The Impact of Social Groups on The Capacity of Wireless Networks

Bita Azimdoost
Department of Electrical Engineering
University of California
Santa Cruz 95064
Email: bazimdoost@soe.ucsc.edu

Hamid R. Sadjadpour
Department of Electrical Engineering
University of California
Santa Cruz 95064
Email: hamid@soe.ucsc.edu

J.J. Garcia-Luna-Aceves
Department of Computer Engineering
University of California
Santa Cruz 95064
Email: jj@soe.ucsc.edu

***Abstract*—The capacity of a wireless network with n nodes is studied when nodes communicate with one another in the context of social groups. Each node is assumed to have at least one local contact in each of the four directions of the plane in which the wireless network operates, and $q(n)$ independent long-range social contacts forming its social group, one of which it selects randomly as its destination. The distance between source and the members of its social group follows a power-law distribution with parameter α, and communication between any two nodes takes place only within the physical transmission range; hence, source-destination communication takes place over multi-hop paths. The order capacity of such a composite network is derived as a function of the number of nodes (n), the social-group concentration (α), and the size of social groups ($q(n)$). It is shown that the maximum order capacity is attained when $\alpha \geq 3$, which makes social groups localized geographically, and that a wireless network can be scale-free when social groups are localized and independent of the number of nodes in the network, i.e., $q(n)$ is independent of n.**

***Index Terms*—Capacity, Social Networks, Communication Networks, Complex Networks.**

I. Introduction

Starting with the work by Gupta and Kumar [1], the order throughput capacity of wireless communication networks has been studied extensively in the recent past, and all this prior work has assumed that sources select their destinations according to a uniform distribution. However, in real wireless networks, the selection of destinations by sources does not follow a uniform distribution, because nodes interact with one another in the context of social groups. It has been observed [2] that the likelihood of having contact with a person decreases with distance and follows a power-law distribution. Consequently, it can be argued that prior results on the order capacity of wireless networks are overly pessimistic, because they inherently assume that source-destination flows involve routes that consume too many communication resources.

On the other hand, as a result of early work by Milgram on the small-world phenomenon [3], the modeling of social networks have received considerable attention. However, this work has not addressed the underlying limitations imposed by the physical layer. For example, Watts and Strogatz [4] divide the edges of a network into local and long-range contacts and assume that there is always an edge between a node and any of its local or long-range social contacts. Dietzfelbinger et al. [7] calculated the average number of steps between any source and target along a ring-based network in which each node is connected to its left and right neighbor and possibly to some further vertices, and the long-range contacts may be selected through any distribution. Fraigniaud et al. [6] assumed that the probability of a node being the long-range contact of a source is proportional to the rank of their distance among the distances from the source to all the other nodes. Kleinberg [5] introduced a model for the characterization of the small-world phenomenon consisting of a two-dimensional extended grid with point-to-point links in which each node has four local contacts and one long-range contact. The source node s selects any other node v as its long-range contact with a probability proportional to $d^{-\alpha}(s,v)$, where $d(s,v)$ is the lattice distance between s and v, and $\alpha \geq 0$ shows the density of the social network. Given that these models neglect the need to consume the resources of the multi-hop paths needed to connect sources with remote destinations in real networks, it can be argued that they render an overly optimistic view of the capacity of social networks.

What is needed to understand the true performance of wireless networks is a model that captures the restrictions imposed by the communication infrastructure, together with the distribution of flows rendered by social groups. In this regard, Li et al. [8] studied the capacity of a wireless network in which source-destination pairs follow a power-law distribution as in Kleinberg's model; however, they provide only upper bounds that need not be tight and provide no insight on the impact of social-group sizes. More recently, Azimdoost et al. [9] studied the interaction between communication and social networks by considering four local contacts and a single long-range contact per node, with the source knowing the location of its four local contacts and the destination. The source-destination pair selection follows a power-law distribution that is a function of the Euclidean distance between the source and the rest of the nodes. While this model is a marked improvement over models that assume a uniform distribution for source-destination pairs, its results are limited in scope because a node usually has more than one long-range social contact in its social group.

In this paper, we study the case of a wireless network in which nodes communicate with others in the context of social groups. Section II introduces the notation and some

978-1-4577-1051-3/11/$26.00 ©2011 IEEE

definitions and results used throughout the paper. If a source communicates with other nodes in the context of a social group, it has multiple remote social contacts, not just one.

Section III shows that the original power-law distribution introduced by Kleinberg [5] cannot be used when the number of long-range contacts q is a function of total number of nodes n in the network. In fact, this limitation was also mentioned by Kleinberg [5] . In addition, a modified power-law distribution is introduced that is applicable for all values of $q(n)$.

The main contribution of this paper is stated in the following theorem, which considers what we call a *wireless social network* . In such a network, each of the n network nodes has a social group consisting of at least one local social contact in each of the four directions of the plane, and $q(n)$ long-range contacts selected independently. Long-range social contacts are selected based on the power-law distribution with parameter α identified in Section III, and one of those long-range contacts is the destination of the node's flow. Communication between any two nodes can take place only if they are within transmission range and interference and such communication succeeds according to the protocol model of multiple access interference [1].

Theorem 1. *The maximum achievable capacity in a wireless social network is*

$$\begin{cases} \Theta(\frac{1}{\sqrt{n\log n}}) & for\ q=\Theta(n) \\ \Theta(\frac{1}{\sqrt{n\log n}}) & for\ (q,\frac{q}{n}) \overset{n\to\infty}{\to} (\infty,0) \\ \Theta(\frac{n-q+1}{n}\frac{1}{\sqrt{n\log n}}) & for\ q<\infty, 0\le\alpha<2 \\ \Theta(\frac{n-q+1}{n^2}(\sqrt{\frac{n}{\log n}})^{\alpha-1}) & for\ q<\infty, 2\le\alpha\le 3 \\ \Theta(\frac{n-q+1}{n}\frac{1}{\log n}) & for\ q<\infty, 3<\alpha \end{cases}$$

Section IV presents the proof of Theorem 1 by deriving upper and lower bounds of the throughput capacity that coincide for the various values of $q(n)$. This result shows that the scaling properties of a wireless network are a function of the spread and size of the social groups. If the size of social groups is proportional to the network size, then the order network capacity is the same as if no social groups existed, which is the same result by Gupta and Kumar [1]. Interestingly, this is the case even when the size of social groups is an insignificant fraction of the number of nodes n as n goes to infinity, and is intuitive by noticing that in such a case source-destination pairs must consume communication resources along large multi-hop paths linking sources with destinations. By contrast, when social group sizes do not grow as fast as the number of nodes in the network n, and their spread is localized ($\alpha\ge 2$), then a wireless network has order capacity increase compared to the case in which no social groups exist. This is an exciting result, because it is representative of most practical wireless networks. Section V concludes the paper by discussing the implications of our results.

