Scaling Properties of Random Networks Under Proximity-based Social Relations

Ali Dabirmoghaddam* J. J. Garcia-Luna-Aceves†*
*University of California, Santa Cruz, CA 95064
†PARC, Palo Alto, CA 94304
{alid, jj}@soe.ucsc.edu

***Abstract*—We present an analytical framework to investigate the interplay between a communication graph and an overlay of social relationships. Particularly, we focus on geographical distance as the key element that interrelates the concept of routing in a communication network with interaction patterns on the social graph. Through this regime, we attempt to identify classes of social relationships that let the ensuing system scale, *i.e.*, accommodate a large number of users given only finite amount of resources. We establish that stochastically localized communication patterns are indispensable to network scalability.**

***Keywords*—random geometric graph, social network, scalability**

I. Introduction

Computer networks can be conceptually organized into several distinct layers that, though logically separate, are operationally interconnected. Within this framework, such constructs are often referred to as *composite (or complex) networks* in which a *communication network* represents the physical communication infrastructure, computing servers, and clients, while a *social network* defines the communication patterns among end users collaborating with one another through applications running in end systems (host computers). An *information network* captures the distribution and relationships among information objects throughout the network.

Due to the involved intricacies, previous work has primarily studied the performance of networks from unidimensional viewpoints of communication, social, or information, while the reciprocal interactions among these layers have largely remained overlooked. Examples of studies of communication networks neglecting the latent social relationships are [1]–[4]. In contrast, several interaction patterns and social paradigms [5]–[7] are independently studied while the restrictions imposed by realistic underlying communication networks are neglected.

The interactions among the communication, social, and information components of composite networks have an undeniable impact on their behavior. Neglecting such an important aspect results in overly simplified models with implications that, though strong in context, are limited in scope and may hardly be extended to more sophisticated real-world scenarios.

We present an analytical framework to help improve the understanding of the relationships between the communication and social networks. We study how the spatial diversity of social connections can affect the scalability of a wireless network. To that end, we focus on the generic family of *proximity-based social models* according to which social relations are established with respect to the geographical vicinity of nodes. Based on this model, nodes are socially inclined to communicate with parties that are geographically closer to them more often than with the ones at farther distances. The relevance of this model to real-world social behaviors of people has been widely studied and verified in both online and offline domains [8]–[10].

The ultimate objective of this analysis is to identify classes of social relationships that allow the underlying communication network to scale. For this purpose, we classify social models in terms of their clustering degrees (the parameter α in our model) such that nodes show higher tendency to communicate within their proximal neighborhood with a larger clustering degree. According to our findings, stochastically localized social interactions (*i.e.*, $\alpha > 3$) are indispensable to scalability of both extended and dense networks.

Section II provides a formal description of our model and the key assumptions. Sections III and IV discuss a framework for characterizing routing dynamics in random networks. Section V uses the resulting model to examine scalability conditions. Finally, Section VI concludes the paper and discusses some avenues for future research.

II. Problem Definition and Assumptions

The term *scalable* usually refers to systems capable of handling a large number of users without incurring significant loss in performance. Here, we present a more objective definition of scalability. We introduce a cost measure that reflects the average amount of resources needed to accommodate a user. In the context of communication networks, a reasonable cost measure is the average number of times a packet needs to be transmitted until it is delivered to its intended destination.

We identify three key types of factors that can influence this measure. *Topological factors* specify the physical connectivity among nodes, *e.g.*, the number of hops separating a source-destination pair on the communication graph. *Social factors* determine the patterns according to which nodes interact with one another, *e.g.*, how a source node chooses its destinations. *Unrestrained factors*, which are related to such physical-layer effects as interference, fading, noise, congestion, and others, which may result in packet losses and incur re-transmissions. Among these factors, the first two can be modeled under a minimal and general set of assumptions, which we discuss below. We do not address unrestrained factors in this paper.

A. The Connectivity Graph Model

We assume a Random Geometric Graph (RGG) as the model for the network topology. Thanks to their simplicity and generality, RGG's have become a de facto standard in the research community to represent the underlying topology of wireless networks. A definition of RGG is provided in the following for future reference.

