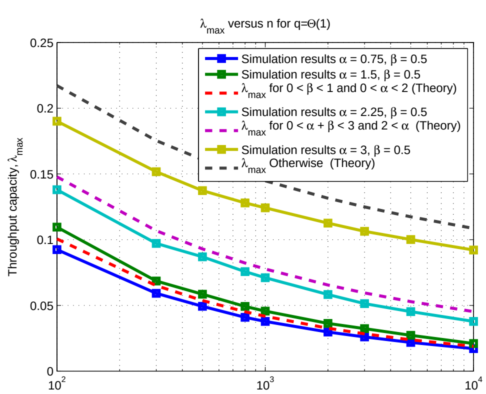

Fig. 6. Maximum throughput capacity versus the number of nodes when $q = \Theta(1)$.

References

[1] F. Boccardi, R. W. Heath, A. Lozano, T. L. Marzetta, and P. Popovski, "Five disruptive technology directions for 5G," *Communications Magazine, IEEE*, vol. 52, no. 2, pp. 74–80, Feb. 2014. [Online]. Available: http://dx.doi.org/10.1109/mcom.2014.6736746

[2] A. Osseiran, F. Boccardi, V. Braun, K. Kusume, P. Marsch, M. Maternia, O. Queseth, M. Schellmann, H. Schotten, H. Taoka, H. Tullberg, M. A. Uusitalo, B. Timus, and M. Fallgren, "Scenarios for 5G mobile and wireless communications: the vision of the METIS project," *Communications Magazine, IEEE*, vol. 52, no. 5, pp. 26–35, May 2014. [Online]. Available: http://dx.doi.org/10.1109/mcom.2014.6815890

[3] P. Gupta and P. Kumar, "The capacity of wireless networks," *IEEE Transactions on Information Theory*, vol. 46, no. 2, pp. 388 – 404, March 2000.

[4] M. Grossglauser and D. N. Tse, "Mobility increases the capacity of ad hoc wireless networks," *IEEE Transactions on Networking*, vol. 10, no. 4, pp. 477–486, August 2002.

[5] M. Gastpar and M. Vetterli, "On the capacity of wireless networks: the relay case," in *Proc. of the 21st Annual Joint Conference of the IEEE Computer and Communications Societies (InfoCom)*, vol. 3, New York, USA, June 23-27 2002, pp. 1577–1586.

[6] B. Liu, Z. liu, and D. Towsley, "On the capacity of hybrid wireless networks," in *Proc. of the 22st Annual Joint Conference of the IEEE Computer and Communications Societies (InfoCom)*, vol. 2, San Francisco, USA, April 1-3 2003, pp. 1543 – 1552.

[7] S. Kulkarn and P. Viswanath, "A deterministic approach to throughput scaling in wireless networks," *IEEE Transactions on Information Theory*, vol. 50, pp. 1041–1049, 2004.

[8] L. Backstrom, E. Sun, and C. Marlow, "Find me if you can: Improving geographical prediction with social and spatial proximity," in *Proceedings of the WWW2010*, Raleigh, USA, April 26 - 30 2010.

[9] B. Latane, J. H. Liu, A. Nowak, M. Bonevento, and L. Zheng, "Distance matters: Physical space and social impact," *Personality and Social Psychology Bulletin*, vol. 21, p. 795, 1995.

[10] R. Albert and A.-L. Barabasi, "Statistical mechanics of complex networks," *Reviews of Modern Physics*, vol. 74, pp. 47 – 97, January 2002.

[11] M. E. J. Newman, "The structure and function of complex networks," *SIAM Review*, vol. 54, pp. 167 – 256, 2003.

[12] S. Boccalettia, V. Latorab, Y. Morenod, M. Chavez, and D. Hwang, "Complex networks: Structure and dynamics," *Elsevier Physics Reports*, vol. 424, pp. 175–308, February 2006.

[13] M. Faloutsos, P. Faloutsos, and C. Faloutsos, "On power-law relationships of the internet topology," in *Proceedings of the conference on Applications, technologies, architectures, and protocols for computer communication (SIGCOMM)*, vol. 29, Cambridge, Massachusetts, USA, October 1999.

[14] B. Azimdoost, H. R. Sadjadpour, and J. Garcia-Luna, "Capacity of composite networks: Combining social and wireless ad hoc networks," in *Proceedings of IEEE Wireless Communications and Networking Conference (WCNC)*, Cancun, Mexico, March 28 - 31 2011.

[15] ——, "Capacity of wireless networks with social behavior," in *IEEE Transaction on Wireless Communications*, January, vol. 12, no 1 2013.

[16] M. D. Penrose, "The longest edge of the random minimal spanning tree," *The Annals of Applied Probability*, vol. 7, no. 2, pp. 340 – 361, 1997.