To the best of our knowledge, this is the first work that considers the interaction between social groups and the underlying wireless communication infrastructure in an analytical framework of the order capacity of wireless networks.

II. Preliminaries

The network is a dense network in a unit square area with n uniformly distributed nodes. We use the protocol model [11] to determine the success of communication in the presence of multiple access interference (MAI). In particular, if χ_i, χ_j and χ_k denote the Cartesian positions in the unit square area for nodes v_i, v_j and v_k, node v_i can successfully transmit to node v_j if $|\chi_i - \chi_j| \le r(n)$, where $r(n)$ is the common transmission range of all the nodes in the network, and for any node $v_k \ne v_i$, that transmits at the same time as v_i, $|\chi_k - \chi_j| \ge (1+\Delta)r(n)$, with $\Delta > 0$ as the guard zone factor. To guarantee connectivity in this network [12], the transmission range ($r(n)$) is assumed to be $r(n) = \Theta(\sqrt{logn/n})$.

As Figure 1 illustrates, a TDMA medium access control scheme is assumed to avoid MAI. The network area is divided into square-lets with side-length $C_1 r(n), (C_1 < \frac{1}{4})$, and at any given time the cells separated by M square-lets distance are the only cells allowed to transmit as shown with a cross sign inside the cells in figure 1 where $M \ge (2+\Delta)/C_1$.

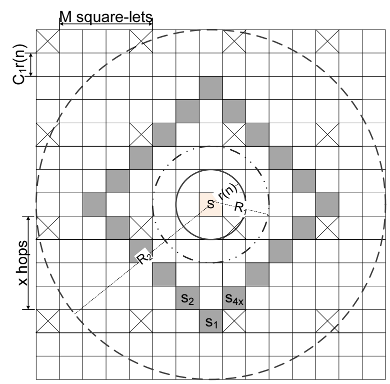

Fig. 1. The solid-line circle shows the transmission range. Dark gray cells (s_i) contain the nodes with $X = x$. R_1 (R_2) are used as the distance of each node in this region instead of their real distances to achieve upper (lower) bounds on $P(X = x)$.

The routing of information is very simple. Each node is assumed to know the locations of its intended destination and its immediate neighbors, and selects as its next hop to the destination that local contact that is closest to the destination. The local contacts are within the radio range since they are the one hop physical neighbors of the node. By assuming that there is at least one local contact in each of the four adjacent cells of the source guarantees that this simple routing protocol converges. If each node has more than four local contacts, i.e., all nodes within transmission range are local contacts, then the order throughput capacity computation does not change and the same results can be derived. The four local contacts assumption was first considered in [5] for grid networks.

We use the notation of [10] to denote the elementary symmetric polynomials of the variables $x = (x_1, ..., x_n)$ by $\sigma_{p,n}, 1 \le p \le n$. In other words,

$$\sigma_{p,n}(x) = \sigma_{p,n}(x_1, ..., x_n) = \sum_{1\le i_1<i_2<..<i_p\le n} x_{i_1}...x_{i_p}.$$

Moreover, we define the elementary symmetric polynomials of the same set of variables except one, x_k, as

$$\sigma^{\overline{k}}_{p,n-1}(x_1,...,x_n) \quad = \quad \sigma_{p,n-1}(x_1,...,x_{k-1},x_{k+1},...,x_n)$$

Lemma 1. Let $x_1,...,x_n$ be non-negative real numbers, $n \geq 2$. Then for $1 \leq p \leq n-1$, we have

$$\sigma_{1,n}\sigma_{p,n} \geq \frac{n(p+1)}{n-p}\sigma_{p+1,n}.$$

The proof of this result is by induction and described in [10]. In Section IV (Lemma 2), we prove that this is a tight bound for values of p that do not grow as fast as n.

The standard notations of O and Ω are used to describe the asymptotic upper and lower bounds respectively. When $f(n) = \Theta(g(n))$, then it is denoted by $f(n) \equiv g(n)$.

III. A Power-Law Distribution for Social Groups

In Kleinberg's model [5], every node s has a directed edge to every other node v_i within lattice distance $p \geq 1$, and directed edges to $q \geq 0$ other nodes using independent random trials. The j^{th} directed edge from s has endpoint $v_i, i = 1,..,n$ with probability proportional to $d_i^{-\alpha} \triangleq d^{-\alpha}(s,v_i)$ and normalizing factor $\sum_{i=1}^{n} d_i^{-\alpha}$. Considering the same probability distribution function for *long-range social contacts* (LSC), the probability that the LSC list contains exactly q independently selected members is

$$\begin{aligned} P(|LSC| = q) &= \sum_{1\leq i_1<...<i_q\leq n} P(LSC = \{v_{i_1},...,v_{i_q}\}) \\ &= \sum_{1\leq i_1<...<i_q\leq n} \prod_{j=1}^{q} P(v_{i_j} \in LSC) \\ &= \sum_{1\leq i_1<...<i_q\leq n} \frac{d_{i_1}^{-\alpha}...d_{i_q}^{-\alpha}}{(\sum_{j=1}^{n} d_j^{-\alpha})^q}. \end{aligned}$$

As can be seen, this probability is close to one for $q = \Theta(1)$, decreases by increasing q, and approaches zero when $q = \Theta(n)$. Kleinberg [5] mentioned that q is a universally constant value and the above derivation proves that the original power-law distribution should be modified to consider those cases when q is a function of n. We assume that each source node has the same number of LSCs $q(n)$ selected in independent random trials.