Definition 1. *$\mathcal{G}(\mathcal{X}; r)$ represents a random geometric graph in which $\mathcal{X}$ is a point process on $\mathbb{R}^k$ that describes the distribution of nodes. Further, an undirected edge connects every pair u and v iff $\|\mathcal{X}_u - \mathcal{X}_v\| \leq r$ for a given $r \in \mathbb{R}^+$.*

Here, $\|.\|$ is a norm of choice on $\mathbb{R}^k$. For simplicity, we choose to use the Euclidean norm in this paper. We consider a Poisson point process (P.P.P.), $\mathcal{X}$, to describe the nodes' geographical distribution in the network. Further, the physical connectivity between nodes is defined according to a Boolean model that assumes nodes as being connected if and only if they are within a distance r from one another.

Two distinct models are commonly used when studying asymptotic behaviors of RGG's.

1) Extended model, in which the node density is fixed, and the network dimensions go to infinity.
2) Dense model, in which the network dimensions are fixed, and the node density goes to infinity.

We construct a general framework that can be used to analyze the scaling properties of these two network models.

B. The Social Communications Model

The social model describes the quality and frequency of inter-node communications in the network, *i.e.*, how sources choose their destinations. In this paper, we consider a proximity-based social model which is defined as follows.

Definition 2. *A communication network follows a proximity-based social model if the probability of every node u and v communicating with each other is inversely proportional to $\|\mathcal{X}_u - \mathcal{X}_v\|^\alpha$ for some arbitrary but fixed exponent $\alpha \in \mathbb{R}_0^+$.*

Definition 2 implies a social model that is power-law distributed with distance. Several studies [8]–[10] have recently verified the relevance of this model to the actual patterns of social interactions in real networks. According to this definition, for a specific realization of the network, the probability of node u choosing v as destination, $P_u(v)$, is obtained as follows:

$$P_u(v) = \frac{d(u,v)^{-\alpha}}{\sum_{w \neq u} d(u,w)^{-\alpha}}\,, \tag{1}$$

where $d(u,v) = \|\mathcal{X}_u - \mathcal{X}_v\|$. The denominator of (1) is in fact a normalizing constant.

According to Equation (1), the geographically closer two nodes are, the more likely they are to communicate. The exception is the case of $\alpha = 0$, which results in a uniform communication model in which a source node is equally likely to choose any other node as its destination, irrespective of their distance. At the other extreme, when $\alpha \to \infty$, every node communicates with its closest neighbor almost surely. In fact, different ranges of α correspond to distinct classes of social relationships with identical scaling behaviors. Identifying such social classes is a primary objective of this paper.

We want to obtain the probability distribution of the event of having a social contact at any given physical distance by considering a RGG on 2-D space. For generality purposes, we choose to calibrate the distance measure by scaling it by the nodes' critical transmission range r. This allows our social model to be equally applicable to both cases of dense and extended topologies. Particularly, for the case of a dense network in which the diameter of the network is fixed and the critical transmission range approaches zero, this adjustment allows having social contacts that are spaced infinitely far away.

Let $x_\circ \leq x \leq d/r$ be such a range-adjusted distance measure, where d is the diameter of the network; that is, the longest possible physical distance between any pair of nodes, and $x_\circ$ denotes the minimum range-adjusted distance between two social contacts. Without loss of generality, we assume that nodes' social contacts are at least one transmission distance away from them, which means $x_\circ = 1$.