[17] F. Xue and P. R. Kumar, *Scaling Laws for Ad Hoc Wireless Networks: An Information Theoretic Approach*. Now Publishers, 2006.

[18] T. P. Mitev, "New inequalities between elementary symmetric polynomials," *Journal of Inequalities in Pure and Applied Mathematics*, vol. 4, no. 2, 2003.

[19] W. Willinger, D. Alderson, and J. C. Doyle, "Mathematics and the internet: A source of enormous confusion and great potential," *Notices American Mathematical Society (AMS)*, vol. 56, no. 5, pp. 586 – 599, May 2009.

[20] M. K. Kiskani, H. R. Sadjadpour, and M. Guizani, "Social interaction increases capacity of wireless networks," in *Proceedings of the 9th IEEE conference on Wireless Communications and Mobile Computing Conference (IWCMC)*, 2013, pp. 467 – 472.

[21] B. Azimdoost and H. R. Sadjadpour, "Capacity of scale free wireless networks," in *Proceedings of the IEEE Global Communications Conference (GLOBECOM)*, 2012, pp. 2379 – 2384.

VIII. Appendix

Lemma 2. *When $\lim_{n\to\infty} q = \infty$, we have $\frac{d_k^{-\alpha}\sigma_{q-1}(\mathbf{d_n^{\bar{k}}})}{\sigma_q(\mathbf{d_n})} \equiv \frac{q}{n}$. specifically, when $q = \Theta(n)$, we have $\frac{d_k^{-\alpha}\sigma_{q-1}(\mathbf{d_n^{\bar{k}}})}{\sigma_q(\mathbf{d_n})} \equiv \Theta(1)$.*

Proof: Define the random variables $Y_i = d_i^{-\alpha}$ and $Z_i = \log Y_i$ for $1 \le i \le n$. Since Y_i's are i.i.d random variables, Z_i's are also i.i.d. random variables. By using the law of large numbers, we have $\lim_{m\to\infty} \frac{1}{m}\sum_{i=1}^m Z_i = \bar{Z}$ where $\bar{Z}$ is the expected value of the random variable Z_i. Hence,

$$\begin{aligned}
\frac{d_k^{-\alpha}\sigma_{q-1}(\mathbf{d_n^{\bar{k}}})}{\sigma_q(\mathbf{d_n})} &\equiv \frac{\sum_{1\le i_1<..<i_q\le n, \exists h: i_h=k} \prod_{j=1}^q Y_{i_j}}{\sum_{1\le i_1<..<i_q\le n} \prod_{j=1}^q Y_{i_j}} \\
&\equiv \frac{\sum_{1\le i_1<..<i_q\le n, \exists h: i_h=k} \exp \sum_{j=1}^q Z_{i_j}}{\sum_{1\le i_1<..<i_q\le n} \exp \sum_{j=1}^q Z_{i_j}} \\
&\equiv \frac{\sum_{1\le i_1<..<i_q\le n, \exists h: i_h=k} \exp q\overline{Z}}{\sum_{1\le i_1<..<i_q\le n} \exp q\overline{Z}} \\
&\equiv \frac{\binom{n-1}{q-1}}{\binom{n}{q}} = \frac{q}{n} \qquad (20)
\end{aligned}$$

Now if $q = \Theta(n)$, we have $\frac{q}{n} \equiv \Theta(1)$ and therefore, $\frac{d_k^{-\alpha}\sigma_{q-1}(\mathbf{d_n^{\bar{k}}})}{\sigma_q(\mathbf{d_n})} \equiv \frac{q}{n} \equiv \Theta(1)$. ■

Lemma 3.

$$\sum_{x=1}^{\frac{1}{r(n)}} x \sum_{l=1}^{4x} \sum_{v_k\in s_l} d_k^{-\beta} \equiv \begin{cases} \Theta\left(nr^{-1}(n)\right), & 0 \le \beta \le 3 \\ \Theta\left(nr^{2-\beta}(n)\right), & 3 \le \beta \end{cases} \qquad (21)$$

Proof: The proof can be found in [20]. ■

Lemma 4.

$$\sigma_1(\mathbf{d_n}) \equiv \begin{cases} \Theta(n), & 0 \le \alpha \le 2 \\ \Theta\left(nr^{2-\alpha}(n)\right), & 2 \le \alpha \end{cases} \qquad (22)$$

Also notice that when $q = \Theta(n)$, we have

$$\sigma_1(\mathbf{d_q}) \equiv \begin{cases} \Theta(n), & 0 \le \beta \le 2 \\ \Theta\left(nr^{2-\beta}(n)\right). & 2 \le \beta \end{cases} \qquad (23)$$