The long-range contacts are selected independently, while closer nodes to the source have a better chance of being selected as a LSC, thus, the probability that a particular q-member set is the LSC set is proportional to the product of the inverse of the distances of its members from the source. This probability can be written as

$$P(LSC = \{v_{i_1},...,v_{i_q}\}) = \frac{d_{i_1}^{-\alpha}...d_{i_q}^{-\alpha}}{N_{\alpha,q}}. \quad (1)$$

The normalization factor $N_{\alpha,q}$ is obtained using the fact that $\sum_{1\leq i_1<...<i_q\leq n} P(LSC = \{v_{i_1},...,v_{i_q}\}) = 1$.

$$N_{\alpha,q} \quad = \sum_{1\leq i_1<...<i_q\leq n} d_{i_1}^{-\alpha}...d_{i_q}^{-\alpha} \quad (2)$$

The probability that a particular node v_k is selected as a LSC (i.e., the probability that v_k is a member of the LSC set) is given by[1]

$$\begin{aligned} P &\quad (v_k \in LSC) \\ &= \sum_{1\leq i_1<...<i_{q-1}\leq n, i_j\neq k} P(LSC = \{v_k, v_{i_1},...,v_{i_{q-1}}\}), \\ &= \frac{\sum_{1\leq i_1<...<i_{q-1}\leq n, i_j\neq k} d_k^{-\alpha} d_{i_1}^{-\alpha}...d_{i_{q-1}}^{-\alpha}}{\sum_{1\leq i_1<...<i_q\leq n} d_{i_1}^{-\alpha}...d_{i_q}^{-\alpha}}. \end{aligned}$$

The above probability function denotes the probability of node v_k being in LSC, and is non-decreasing in q. It also guarantees that the described process ends up with a q-member LSC set for each source node.

Let ϑ_t be a random variable denoting the destination node. Then, for each particular $v_k \in V$ (the set of nodes except source), we have

$$\begin{aligned} P(\vartheta_t = v_k) &= P(\vartheta_t = v_k \mid v_k \in LSC) \times P(v_k \in LSC\ ) \\ &+ P(\vartheta_t = v_k \mid v_k \notin LSC) \times P(v_k \notin LSC). \end{aligned}$$

Given that the destination is only selected from LSCs, $P(v_k \notin LSC) = 0$. Furthermore, the selection of destination from LSCs has a uniform distribution.

$$\begin{aligned} P(\vartheta_t = v_k) = &\quad \tfrac{1}{q}P(v_k \in LSC) \\ &= \frac{\sum_{1\leq i_1<...<i_{q-1}\leq n, i_j\neq k} d_k^{-\alpha} \prod_{j=1}^{q-1} d_{i_j}^{-\alpha}}{q\sum_{1\leq i_1<...<i_q\leq n} \prod_{j=1}^{q} d_{i_j}^{-\alpha}} \end{aligned}$$

Let $v = (v_1,...,v_n)$ denote $(d_1^{-\alpha},...,d_n^{-\alpha})$, then the above equation can be written as

$$P(\vartheta_t = v_k) = \frac{d_k^{-\alpha}\sigma^{\overline{k}}_{q-1,n-1}(v)}{q\sigma_{q,n}(v)}. \quad (3)$$

IV. Throughput Capacity Analysis

Let λ denote the data rate for each node and X be the number of hops traveled by each bit from source to destination. The total number of concurrent transmissions in such a network is then $n\lambda E[X]$, where $E[X]$ is the average number of hops in a route for any given source-destination pair. This value is upper bounded by the total bandwidth W available, divided by the number of non-interfered groups in the TDMA scheme as shown in Figure 1 ($i.e., \frac{W}{M^2C_1^2r^2(n)}$). Therefore, using the minimum transmission range necessary to guarantee connectivity, the maximum data rate for each node is [9]

$$\lambda \leq \lambda_{max} = \Theta(\frac{1}{\log nE[X]}). \quad (4)$$

The average number of hops can be computed as

$$E[X] = \sum_{x=1}^{x_{max}} xP(X=x) = P(X=1) + \sum_{x=2}^{x_{max}} xP(X=x).$$

$P(X=1)$ is the probability that the packets travel just one hop from source to destination. This probability is a positive

[1]Again, we assume that |LSC| is equal to q for all sources.

number smaller than one, so we can ignore it when deriving the order of expected number of hops.

To compute $P(X = x)$ for $x > 1$, we need to consider the long-range contacts outside the circle with radius $r(n)$ centered at the source node. Given that all the nodes inside the transmission range of a source receive the data transmitted from it in just one hop, $P(X = x) = 0$ for $1 < x < \lceil \frac{1}{C_1} + 1 \rceil$. The information between source and destination located on two opposite corners of the network area passes through the maximum number of hops which is $\lceil \frac{2}{C_1 r(n)} \rceil$. Thus, $P(X = x)$ can be calculated as

$$E[X] \equiv \sum_{\lceil \frac{1}{C_1} + 1 \rceil}^{\lceil \frac{2}{C_1 r(n)} \rceil} xP(X = x).$$

To compute $P(X = x)$ for $x = \lceil \frac{1}{C_1} + 1 \rceil, ..., \lceil \frac{2}{C_1 r(n)} \rceil$, we need to compute the number of nodes at a distance of x hops from the source and their corresponding Euclidean distances from the source. The geometric place of such nodes is a rhombus around the source node as shown in Figure 1 and explained in [9]. The probability that the number of hops between source and destination is x hops equals the probability that the destination is located in one of the cells on the boundaries of this rhombus. Hence,

$$\begin{aligned} P \quad (X = x) &= \sum_{l=1}^{4x} P(destination\ is\ inside\ s_l) \\ &= \sum_{l=1}^{4x} \sum_{v_k\ in\ s_l} P(\vartheta_t = v_k) \end{aligned}$$

Therefore,

$$\begin{aligned} E[X] &\equiv \sum_{\lceil \frac{1}{C_1} + 1 \rceil}^{\lceil \frac{2}{C_1 r(n)} \rceil} x \sum_{l=1}^{4x} \sum_{v_k\ in\ s_l} P(\vartheta_t = v_k) \\ &\equiv \sum_{\lceil \frac{1}{C_1} + 1 \rceil}^{\lceil \frac{2}{C_1 r(n)} \rceil} x \sum_{l=1}^{4x} \sum_{v_k\ in\ s_l} \frac{d_k^{-\alpha} \sigma_{q-1,n-1}^{\overline{k}}(v)}{q\sigma_{q,n}(v)} \qquad (5) \end{aligned}$$

We now compute the average number of hops based on different values of q as a function of n.