Define $F_\alpha(x) = \Pr\{\text{having a social contact at distance } \leq rx\}$. According to Definition 2, assume that such density function is a power-law on distance. Also, by Poisson approximation, we know that the number of potential social contacts at any distance x in 2-D space is linearly proportional to x. Therefore, we define the corresponding p.d.f. as follows:

$$f_\alpha(x) = crx \cdot (rx)^{-\alpha} = c\,(rx)^{1-\alpha}\,,$$

in which c is a constant independent of x. To obtain the value of c note that

$$1 = \int_1^{d/r} f_\alpha(x)\,\mathrm{d}x = \frac{c\,r^{1-\alpha}}{2-\alpha} \times x^{2-\alpha}\Big|_1^{d/r}\,. \tag{2}$$

Given that we are investigating the scaling behavior of a network when $d/r \to \infty$, a natural requirement is to have $\alpha > 2$, as otherwise, the right-hand side of (2) would diverge. Thus, from Equation (2), for any $\alpha > 2$, we obtain that

$$c = r\,(2-\alpha)\left(d^{2-\alpha} - r^{2-\alpha}\right)^{-1}.$$

The p.d.f. $f_\alpha(x)$ provides a description of our proximity-based social model. We use this model to define our cost measure as discussed in the following subsection.

C. Expected Social Path Length (ESPL)

Recall that the cost that every packet imposes to the network is measured by the expected number of times it has to be transmitted in the network until delivery. Knowing the average number of hops each packet travels considering the underlying social relations, we define the expected social path length (ESPL) as follows.

Definition 3. *The* ESPL *is the expected number of hops, $\bar{h}(x)$, separating a source-destination pair on a proximity-based social network identified by $f_\alpha(x)$ and is calculated as*

$$\mathbb{E}[\mathscr{L}_\alpha] = \int_1^{d/r} f_\alpha(x)\,\bar{h}(x)\,\mathrm{d}x\,. \tag{3}$$

Definition 3 exploits the notion of geographical distance to combine the routing on the connectivity graph of the network with the concept of social relations. In view of that, ESPL is a cost measure reflecting the amount of resources that every node consumes on average, while accounting for both topological and social considerations.

Evidently, ESPL is a non-decreasing function of the network size; nevertheless, the network cannot obviously sustain a continuously increasing load forever as more nodes join in. Hence, we present the following definition for the class of social relations that allow the underlying communication network scale appropriately without significant loss in performance.

Definition 4. *A communication network with proximity-based social relations exhibits scalability if* $\mathbb{E}[\mathscr{L}_\alpha] < \infty$ *when the number of nodes* $n \to \infty$.

Based on Definition 4, a necessary condition for scalability is that the network performs, on average, a finite number of transmissions per packet, no matter how large the network would grow. In the sequel, we investigate to see how different values of α affect the growth of ESPL as the network grows larger. To that end, we first introduce a methodology to compute the average number of hops, $\bar{h}(x)$, that a routing algorithm takes over any given distance x. This analysis primarily forms the content of the next section.

III. Progressive Walks on Random Graphs

As seen from Definition 3, an accurate evaluation of ESPL depends at least in part on the performance of the employed routing algorithm. Conventionally, it is preferred to characterize the behavior of the system under idealistic conditions to obtain a reasonable upper-bound on the achievable performance limits. As such, in most practical settings, the underlying routing algorithm is usually assumed to be optimal, *i.e.*, the minimum weight routing (a.k.a. shortest path routing).

Although finding optimal paths on deterministic graphs is algorithmically straightforward, in the context of random graphs, it turns out to be a highly challenging problem. Most of this complexity stems from the random nature of the underlying topology. In essence, an optimal routing algorithm requires global and exact information about the network structure which is virtually non-existent when speaking of RGG's.

Despite theoretical difficulties in analysis of optimal routing in random configurations, more tractable solutions with near-optimal performance can still be conceived. One such routing strategy is known as *greedy forwarding* in which every relay attempts to push the packet some distance closer to the destination (see for example [11]). With this policy, even though the global structure of the routes will/should not be necessarily optimized, a sub-optimal path can still be found through making locally optimized decisions when choosing subsequent relays along the path.

Various criteria for optimizing local decisions have been studied in the existing literature, and this is, essentially, what makes different variations of greedy forwarding. We abstract away such functional details by introducing the notion of *progressive walk* that captures the gist of greedy forwarding.

Definition 5. *We say a walk* $\langle s, \ldots, t\rangle$ *on* $\mathcal{G}(\mathcal{X}; r)$ *is a progressive walk from* s *to* t *and denote it with* $s \rightsquigarrow t$ *iff* $\|\mathcal{X}_u - \mathcal{X}_t\| \geq \|\mathcal{X}_v - \mathcal{X}_t\|$ *for all ordered pairs* (u, v) *on* $s \rightsquigarrow t$.