Proof: The proof can be found in [20]. ■

Proof of theorem 1: This theorem is a direct result of lemmas 2 and 3. ■

Lemma 5. *Let $\mathbf{\Psi} = \{\psi_1, ..., \psi_n\}$ be a set of $n \ge 2$ non-negative real numbers. Then for $1 \le p \le n-1$ we have $\sigma_1(\mathbf{\Psi})\sigma_p(\mathbf{\Psi}) \ge \frac{n(p+1)}{n-p}\sigma_{p+1}(\mathbf{\Psi})$.*

Proof: The proof can be found in [18]. ■

Lemma 6. *If $\mathbf{\Psi} = \{\psi_1, ..., \psi_n\}$ be a set of $n \ge 2$ non-negative real numbers, then for a finite p and when $n \to \infty$, we have $\frac{\sigma_1(\mathbf{\Psi})\sigma_p(\mathbf{\Psi})}{(p+1)\sigma_{p+1}(\mathbf{\Psi})} = \Theta\left(\frac{n}{n-p}\right) = \Theta(1)$.*

Proof: This is the lemma 4.1 in [15]. The proof can be found there. ■

Lemma 7. *When $q = \Theta(1)$ or $q = \Theta(g(n))$ where $\lim_{n\to\infty} \frac{g(n)}{n} = 0$, then $\sigma_1(\mathbf{d_q})$ has the order of $\Theta\left(r(n)^{-\beta}\right)$.*

Proof: Suppose that the i-th member of the long-range social group is located in the distance of x_{q_i} hops from the source, then we can say that

$$\sigma_1(\mathbf{d_q}) = \sum_{i=1}^q \left(c_i r(n) x_{q_i}\right)^{-\beta} = (r(n))^{-\beta} \sum_{i=1}^q \left(c_i x_{q_i}\right)^{-\beta}. \quad (24)$$

Since, x_{q_i} can be every integer between one and $\frac{1}{r(n)}$, the order of $\sigma_1(\mathbf{d_q})$ may range from $\Theta(1)$ to $\Theta(r(n)^{-\beta})$. However, note that when n goes to infinity, with probability approaching one at least one of the long-range contacts lies within a lattice distance of $\Theta(1)$ to the source.

To prove this, it is enough to show that with probability approaching zero, all of the long-range contacts lie outside a lattice distance of $f(n) = \Omega(1)$ to the source. Assuming $q = \Theta(1)$ or $q = \Theta(g(n))$ where $\lim_{n\to\infty} \frac{g(n)}{n} = 0$, we can argue that the probability of selecting long-range social contacts is independent of each other. Thus, using lemma 4 we have

$$\begin{aligned}
&\Pr(x_{q_1} = \Theta(f(n)), x_{q_2} = \Theta(f(n)), ..., x_{q_q} = \Theta(f(n))) \\
=& \prod_{i=1}^q \Pr(x_{q_i}\Theta(f(n)) \equiv \prod_{i=1}^q \frac{(f(n)r(n))^{-\alpha}}{\sigma_1(\mathbf{d_n})} \equiv O(\frac{(r(n))^{-q\alpha}}{(\sigma_1(\mathbf{d_n}))^q}) \\
\equiv& \begin{cases} O\left((nr^\alpha(n))^{-q}\right), & 0 \le \alpha \le 2 \\ O\left((\log n)^{-q}\right). & 2 \le \alpha \end{cases}
\end{aligned}$$

It is not difficult to verify that the right hand side which is an upper bound for this probability goes to zero as n approaches infinity thus the aforementioned probability tends to zero. Thus with probability approaching one, there exists at least one long-range contact in the lattice distance of $\Theta(1)$ to the source which will be the dominant term in $\sigma_1(\mathbf{d_q})$. Therefore, $\sigma_1(\mathbf{d_q}) = \Omega\left((r(n))^{-\beta}\right)$ and since in the case of $q = \Theta(1)$, $\sigma_1(\mathbf{d_q})$ is only composed of $\Theta(1)$ terms we have $\sigma_1(\mathbf{d_q}) = \Theta\left((r(n))^{-\beta}\right)$.