A. Case I: q grows with n

If $q = n$, then $E[X]$ can be rewritten as

$$E[X] \equiv \sum_{x=\lceil \frac{1}{C_1} + 1 \rceil}^{\lceil \frac{2}{C_1 r(n)} \rceil} x \sum_{l=1}^{4x} \sum_{v_k\ in\ s_l} \frac{d_k^{-\alpha} \sigma_{n-1,n-1}^{\overline{k}}(v)}{n\sigma_{n,n}(v)}.$$

Since

$$\begin{aligned} d_k^{-\alpha} \sigma_{n-1,n-1}^{\overline{k}}(v) &= d_k^{-\alpha} \prod_{i=1, i \neq k}^{n} d_i^{-\alpha} \\ = \prod_{i=1}^{n} d_i^{-\alpha} &= \sigma_{n,n}(v) \end{aligned}$$

$E[X] \equiv \sum_{x=\lceil \frac{1}{C_1}+1 \rceil}^{\lceil \frac{2}{C_1 r(n)} \rceil} x \sum_{l=1}^{4x} \sum_{v_k\ in\ s_l} \frac{1}{n}$.

Because nodes are uniformly distributed over the network area, there are $nC_1^2 r^2(n)$ nodes inside each cell s_l with high probability. Thus[2]

$$\begin{aligned} E[X] &\equiv \sum_{x=\lceil \frac{1}{C_1} + 1 \rceil}^{\lceil \frac{2}{C_1 r(n)} \rceil} 4x^2 C_1^2 r^2(n) \\ &\equiv r^2(n) \int_{\lceil \frac{1}{C_1} + 1 \rceil}^{\lceil \frac{2}{C_1 r(n)} \rceil} u^2 du \equiv \frac{1}{r(n)} \equiv \sqrt{\frac{n}{\log n}}. \end{aligned}$$

Hence, the per-node throughput capacity is $\frac{1}{\sqrt{n \log n}}$, which is the same as the result by Gupta and Kumar [1]. This result is consistent, because the number of social contacts is equal to the total number of nodes in the network, and one of these nodes is selected randomly and uniformly as the destination, which is a similar assumption to that of the original work by Gupta and Kumar [1].

The second case is when $q = \Theta(n)$ but $q \neq n$. Define i.i.d. random variables $Y_i = d_i^{-\alpha}$ for $1 \leq i \leq n$ and define the sequence $Z_i = \log Y_i$ for all values of i. It is obvious that Z_i are i.i.d. as well. Utilizing the law of large numbers, we have $\lim_{m \to \infty} \frac{1}{m} \sum_{i=1}^{m} Z_i = \overline{Z}$ where $\overline{Z}$ is the expected value of random variable Z_i. Thus Eq. (3) can be computed as

$$\begin{aligned} P(\vartheta_t = v_k) &\equiv \frac{\sum_{1 \leq i_1 < .. < i_q \leq n, \exists h: i_h = k} \prod_{j=1}^{q} Y_{i_j}}{q \sum_{1 \leq i_1 < .. < i_q \leq n} \prod_{j=1}^{q} Y_{i_j}} \\ &\equiv \frac{\sum_{1 \leq i_1 < .. < i_q \leq n, \exists h: i_h = k} \exp \sum_{j=1}^{q} Z_{i_j}}{q \sum_{1 \leq i_1 < .. < i_q \leq n} \exp \sum_{j=1}^{q} Z_{i_j}} \\ &\equiv \frac{\sum_{1 \leq i_1 < .. < i_q \leq n, \exists h: i_h = k} \exp q\overline{Z}}{q \sum_{1 \leq i_1 < .. < i_q \leq n} \exp q\overline{Z}} \\ &\equiv \frac{\binom{n-1}{q-1}}{q\binom{n}{q}} = \frac{1}{n} \end{aligned}$$

Therefore, the value of $E[X]$ is similar to the case $q = n$.

$$E[X] \equiv \sum_{x=\lceil \frac{1}{C_1} + 1 \rceil}^{\lceil \frac{2}{C_1 r(n)} \rceil} x \sum_{l=1}^{4x} \sum_{v_k\ in\ s_l} \frac{1}{n} \equiv \sqrt{\frac{n}{\log n}}.$$

Using Eq. (4) provides the maximum capacity as

$$\lambda_{max} = \Theta(\frac{1}{\sqrt{n \log n}}).$$

B. Case II: n grows much faster than q

In this case, the expected number of hops between source and destination is obtained when $\lim_{n \to \infty} \frac{q}{n} = 0$, and two mutually exclusive situations must be considered, namely: $lim_{n \to \infty} q = \infty$ and $\lim_{n \to \infty} q < \infty$.

[2]Note that we are computing the order of $E[X]$ dropping constant factors.

When $\lim_{n\to\infty} q = \infty$, we can use law of large numbers and a similar procedure as before to arrive at

$$E[X] = \Theta\left(\sqrt{\frac{n}{\log n}}\right), \qquad \lambda_{max} = \Theta\left(\frac{1}{\sqrt{n \log n}}\right).$$

When each node has finite number of contacts ($\lim_{n\to\infty} q < \infty$), the numerator of $P(\vartheta_t = v_k)$ can be expanded as

$$\begin{aligned} & d_k^{-\alpha}\sigma^{\overline{k}}_{q-1,n-1}(v) \\ &= d_k^{-\alpha}(\sigma_{q-1,n}(v) - d_k^{-\alpha}\sigma^{\overline{k}}_{q-2,n-1}(v)) \\ &= d_k^{-\alpha}(\sigma_{q-1,n}(v) - d_k^{-\alpha}(\sigma_{q-2,n}(v) - d_k^{-\alpha}\sigma^{\overline{k}}_{q-3,n-1}(v))) \end{aligned}$$

Note that $d_k^{-\alpha}$ and $\sigma_{q-i,n-j}$ are positive values; therefore, the upper and lower bounds for $P(\vartheta_t = v_k)$ are obtained as

$$P_{lower} \leq P(\vartheta_t = v_k) \leq P_{upper}, \tag{6}$$

where $P_{lower} = d_k^{-\alpha}\frac{\sigma_{q-1,n}(v) - d_k^{-\alpha}\sigma_{q-2,n}(v)}{q\sigma_{q,n}(v)}$ and $P_{upper} = \frac{d_k^{-\alpha}\sigma_{q-1,n}(v)}{q\sigma_{q,n}(v)}$.