In fact, for a given pair of source-destination, a greedy forwarding algorithm outputs a progressive walk on the communication graph providing the existence of such a walk, and that the algorithm can find it. The expected number of hops on a greedy route is equivalent to the expected length of the corresponding walk.

Note that the former requirement on the existence of such a walk is trivial when the expected length of the walk is a parameter of interest. The latter condition, on the other hand, implicitly assumes that the routing algorithm is always able to find a progressive walk with high probability (w.h.p.). Of course, in practice, this might not necessarily be true. To elucidate, let us first have a closer look at the core mechanism of the greedy routing algorithm, *i.e.*, progressive forwarding, through the following definition. Here, $\mathcal{B}(\mathcal{X}_c; r)$ denotes the ball of radius r centered at $\mathcal{X}_c$.

Definition 6. *We define the hand-off region of a relay* u *for a final destination* t *as* $\mathcal{H}_t(u) \triangleq \mathcal{B}(\mathcal{X}_u; r) \cap \mathcal{B}(\mathcal{X}_t; x)$, *where* $x = \|\mathcal{X}_u - \mathcal{X}_t\|$. *Further, we say node* v *is a potential next-hop for* $u \rightsquigarrow t$ *iff* $\mathcal{X}_v \in \mathcal{H}_t(u)$.

In accordance with the requirements of Definition 5, the hand-off region defines the subset of nodes that can be considered by a relay as potential next-hops to further the packet towards its destination. The convergence of the progressive walk relies upon having at least one potential next-hop in each and every hand-off region along the walk. If the packet comes at a relay with a void hand-off region, *i.e.*, a dead-end, the progressive walk stalls as no further progress is allowed. For the time being, we continue with the assumption that the greedy algorithm converges w.h.p. However, we shall later relax this assumption by slightly modifying our routing algorithm to circumvent dead-ends should one be encountered.

A. Greedy Forwarding with Almost Sure Convergence

As implied by Definition 5, the key element of a progressive walk is to progressively reduce the remaining distance to the destination along the walk. In fact, at every stage of the walk, the packet is pushed some distance closer to the destination on the Euclidean plane. In view of this, a progressive walk can be perceived as a drifted random walk on the communication graph. The distance traveled by the packet at every hop is a random variable determined by the process specifying the topology of the communication graph as well as the optimization criteria of the greedy routing algorithm. Exploiting results from the theory of martingales, Theorem 1 provides a useful model that describes the relationship between the physical distance and the average hop-count on a RGG, under certain conditions when a greedy forwarding algorithm is considered.

Theorem 1. *Consider a source* s *and a destination* t *spaced at distance* $x = \|\mathcal{X}_s - \mathcal{X}_t\| > r$ *in a* RGG $\mathcal{G}(\mathcal{X}; r)$. *Node* s *sends a packet to* t *through multiple intermediate hops employing a geographical greedy forwarding algorithm. Let* ξ_δ *be a*

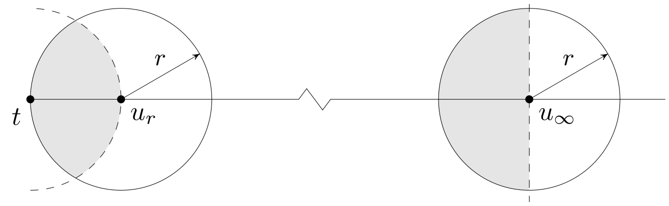

Fig. 1: A 2-D illustration of how the hand-off region, *i.e.*, shaded area, shrinks as the remaining distance to the destination is reduced. Here t is the destination, and u_r and u_∞ represent relays at distances r and ∞ from destination, respectively.

random variable denoting the progress towards destination if a transmission at distance δ from destination takes place. Provided that ξ_δ's are independent, and the routing algorithm converges w.h.p.,

$$\lim_{\delta \to r^+} \mathbb{E}[\xi_\delta] < \frac{x}{\bar{h}(x)} < \lim_{\delta \to \infty} \mathbb{E}[\xi_\delta]\,. \tag{4}$$

Proof: Let $S_\delta(t) = \sum_{i=1}^{t} \xi_\delta(i)$ be a random walk where $\xi_\delta(i)$ is a stochastic process with respect to i representing the progress towards destination when at distance δ from it. In fact, S_δ is a progressive walk that assumes all relays have similar-sized hand-off regions as if they are all at distance δ from destination.