For the case of $q = \Theta(g(n))$ we know that at least one long range contact exists within a distance of $\Theta(1)$ to the source. Therefore, $\sigma_1(\mathbf{d_q})$ can have the order of $\Theta\left((r(n))^{-\beta}\right)$ when it only has $\Theta(1)$ social contacts within a distance of $\Theta(1)$ to the source or it can have the order of $\Theta\left(g(n)(r(n))^{-\beta}\right)$ when almost all of the $\Theta(g(n))$ social contacts lie within a distance of $\Theta(1)$ to the source. We will now show that with a probability close to one the latter almost never happens and therefore in the case of $q = \Theta(g(n))$, almost surely we have, $\sigma_1(\mathbf{d_q}) = \Theta\left((r(n))^{-\beta}\right)$. To prove this, using the same approach as above, we will compute the probability that almost all of the social contacts lie within a distance $\Theta(1)$ to the source,

$$\begin{aligned}
&\Pr(x_{q_1} = \Theta(1), x_{q_2} = \Theta(1), ..., x_{q_q} = \Theta(1)) \\
=& \prod_{i=1}^q \Pr(x_{q_i} = \Theta(1)) \equiv \prod_{i=1}^q \frac{(r(n))^{-\alpha}}{\sigma_1(\mathbf{d_n})} \\
\equiv& \Theta(\frac{(r(n))^{-q\alpha}}{(\sigma_1(\mathbf{d_n}))^q}) = \begin{cases} \Theta\left((nr^\alpha(n))^{-q}\right), & 0 \le \alpha \le 2 \\ \Theta\left((\log n)^{-q}\right). & 2 \le \alpha \end{cases}
\end{aligned}$$

When n is a large number, this probability goes to zero and therefore this scenario almost surely never happens. ■

Lemma 8. *The following inequalities hold.*

$$\sigma_{q-1}(\mathbf{d_n}) - d_k^{-\alpha}\sigma_{q-2}(\mathbf{d_n}) \leq \sigma_{q-1}(\mathbf{d_n^{\bar{k}}}) \leq \sigma_{q-1}(\mathbf{d_n}) \quad (25)$$

Proof: This lemma can be proved by expanding the polynomials and considering the non-negativity of elements in $\mathbf{d_n}$. We will use this lemma to find the upper and lower bounds for $E[X]$. ■

Proof of theorem 3: We can use lemma 2 to simplify equation (7) as

$$E[X] \equiv \frac{q}{n\sigma_1(\mathbf{d_q})} \sum_{x=1}^{\frac{1}{r(n)}} x \sum_{l=1}^{4x} \sum_{v_k \in s_l} d_k^{-\beta}. \quad (26)$$

Now, using lemmas 3 and 7 and replacing q with $\Theta(f(n))$ proves the theorem. ■

Proof of theorem 4: Since each node can receive or transmit just one flow at a time, the maximum rate a node (and a cell) can support is $\Theta(1)$. Each node carries traffic during transmission, reception, or relaying of the data. The maximum value of this traffic should not exceed the maximum supportable traffic of $\Theta(1)$. We will consider three different scenarios:

(i) Nodes in transmission mode: Each node transmit at maximum rate of λ_{max} which is much less than one for all the obtained capacity regions. It has been shown [7] that there are $\Theta(\log n)$ nodes in each cell which results in maximum generated traffic by each cell as $\Theta(\lambda_{max} \log n)$. Since λ_{max} does not exceed $\Theta(\frac{1}{\log n})$, then the maximum traffic generated by each cell cannot exceed $\Theta(1)$. Therefore, the traffic generate in transmission mode does not create any bottleneck.

(ii) Nodes in relay mode: A path of length x-hops consists of exactly x cells in our model. Since we have a total of $\frac{1}{r^2(n)}$ cells, the probability that a cell is selected from a group of x specific cells is equal to $xr^2(n)$. The probability that a source-destination path of length x-hops passes through a specific cell is always less than $xr^2(n)$. Thus, the probability of a source-destination path L_i passing through a specific cell S_0 is

$$\begin{aligned} & \Pr(L_i \text{ intersects } S_0) \\ = & \sum_x \Pr(L_i \text{ intersects } S_0 | X_i = x)\Pr(X_i = x) \\ \leq & \sum_x xr^2(n)\Pr(X_i = x), \end{aligned}$$

where X_i is the number of hops the path L_i is passing through. Therefore, $\Pr(L_i \text{ intersects } S_0) \leq E[X]r^2(n)$. Since we only consider unicast communications, there are at most a total of $\Theta(n)$ source-destination pairs. Therefore, using the union bound, the maximum number of paths intersecting a specific cell is $\Theta(nE[X]r^2(n))$. Consequently, the maximum traffic load of a relay cell is $\Theta(nE[X]r^2(n)\lambda_{max})$ which is $\Theta(1)$ in all regions of the throughput capacity obtained in this paper. Therefore no cell will carry more than what it can support when it is in relay mode.

A relay node in a cell consisting of $\Theta(\log n)$ nodes is selected with a uniform distribution. Hence, the probability that a specific node is a relay equals the probability that the corresponding cell is a relay, divided by the number of nodes in that cell. This probability is smaller than $\Theta(E[X]\lambda_{max})$ which is less than $\Theta(1)$. It is concluded that the relay nodes will never cause bottleneck in the network.