Lemma 2. Let $\Psi = \{\psi_1, ..., \psi_n\}$ be a set of $n \geq 2$ non-negative real numbers. Then for a finite p, i.e., $\lim_{n\to\infty} p < \infty$, we have

$$\frac{\sigma_{1,n}(\Psi)\sigma_{p,n}(\Psi)}{(p+1)\sigma_{p+1,n}(\Psi)} = \Theta\left(\frac{n}{n-p}\right). \tag{7}$$

Proof: Define random variables $U_i^p = \psi_{i_1}...\psi_{i_p}$ for $i = 1, .., \binom{n}{p}$ where $1 \leq i_1 < .. < i_p \leq n$. Due to symmetry, these random variables are identically distributed. Moreover, their mean $\overline{U_p}$ is a function of p. It can easily be seen that these random variables are not independent, as they may have common factors of ψ_{i_j}. We partition the set Ψ into p-member subsets. Assume that T^p is the set of all possible such partitionings (each denoted by T_i^p) with no common member, i.e., $T_i^p \cap T_j^p = \phi$. For a finite p, the number of T^p members is $|T^p| \equiv \binom{n}{p}/\left(\frac{n}{p}\right) = \binom{n-1}{p-1}$.

Now we can expand $\sigma_{p,n}(\Psi)$ to separate summations over different partitions described above. Thus,

$$\sigma_{p,n} = \sum_{1\leq i_1<..<i_p\leq n} \psi_{i_1}..\psi_{i_p} = \sum_{j=1}^{|T^p|} \sum_{\{\psi_{i_1}..\psi_{i_p}\}\in T_j^p} \psi_{i_1}..\psi_{i_p}$$

Because each inner summation is applied over one possible partitioning of Ψ, it is performed over $\frac{n}{p}$ of independent U_i as described before. The law of large numbers can be applied here.

$$\lim_{n\to\infty} \sum_{\{\psi_{i_1}..\psi_{i_p}\}\in T_j^p} \psi_{i_1}..\psi_{i_p} = \lim_{n\to\infty} \sum_{\{\psi_{i_1}..\psi_{i_p}\}\in T_j^p} U_i^p = \frac{n}{p}\overline{U_p}$$

Thus,

$$\sigma_{p,n} = \sum_{j=1}^{|T^p|} \frac{n}{p}\overline{U_p} = \binom{n}{p}\overline{U_p}.$$

A similar formulation can be derived for $\sigma_{p+1,n}(\Psi)$.

$$\sigma_{p+1,n} = \sum_{j=1}^{|T^{p+1}|} \frac{n}{p+1}\overline{U_{p+1}} = \binom{n}{p+1}\overline{U_{p+1}}$$

Therefore,

$$\frac{\sigma_{1,n}\sigma_{p,n}}{(p+1)\sigma_{p+1,n}} = \frac{\sigma_{1,n}\binom{n}{p}\overline{U_p}}{(p+1)\binom{n}{p+1}\overline{U_{p+1}}}.$$

Note that U_i^p have identical distribution and ψ_i are i.i.d.. Therefore, the expected value $\overline{U_{p+1}}$ can be expressed in terms of $\overline{U_p}$

$$\begin{aligned} \overline{U_{p+1}} &= E[U_i^{p+1}] = E[\psi_{i_1}...\psi_{i_{p+1}}] \\ &= \sum_{\psi_{i_{p+1}}} E[\psi_{i_1}...\psi_{i_p}\psi_{i_{p+1}}|\psi_{i_{p+1}}]p(\psi_{i_{p+1}}) \\ &= \sum_{\psi_{i_{p+1}}} \psi_{i_{p+1}} E[\psi_{i_1}...\psi_{i_p}]p(\psi_{i_{p+1}}) \\ &= \overline{U_p}\sum_{\psi_{i_{p+1}}} \psi_{i_{p+1}} p(\psi_{i_{p+1}}) \\ &= \overline{U_p}.\overline{\psi_{p+1}} = \overline{U_p}.\overline{\psi} \end{aligned}$$

Furthermore, by utilizing law of large numbers for $\sigma_{1,n}$ results in $\sigma_{1,n}(\Psi) \to n\overline{\psi}$. Thus

$$\frac{\sigma_{1,n}(\Psi)\sigma_{p,n}(\Psi)}{(p+1)\sigma_{p+1,n}(\Psi)} \equiv \frac{n\binom{n}{p}}{(p+1)\binom{n}{p+1}} = \frac{n}{n-p}.$$

■

Returning to the case of finite contacts, we use Lemma 2 (for $p = q-1$) and inequality (6) to obtain an upper bound for $E[X]$ in eq. (5).

$$\begin{aligned} E[X] &\leq \sum_{\lceil\frac{1}{C_1}+1\rceil}^{\lceil\frac{2}{C_1 r(n)}\rceil} x \sum_{l=1}^{4x} \sum_{v_k \ in \ s_l} \frac{d_k^{-\alpha}\sigma_{q-1,n}(v)}{q\sigma_{q,n}(v)} \\ &\equiv \frac{n}{n-q+1} \sum_{\lceil\frac{1}{C_1}+1\rceil}^{\lceil\frac{2}{C_1 r(n)}\rceil} x \sum_{l=1}^{4x} \sum_{v_k \ in \ s_l} \frac{d_k^{-\alpha}}{\sigma_{1,n}} \end{aligned} \tag{8}$$

Referring to the results presented in [9], it can be observed that the average number of hops in this case is $\frac{n}{n-q+1}$ times more than the case when there is only one long-range contact for each source. To calculate the above summation, we need to compute the distance between each node in s_i and the source. To simplify the problem, we use distances $R_1 = xC_1r(n)/A_1$ and $R_2 = A_2xC_1r(n)$ $(A_1, A_2 > 1)$ for all such nodes to reach upper and lower bounds for this summation (see figure 1).

$$\begin{aligned} \sum_{l=1}^{4x} \sum_{v_k \ in \ s_l} (A_2xC_1r(n))^{-\alpha} &\leq \sum_{l=1}^{4x} \sum_{v_k \ in \ s_l} d_k^{-\alpha} \\ &\leq \sum_{l=1}^{4x} \sum_{v_k \ in \ s_l} (xC_1r(n)/A_1)^{-\alpha} \end{aligned}$$

By replacing the number of nodes in each cell by $nC_1^2r^2(n)$ and ignoring the constant values in the above inequality, we can see that the order of both upper and lower bounds are the same.

$$\begin{aligned}\sum_{\lceil \frac{1}{C_1}+1\rceil}^{\lceil \frac{2}{C_1 r(n)}\rceil} x \sum_{l=1}^{4x} \sum_{v_k \ in \ s_l} d_k^{-\alpha} &\equiv nr^{2-\alpha}(n) \sum_{\lceil \frac{1}{C_1}+1\rceil}^{\lceil \frac{2}{C_1 r(n)}\rceil} x^{2-\alpha} \\ &\stackrel{a}{\equiv} nr^{2-\alpha}(n) \int_{\lceil \frac{1}{C_1}+1\rceil}^{\lceil \frac{2}{C_1 r(n)}\rceil+1} u^{2-\alpha} du\end{aligned}$$

The last equality (a) is obtained by replacing the sum by its integral approximation. After computing that integral for a sufficiently large value of n we arrive at

$$\sum_{\lceil \frac{1}{C_1}+1\rceil}^{\lceil \frac{2}{C_1 r(n)}\rceil} x \sum_{l=1}^{4x} \sum_{v_k \ in \ s_l} d_k^{-\alpha} \equiv \begin{cases} \Theta(n\sqrt{\frac{n}{\log n}}) & , for \ 0 \leq \alpha \leq 3 \\ \Theta(n(\sqrt{\frac{n}{\log n}})^{\alpha-2}) & , for \ 3 \leq \alpha \end{cases} \quad (9)$$

Moreover, $\sigma_{1,n}$ can be written as

$$\sigma_{1,n} = \sum_{v_k} d_k^{-\alpha} \equiv \int_{r(n)}^{\gamma d_{max}} nu^{1-\alpha} du,$$

where d_{max} is the maximum distance between any two nodes in the network, and $\gamma \leq 1$. Calculating the integral for a sufficiently large value of n leads to

$$\sigma_{1,n} \equiv \begin{cases} \Theta(n) & for \ 0 \leq \alpha \leq 2 \\ \Theta(n(\sqrt{\frac{n}{\log n}})^{\alpha-2}) & for \ 2 \leq \alpha \end{cases} \quad (10)$$

The derivations of Eqs. (9) and (10) are described in the Appendix.

Now we can use these results in Eq. (8) to obtain the following upper bound for $E[X]$. Note that $E[X] \geq 1$; therefore, if the computation ends up with $E[X] < 1$, we replace it with 1.

$$E[X] = \begin{cases} O(\frac{n}{n-q+1}\sqrt{\frac{n}{\log n}}) & for \ 0 \leq \alpha < 2 \\ O(\frac{n}{n-q+1}(\sqrt{\frac{n}{\log n}})^{3-\alpha}) & for \ 2 \leq \alpha \leq 3 \\ O(\frac{n}{n-q+1}) & for \ 3 < \alpha \end{cases}$$

The lower bound capacity follows immediately.

$$\lambda_{max} = \begin{cases} \Omega(\frac{n-q+1}{n}\frac{1}{\sqrt{n\log n}}) & for \ 0 \leq \alpha < 2 \\ \Omega(\frac{n-q+1}{n^2}(\sqrt{\frac{n}{\log n}})^{\alpha-1}) & for \ 2 \leq \alpha \leq 3 \\ \Omega(\frac{n-q+1}{n}\frac{1}{\log n}) & for \ 3 < \alpha \end{cases}$$

Thus, these are the upper bounds of $E[X]$ and the lower bounds on the capacity if the number of long-range contacts is a finite number greater than one.

To compute the lower bound for $E[X]$, we will study the lower bound of $P(\vartheta_t = v_k)$ in Eq. (6). First, we calculate the order of $\frac{\sigma_{q-2,n}(v)}{q\sigma_{q,n}(v)}$. This value is obtained by replacing $p = q-1$ and $p = q-2$ in Eq. (7).

$$\frac{\sigma_{1,n}\sigma_{q-1,n}}{q\sigma_{q,n}} = \Theta\left(\frac{n}{n-q+1}\right)$$
$$\frac{\sigma_{1,n}\sigma_{q-2,n}}{(q-1)\sigma_{q-1,n}} = \Theta\left(\frac{n}{n-q+2}\right)$$

By multiplying these two equations and combining with Eq. (10), we arrive at

$$\begin{aligned}\frac{\sigma_{q-2,n}}{q\sigma_{q,n}} &= \Theta(\frac{(q-1)n^2}{(n-q+1)(n-q+2)\sigma_{1,n}^2}) \\ &= \begin{cases} \Theta(\frac{(q-1)}{(n-q+1)(n-q+2)}) & for \ 0 \leq \alpha < 2 \\ \Theta(\frac{(q-1)(\log n)^{\alpha-2}}{(n-q+1)(n-q+2)n^{\alpha-2}}) & for \ 2 < \alpha \end{cases}\end{aligned} \quad (11)$$

The lower bound for $E[X]$ is derived by combining Eqs. (5) and (6).

$$\begin{aligned}E[X] &\geq \sum_{\lceil \frac{1}{C_1}+1\rceil}^{\lceil \frac{2}{C_1 r(n)}\rceil} x \sum_{l=1}^{4x} \sum_{v_k \ in \ s_l} \frac{d_k^{-\alpha}\sigma_{q-1,n}(v) - d_k^{-2\alpha}\sigma_{q-2,n}(v)}{q\sigma_{q,n}(v)}, \\ &= \frac{\sigma_{q-1,n}(v)}{q\sigma_{q,n}(v)} \sum_{\lceil \frac{1}{C_1}+1\rceil}^{\lceil \frac{2}{C_1 r(n)}\rceil} x \sum_{l=1}^{4x} \sum_{v_k \ in \ s_l} d_k^{-\alpha}, \\ &- \frac{\sigma_{q-2,n}(v)}{q\sigma_{q,n}(v)} \sum_{\lceil \frac{1}{C_1}+1\rceil}^{\lceil \frac{2}{C_1 r(n)}\rceil} x \sum_{l=1}^{4x} \sum_{v_k \ in \ s_l} d_k^{-2\alpha}.\end{aligned}$$

If we replace the terms in the negative part of the above formula with their equivalents from Eqs. (9) and (11), it appears that this part will be of an order less than one. Thus, it can be ignored when comparing the positive part of this formula and the lower bound for $E[X]$ are the same as its upper bound. Therefore, the obtained lower bounds on capacity are indeed tight bounds.