(iii) Node is in receive mode: Similar to previous section argument, we conclude that receiver cells do not cause bottleneck in the network. Since the selection of friends for each node follows power-law distribution that may make the distribution of the destination nodes non-uniform. In case of $q = \Theta(1)$, each node has only $q = \Theta(1)$ social contacts and it consumes a constant bandwidth and does not cause bottleneck. For $q = \Theta(n)$, we prove that this distribution is still uniform for large n and similar to the relay nodes, the destination nodes does not create any bottleneck.

The source nodes are uniformly distributed in the network. Thus the probability that a specific node v_k is the destination can be written as

$$\begin{aligned} \Pr(T = v_k) &= \sum_{i=1}^{n} \Pr(T = v_k | v_i \text{ is source})\Pr(v_i \text{ is source}), \\ &= \frac{1}{n}\sum_{i=1}^{n} \Pr(T = v_k | v_i \text{ is source}). \end{aligned}$$

Let d_{ki} be the distance between v_k and v_i and G_i be the set of social contacts if node v_i is the source. Let's define $\mathbf{d_{q_i}} = (d_{g_1 i}^{-\beta}, ..., d_{g_q i}^{-\beta})$ and $\mathbf{d_{n_i}} = (d_{g_1 i}^{-\alpha}, ..., d_{g_n i}^{-\alpha})$. Now, similar to equations (2) and (4) which has been written for one specific source node, we have

$$\begin{aligned} & \Pr(T = v_k | v_i \text{ is source}) \\ = & \Pr(T = v_k | v_i \text{ is source}, v_k \in G_i)\Pr(v_k \in G_i) \\ = & \frac{d_{ki}^{-\beta}}{\sigma_1(\mathbf{d_{q_i}})} \frac{d_{ki}^{-\alpha}\sigma_{q-1}(\mathbf{d_{n_i}^{\bar{k}}})}{\sigma_q(\mathbf{d_{n_i}})}. \end{aligned}$$

Using lemma 2, $\Pr(T = v_k | v_i \text{ is source}) \equiv \frac{d_{ki}^{-\beta}}{\sigma_1(\mathbf{d_{q_i}})}$. Therefore,

$$\begin{aligned} \Pr(T = v_k) &= \frac{1}{n}\sum_{i=1}^{n} \Pr(T = v_k | v_i \text{ is source}) \\ &= \frac{1}{n}\sum_{i=1}^{n} \frac{d_{ki}^{-\beta}}{\sigma_1(\mathbf{d_{q_i}})} = \frac{1}{n}. \end{aligned} \quad (27)$$

So the destinations are distributed uniformly similar to the relay nodes, and no node in receive mode will be a bottleneck. Notice that since for the case of $q = \Theta(n)$ no node will become bottleneck, for the case of $q = \Theta(f(n))$ also no node will become bottleneck when $f(n) = O(n)$ as in our case. ■

Proof of theorem 7: The proof of this theorem is very similar to the proof of theorem 4. For relay and transmit modes we can readily use the same proof as in theorem 4. For receive mode, we only need to prove that the destinations will have a uniform distribution.

The source nodes are uniformly distributed in the network. Thus the probability that a specific node v_k is the destination can be written as

$$\begin{aligned}\Pr(T=v_k) &= \sum_{i=1}^{n}\Pr(T=v_k|v_i\text{is source})\Pr(v_i \text{ is source})\\ &= \frac{1}{n}\sum_{i=1}^{n}\Pr(T=v_k|v_i\text{is source}).\end{aligned}$$

Let G_i be the set of social contacts if node v_i is the source, and Q_i be the number of social contacts of source node v_i. Using equations (2) and (4) which has been written for one specific source node, we have

$$\begin{aligned}&\Pr(T=v_k|v_i \text{ is source})\\ &= \sum_{q=1}^{n}\Pr(T=v_k|v_i \text{ is source}, v_k\in G_i, Q_i=q)\\ &\qquad \times \Pr(v_k\in G_i, Q_i=q)\\ &= \sum_{q=1}^{n}\Pr(T=v_k|v_i \text{ is source}, v_k\in G_i, Q_i=q)\\ &\qquad \times \Pr(v_k\in G_i|Q_i=q)\Pr(Q_i=q)\\ &= \sum_{q=1}^{n}\frac{q^{-\gamma}}{q\sigma_1(\mathbf{b})}\frac{d_k^{-\alpha}\sigma_{q-1}(\mathbf{d}_\mathbf{n}^{\bar{\mathbf{k}}})}{\sigma_q(\mathbf{d_n})} \qquad (28)\end{aligned}$$

Now let P_1 and P_2 represent $\sum_{q=q_0+1}^{n}\frac{q^{-\gamma}}{q\sigma_1(\mathbf{b})}\frac{d_k^{-\alpha}\sigma_{q-1}(\mathbf{d}_\mathbf{n}^{\bar{\mathbf{k}}})}{\sigma_q(\mathbf{d_n})}$ and $\sum_{q=1}^{q_0}\frac{q^{-\gamma}}{q\sigma_1(\mathbf{b})}\frac{d_k^{-\alpha}\sigma_{q-1}(\mathbf{d}_\mathbf{n}^{\bar{\mathbf{k}}})}{\sigma_q(\mathbf{d_n})}$, respectively. Using the results from Theorem 2 for $q>q_0$ [20], we have $\frac{d_k^{-\alpha}\sigma_{q-1}(\mathbf{d}_\mathbf{n}^{\bar{\mathbf{k}}})}{\sigma_q(\mathbf{d_n})}\equiv\frac{q}{n}$. Also using results from [21] and lemma 6 for $q\le q_0$ we have $\frac{d_k^{-\alpha}\sigma_{q-1}(\mathbf{d}_\mathbf{n}^{\bar{\mathbf{k}}})}{\sigma_q(\mathbf{d_n})}<\frac{d_k^{-\alpha}q}{\sigma_1(\mathbf{d_n})}$. Therefore,

$$\begin{aligned}P_1 &= \Theta(\frac{1}{n\sigma_1(\mathbf{b})}\sum_{q=q_0+1}^{n}q^{-\gamma})\\ P_2 &= O(\frac{d_k^{-\alpha}}{\sigma_1(\mathbf{b})\sigma_1(\mathbf{d_n})}\sum_{q=1}^{q_0}q^{-\gamma}) \qquad (29)\end{aligned}$$

For large values of γ, $\sigma_1(\mathbf{b})$, $\sum_{q=q_0+1}^{n}q^{-\gamma}$, and $\sum_{q=1}^{q_0}q^{-\gamma}$ are all $\Theta(1)$. Hence, we have $P_1\equiv\frac{1}{n}$ and $P_2=O(\frac{d_k^{-\alpha}}{\sigma_1(\mathbf{d_n})})$. Then,

$$\begin{aligned}\Pr(T=v_k) &= \frac{1}{n}\sum_{i=1}^{n}(P_1+P_2)\\ &= \frac{1}{n}\sum_{i=1}^{n}\Theta(\frac{1}{n})+O(\frac{d_k^{-\alpha}}{\sigma_1(\mathbf{d_n})}), \qquad (30)\end{aligned}$$

where d_k in the above formulation is the distance from v_k to the source node v_i which can be shown as d_{k_i}. Thus using similar notation for $\mathbf{d_{n_i}}$ we have

$$\sum_{i=1}^{n}O(\frac{d_k^{-\alpha}}{\sigma_1(\mathbf{d_{n_i}})})=O(\frac{\sigma_1(\mathbf{d_{n_k}})}{\sigma_1(\mathbf{d_{n_i}})})=O(1), \qquad (31)$$

that results in $\Pr(T=v_k)\equiv\frac{1}{n}$. Therefore, the destinations are distributed uniformly similar to the relay nodes, and no node in receive mode will be a bottleneck. ■

Proof of theorem 8: To simplify the equation (17), like the process in the proof of theorem 5, we break $E[X]$ into the following two parts,

$$E_1 \equiv \sum_{x=1}^{\frac{1}{r(n)}}x\sum_{l=1}^{4x}\sum_{v_k\in s_l}\sum_{q=q_0}^{n-1}\frac{q^{-\gamma}d_k^{-\alpha-\beta}\sigma_{q-1}(\mathbf{d}_\mathbf{n}^{\bar{\mathbf{k}}})}{\sigma_1(\mathbf{b})\sigma_1(\mathbf{d_q})\sigma_q(\mathbf{d_n})} \qquad (32)$$

and

$$E_2 \equiv \sum_{x=1}^{\frac{1}{r(n)}}x\sum_{l=1}^{4x}\sum_{v_k\in s_l}\sum_{q=1}^{q_0-1}\frac{q^{-\gamma}d_k^{-\alpha-\beta}\sigma_{q-1}(\mathbf{d}_\mathbf{n}^{\bar{\mathbf{k}}})}{\sigma_1(\mathbf{b})\sigma_1(\mathbf{d_q})\sigma_q(\mathbf{d_n})} \qquad (33)$$