However, it is important to compute the traffic carried in each cell and find out if this throughput capacity can be supported for each cell. It can be proved that the total traffic to each cell is upper bounded by $\log n$. Therefore, the maximum throughput capacity is upper bounded by the inverse of this traffic [13], i.e., $\lambda_{max} \leq \frac{1}{\log n}$, which does not violate the throughput capacity bounds we derived earlier.

V. Discussion and Future Work

This paper presents the first modeling framework for the capacity of a wireless network in which nodes communicate in the context of social groups and successful transmissions can occur only between nodes within transmission range of each other. The model characterizes a wireless network of n nodes with each social group has a size that is a function of the number of nodes n, the probability of a node being a long-range social contact of a source is inversely proportional to their Euclidean distance with power factor α, and MAI is modeled according to the protocol model.

Figure 2 illustrates the results of Theorem 1 by plotting the network capacity as a function of n for different values of α when the number of long-range contacts is a fixed number, i.e., $q(n) = 5$. The capacity order decreases exponentially as the number of nodes increases. However, increasing the value of α affects the rate of this capacity decrease. Small values of α correspond to the case in which the social groups are highly distributed in the wireless network, and lead to a rate of order-capacity decrease similar to the results derived by Gupta and Kumar [1], in which no social groups exist. In contrast, for large values of α, social groups are localized, the paths from sources to destinations involve only $\Theta(1)$ hops, and the maximum throughput capacity is achieved. Furthermore, rate of order-capacity decrease is much smaller than with small values of α.

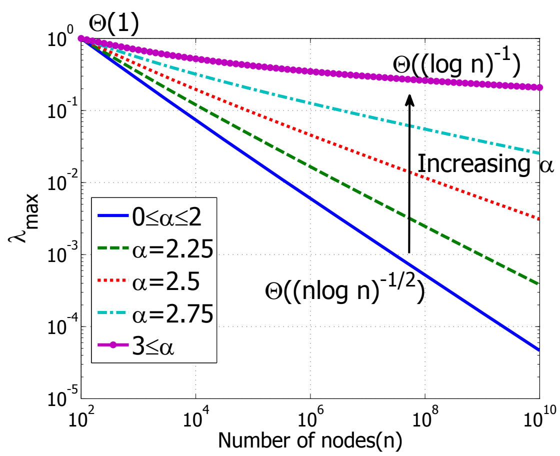

Fig. 2. Throughput capacity vs. the number of nodes when $q = 5$.

Figure 3 shows the throughput capacity versus the power law exponent (α) for two types of values of $q(n)$. In one case, $q(n)$ is a function of n, i.e., $q(n) = f(n)$, where $f(n)$ is an increasing function of n, and in the second case $q(n)$ is a constant value, i.e., $q(n) = 100$. As the figure illustrates, if the number of long-range contacts is not a function of the number of nodes, the resulting capacity changes with the parameter α. If α assumes small values ($\alpha \leq 2$), the network behaves as if there were no social groups. For medium values of α ($2 < \alpha < 3$), an exponential growth is observed in the throughput capacity from $\Theta(1/\sqrt{n \log n})$ to $\Theta(1/\log n)$. For large values of α ($\alpha \geq 3$), each source selects its destination along a path involving only $\Theta(1)$ hops w.h.p. and the resulting capacity is the maximum capacity that can be obtained. We also observe that the rate of capacity increase is very slow for $\alpha > 4$.

However, if the number of long-range social contacts $q(n)$ grows proportional to the number of nodes n, the network behaves as if the network had no social groups, independently of the rate of growth for $q(n)$, and each node selects its destination randomly from all the other network nodes. In this case, the throughput capacity does not change with parameter α, and this is true even if $q(n)$ is much smaller than n, i.e., $q(n) = \log\log(n)$ which is a small number even when n is a very large number.

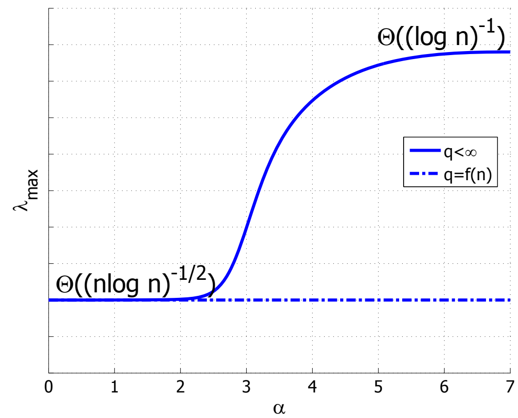

Fig. 3. Throughput capacity changes with α.

While the above results may appear somewhat in the abstract, they are actually great news for real wireless networks, because the size of social groups of individuals are finite (and could be argued relatively small) and become independent of the total number of nodes as the latter grows to infinity.

In this work we have made many assumptions to simplify our analytical framework. For example, we have assumed that all the nodes have social groups with the same size and dispersion, that each source unicasts with a single destination in its social group, that the protocol model is used to model MAI, and that all radios are similar. In addition, we have not addressed the role of content popularity or common interest in content within social groups. We hope to relax these assumptions in our future work, and that this paper will inspire more modeling work on the impact of social groups in wireless networks.

Appendix

Detailed derivation of equation (9)

$$\sum_{\lceil \frac{1}{C_1}+1 \rceil}^{\lceil \frac{2}{C_1 r(n)} \rceil} x \sum_{l=1}^{4x} \sum_{v_k \ in \ s_l} d_k^{-\alpha}$$

$$\equiv \quad n r^{2-\alpha}(n) \int_{\lceil \frac{1}{C_1}+1 \rceil}^{\lceil \frac{2}{C_1 r(n)} \rceil +1} u^{2-\alpha} du$$

$$= \quad \frac{n r^{2-\alpha}(n)}{3-\alpha}((\lceil \frac{2}{C_1 r(n)} \rceil + 1)^{3-\alpha} - (\lceil \frac{1}{C_1} + 1 \rceil)^{3-\alpha})$$

If the transmission range is the minimum range required for the network connectivity, i.e. $r(n) = \Theta(\sqrt{\frac{\log n}{n}})$, then for sufficiently large n, we have

$$\left(\lceil\frac{2}{C_1r(n)}\rceil+1\right)^{3-\alpha}=\Theta((\sqrt{\frac{n}{\log n}})^{3-\alpha})$$