We can use the argument in the proof of theorem 5 to simplify E_1 as

$$\begin{aligned}E_1 &\equiv \sum_{x=1}^{\frac{1}{r(n)}}x\sum_{l=1}^{4x}\sum_{v_k\in s_l}\sum_{q=q_0}^{n}\frac{q^{-\gamma}d_k^{-\beta}}{\sigma_1(\mathbf{b})\sigma_1(\mathbf{d_q})}\frac{q}{n}\\ &\equiv \sum_{x=1}^{\frac{1}{r(n)}}x\sum_{l=1}^{4x}\sum_{v_k\in s_l}\frac{d_k^{-\beta}}{n\sigma_1(\mathbf{b})}\sum_{q=q_0}^{n}\frac{q^{-\gamma+1}}{\sigma_1(\mathbf{d_q})} \qquad (34)\end{aligned}$$

Since q_0 is a very large number, law of large numbers ensures that $\frac{1}{q}\sigma_1(\mathbf{d_q})$ lies in the interval $(E[\mathbf{d_q}]-\epsilon, E[\mathbf{d_q}]+\epsilon)$ with probability one thus it can be replaced by $E[\mathbf{d_q}]$ in our work.

$$\sum_{q=q_0}^{n}\frac{q^{-\gamma+1}}{\sigma_1(\mathbf{d_q})}=\frac{1}{E[\mathbf{d_q}]}\sum_{q=q_0}^{n}q^{-\gamma} \qquad (35)$$

To find $E[\mathbf{d_q}]$ notice that according to the proof of lemma 7 we know that with probability close to one when n approaches infinity, there exists a long-range social contact within the lattice distance of $\Theta(1)$ from the source thus

$$E[\mathbf{d_q}]\equiv\sum_{x=1}^{\frac{1}{r(n)}}\Pr(X=x)(xr(n))^{-\beta}\equiv(r(n))^{-\beta} \qquad (36)$$

Now if $\gamma>1$, we have $\sum_{q=q_0}^{n}q^{-\gamma}\le\sigma_{1,n}(\mathbf{b}^{-\gamma})\le\sum_{q=1}^{\infty}q^{-\gamma}=\zeta(\gamma)\equiv\Theta(1)$. Therefore (35) can be simplified to

$$\sum_{q=q_0}^{n}\frac{q^{-\gamma+1}}{\sigma_1(\mathbf{d_q})}=(r(n))^{\beta} \qquad (37)$$

and if $0\le\gamma\le1$ we have $\sum_{q=q_0}^{n}q^{-\gamma}\equiv\sigma_{1,n}(\mathbf{b}^{-\gamma})\equiv\dfrac{n^{-\gamma+1}}{-\gamma+1}\equiv n^{-\gamma+1}$. Thus in this case (35) simplifies to

$$\sum_{q=q_0}^{n}\frac{q^{-\gamma+1}}{\sigma_1(\mathbf{d_q})}\equiv n^{-\gamma+1}(r(n))^{\beta} \qquad (38)$$

Using the previous equations of (21) and (22)

$$\begin{aligned}&\frac{1}{n\sigma_1(\mathbf{b})}\sum_{x=1}^{\frac{1}{r(n)}}x\sum_{l=1}^{4x}\sum_{v_k\in s_l}d_k^{-\beta}\\ &\equiv \begin{cases}\Theta\left(n^{\gamma-1}r(n)^{-1}\right), & 0\le\beta\le3,\ 0\le\gamma\le1\\ \Theta\left(n^{\gamma-1}r(n)^{2-\beta}\right), & 3\le\beta, \quad 0\le\gamma\le1\\ \Theta\left(r(n)^{-1}\right), & 0\le\beta\le3, \quad \gamma>1\\ \Theta\left(r(n)^{2-\beta}\right), & 3\le\beta, \quad \gamma>1\end{cases} \qquad (39)\end{aligned}$$

Therefore using (37), (38) and (39) we have

$$E_1 \equiv \begin{cases} \Theta\left(r(n)^{-1+\beta}\right), & 0 \le \beta \le 3 \\ \Theta\left(r(n)^2\right), & 3 \le \beta \end{cases} \equiv \begin{cases} \Theta\left(r(n)^{-1+\beta}\right), & 0 \le \beta \le 1 \\ \Theta(1), & 1 \le \beta \end{cases} \quad (40)$$

Notice that since $E[X]$ cannot be smaller than one, thus we can replace $r(n)^{-1+\beta}$ for $1 \le \beta \le 3$, and $r(n)^2$ with 1, thus, the second equality holds. Now we use lemma 8 and equation (25) to prove that the order of E_1 is dominant in the summation $E[X] = E_1 + E_2$. Using the right hand side of (25) we have

$$E_2 \le \sum_{x=1}^{\frac{1}{r(n)}} x \sum_{l=1}^{4x} \sum_{v_k \in s_l} \sum_{q=1}^{q_0-1} \frac{q^{-\gamma} d_k^{-\alpha-\beta} \sigma_{q-1}(\mathbf{d_n})}{\sigma_1(\mathbf{b})\sigma_1(\mathbf{d_q})\sigma_q(\mathbf{d_n})} \quad (41)$$

Since $q \le q_0$, it is a finite number and we can use lemma 5 to get

$$\frac{\sigma_{q-1}(\mathbf{d_n})}{\sigma_q(\mathbf{d_n})} \equiv \frac{1}{\sigma_1(\mathbf{d_n})}\Theta(\frac{nq}{n-q+1}) \equiv \frac{1}{\sigma_1(\mathbf{d_n})} \quad (42)$$

Thus

$$E_2 \le \sum_{x=1}^{\frac{1}{r(n)}} x \sum_{l=1}^{4x} \sum_{v_k \in s_l} \frac{d_k^{-\alpha-\beta}}{\sigma_1(\mathbf{d_n})\sigma_1(\mathbf{b})} \sum_{q=1}^{q_0-1} \frac{q^{-\gamma}}{\sigma_1(\mathbf{d_q})}. \quad (43)$$

Notice that using the argument in the proof of lemma 7 for very large n, there exists a long-range contact in the lattice distance of $\Theta(1)$ to the source, with high probability, which will be the dominant term in the summation $\sigma_1(\mathbf{d_q})$ thus $\sigma_1(\mathbf{d_q})$ scales as $r(n)^{-\beta}$ and hence,

$$\sum_{q=1}^{q_0-1} \frac{q^{-\gamma}}{\sigma_1(\mathbf{d_q})} \equiv \frac{1}{r(n)^{-\beta}} \sum_{q=1}^{q_0-1} q^{-\gamma} \equiv r(n)^{\beta} \quad (44)$$

Therefore, $E_2 \le \frac{r(n)^{\beta}}{\sigma_1(\mathbf{d_n})\sigma_1(\mathbf{b})} \sum_{x=1}^{\frac{1}{r(n)}} x \sum_{l=1}^{4x} \sum_{v_k \in s_l} d_k^{-\alpha-\beta}$. Thus for $\gamma > 1$ we have

$$E_2 \equiv \begin{cases} O\left(r(n)^{-1+\beta}\right), & 0 \le \alpha+\beta \le 3, 0 \le \alpha \le 2 \\ O\left(r(n)^{\alpha+\beta-3}\right), & 0 \le \alpha+\beta \le 3, \alpha \ge 2 \\ O\left(r(n)^{2-\alpha}\right), & 3 \le \alpha+\beta, 0 \le \alpha \le 2 \\ O(1), & 3 \le \alpha+\beta, \alpha \ge 2 \end{cases} \quad (45)$$

and for $0 \le \gamma \le 1$, E_2 will have a scaling factor of $n^{1-\gamma}$ multiplied by the above equation. It can be verified that E_1 is the dominant term compared to E_2 and therefore theorem 8 is proved. ■

Mohsen Karimzadeh Kiskani received his bachelors degree in mechanical engineering from Sharif University of Technology in 2008. He got his masters degree in electrical engineering from Sharif University of Technology in 2010. He is currently a PhD candidate in electrical engineering department at University of California, Santa Cruz. His main areas of interest include wireless communications, wireless social networks, and applications of network coding and multi hop index coding in wireless and caching networks. He is also interested in the complexity study of Constraint Satisfaction Problems (CSP) in computer science.

Bita Azimdoost received her B.S. and M.S. degrees from Sharif University of Technology in 1998 and 2000, respectively. Since 2010, she has been pursuing Ph.D. degree with the Department of Electrical Engineering at University of California, Santa Cruz, USA. Her research interests span over wireless ad hoc networks, social networking, and new emerging architectures including information-centric networks.

Hamid Sadjadpour (S94-M95-SM00) received his B.S. and M.S. degrees from Sharif University of Technology and Ph.D. degree from University of Southern California in, respectively. After graduation, he joined AT&T as a member of technical staff and finally Principal member of technical staff at AT&T research Lab. in Florham Park, NJ until 2001. In fall 2001, he joined University of California, Santa Cruz (UCSC) where he is now a Professor. Dr. Sadjadpour has served as technical program committee member and chair in numerous conferences. He has been also Guest editor of EURASIP in 2003 and 2006, and is currently member of Editorial board of Wiley Wireless Communications and Mobile Computing Journal (WCMC). He has published more than 170 publications and has awarded 17 patents. His research interests are in the general area of wireless communications and networks. He is the co-recipient of best paper awards at 2007 International Symposium on Performance Evaluation of Computer and Telecommunication Systems (SPECTS), 2008 Military Communications (MILCOM) conference, and 2010 European Wireless Conference.