If $\alpha < 3$,[3]

$$\begin{aligned}&(\lceil\frac{2}{C_1r(n)}\rceil+1)^{3-\alpha}-(\lceil\frac{1}{C_1}+1\rceil)^{3-\alpha}\\ \equiv\;&\Theta((\sqrt{\frac{n}{\log n}})^{3-\alpha})-\Theta(1)\\ =\;&\Theta\left((\sqrt{\frac{n}{\log n}})^{3-\alpha}\right)\end{aligned}$$

Therefore,

$$\begin{aligned}&\sum_{\lceil\frac{1}{C_1}+1\rceil}^{\lceil\frac{2}{C_1r(n)}\rceil}x\sum_{l=1}^{4x}\sum_{v_k\ in\ s_l}d_k^{-\alpha}\\ \equiv\;&\frac{nr^{2-\alpha}(n)}{3-\alpha}\Theta((\sqrt{\frac{n}{\log n}})^{3-\alpha})\equiv\Theta(n\sqrt{\frac{n}{\log n}})\end{aligned}$$

For dense social networks in which $\alpha > 3$ we have

$$\begin{aligned}&\sum_{\lceil\frac{1}{C_1}+1\rceil}^{\lceil\frac{2}{C_1r(n)}\rceil}x\sum_{l=1}^{4x}\sum_{v_k\ in\ s_l}d_k^{-\alpha}\\ \equiv\;&\frac{nr^{2-\alpha}(n)}{\alpha-3}((\frac{1}{\lceil\frac{1}{C_1}+1\rceil})^{\alpha-3}-(\frac{1}{\lceil\frac{2}{C_1r(n)}\rceil+1})^{\alpha-3})\end{aligned}$$

and for large n

$$\begin{aligned}&(\frac{1}{\lceil\frac{1}{C_1}+1\rceil})^{\alpha-3}-(\frac{1}{\lceil\frac{2}{C_1r(n)}\rceil+1})^{\alpha-3}\\ \equiv\;&\Theta(1)-\Theta((\sqrt{\frac{\log n}{n}})^{\alpha-3})\equiv\Theta(1)\end{aligned}$$

Thus, the above summation is equivalent to

$$\frac{nr^{2-\alpha}(n)}{\alpha-3}\Theta(1)\equiv\Theta\left(n(\sqrt{\frac{n}{\log n}})^{\alpha-2}\right)$$

Detailed derivation of equation (10)

For large n with minimum transmission range:

$$\begin{aligned}\sigma_{1,n}\;&\equiv\;\int_{r(n)}^{\gamma d_{max}}nu^{1-\alpha}du\\ &=\;\frac{n}{2-\alpha}((\gamma d_{max})^{2-\alpha}-r^{2-\alpha}(n))\end{aligned}$$

If $\alpha < 2$, for large n, the transmission range is very small; therefore,

$$\sigma_{1,n}\equiv\frac{n}{2-\alpha}(\gamma d_{max})^{2-\alpha}\equiv\Theta(n)$$

[3]Note that for $\alpha = 3$, both Cases I and II give the same result.

And for larger α, $\sigma_{1,n}$ is

$$\begin{aligned}\sigma_{1,n}\;&\equiv\;\frac{n}{\alpha-2}((\sqrt{\frac{n}{\log n}})^{\alpha-2}-(\frac{1}{\gamma d_{max}})^{\alpha-2})\\ &\equiv\;\frac{n}{\alpha-2}(\sqrt{\frac{n}{\log n}})^{\alpha-2}\equiv\Theta\left(n(\sqrt{\frac{n}{\log n}})^{\alpha-2}\right)\end{aligned}$$

Acknowledgment

This research was partially sponsored by the U.S. Army Research Laboratory under the Network Science Collaborative Technology Alliance, Agreement Number W911NF-09-0053, by the Army Research Office under agreement number W911NF-05-1-0246, by the National Science Foundation under grant CCF-0729230, and by the Baskin Chair of Computer Engineering. The views and conclusions contained in this document are those of the author(s) and should not be interpreted as representing the official policies, either expressed or implied, of the U.S. Army Research Laboratory or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation hereon.

References

[1] P. Gupta, P. R. Kumar, "The capacity of wireless networks," *IEEE Transaction on Information Theory*, Vol. 46, No. 2, pp. 388-404, 2000.

[2] L. Backstrom, E. Sun,C. Marlow, "Find Me If You Can: Improving Geographical Prediction with Social and Spatial Proximity," *Proc. WWW'10*, Raleigh, NC, 2010.

[3] S. Milgram, "The small world problem," *Psychology Today*, No. 1, Vol. 61, 1967.

[4] D. Watts, S. Strogatz, "Collective dynamics of small-world networks," *Nature*, pp. 393-440, 1998.

[5] J. Kleinberg, "The small-world phenomenon: an algorithm perspective," *Proc. 32nd Annual ACM Symposium on Theory of Computing*, May 21-23, 2000, Portland, Oregon, USA.

[6] P. Fraigniaud, G. Giakkoupis, "On The Searchability of Small-World Networks with Arbitrary Underlying Structure," *Proc. ACM STOC'10*, June 2010, Cambridge, Massachusetts, USA

[7] M. Dietzfelbinger, P. Woelfel, "Tight Lower Bounds for Greedy Routing in Uniform Small World Rings," *Proc. ACM STOC 09*, pp. 591-600, 2009.

[8] J. Li, Ch. Blake, D. S. J. De Couto, H. I. Lee, R. Morris, "Capacity of ad hoc wireless networks," *Proc. 7th Annual Int'l Conf. on Mobile Computing and Networking*, July 2001, Rome, Italy.

[9] B. Azimdoost, H. R. Sadjadpour, J.J. Garcia-Luna Aceves, "Capacity of Composite Networks: Combining Social and Wireless Ad Hoc Networks," *Proc. IEEE WCNC 2011*, Cancun, Mexico, March 28 - 31, 2011.

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

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

[12] S. Kulkarn, P. Viswanath, "A Deterministic Approach to Throughput Scaling in Wireless Networks," *IEEE Trans. Information Theory*, Vol. 50, pp. 1041-1049, 2004.

[13] B. Liu, D. Towsley, A. Swami, "Data Gathering Capacity of Large Scale Multihop Wireless Networks," *IEEE International Conference on Mobile Ad-hoc and Sensor Systems (MASS)*, 2008.