Let $T_\delta = \inf\{t : S_\delta(t) \geq x\}$ be the first time $S_\delta(t)$ hits the target distance x. Note that $0 \leq \xi_\delta(i) \leq r$ and $\mathbb{E}[\xi_\delta] > 0$; thus, $P(T_\delta < \infty) = 1$. Also, $\{t < T_\delta\} = \{S_\delta(1), \ldots, S_\delta(t) < x\}$ which is clearly independent of $S_\delta(t')$ for $t' > T_\delta$. Therefore, T_δ is a stopping time with respect to $S_\delta(t)$.

Fix a δ such that $r < \delta < x$, and consider a relay at distance δ from destination. The measure of hand-off region is a monotonically decreasing function of δ (see Fig. 1); therefore,

$$\lim_{\delta \to r^+} \mathbb{E}[\xi_\delta] < \mathbb{E}[\xi_\delta] < \lim_{\delta \to \infty} \mathbb{E}[\xi_\delta] \quad \text{for all } \delta > r\,. \tag{5}$$

Now, consider the process $M_\delta(t) = S_\delta(t) - t\,\mathbb{E}[\xi_\delta]$. Note that,

$$\begin{aligned}
\mathbb{E}[M_\delta(t)] &= \mathbb{E}\Big[S_\delta(t) - t\,\mathbb{E}[\xi_\delta]\Big] = \mathbb{E}\Big[\sum_{i=1}^{t} \xi_\delta(i) - t\,\mathbb{E}[\xi_\delta]\Big] \\
&= \mathbb{E}\Bigg[\sum_{i=1}^{t} \Big(\xi_\delta(i) - \mathbb{E}[\xi_\delta]\Big)\Bigg] = \sum_{i=1}^{t} \mathbb{E}\Big[\xi_\delta(i) - \mathbb{E}[\xi_\delta]\Big] \\
&= \sum_{i=1}^{t} \Big(\mathbb{E}[\xi_\delta] - \mathbb{E}[\xi_\delta]\Big) = 0 < \infty\,.
\end{aligned}$$

Also, $\mathbb{E}[M_\delta(t+1) - M_\delta(t)] = \mathbb{E}[M_\delta(t+1)] - \mathbb{E}[M_\delta(t)] = 0$. Therefore, $M_\delta(t)$ is a martingale with respect to ξ_δ. According to the optional stopping theorem, $M_\delta(T_\delta \wedge t)$ is also a martingale with respect to ξ_δ, where $(T_\delta \wedge t)$ is the minimum of T_δ and t. Hence,

$$\begin{aligned}
\mathbb{E}[M_\delta(T_\delta)] &= \mathbb{E}\Big[S_\delta(T_\delta) - T_\delta\,\mathbb{E}[\xi_\delta]\Big] \\
&= \mathbb{E}[S_\delta(T_\delta)] - \mathbb{E}[T_\delta] \cdot \mathbb{E}[\xi_\delta] = 0\,,
\end{aligned}$$

which yields

$$\mathbb{E}[S_\delta(T_\delta)] = \mathbb{E}[T_\delta] \cdot \mathbb{E}[\xi_\delta]\,. \tag{6}$$

Now, consider the process $S(t) = \sum_{i=1}^{t} \xi_y(t)$, where $y = \max\big(x - S(t-1), r\big)$ and $S(0) = 0$. Let $T = \min\{t : S(t) \geq x\}$ be a stopping time. From Equation (5), for all $y > r$ we have that

$$\begin{aligned}
\lim_{\delta \to r^+} \mathbb{E}[\xi_\delta] \cdot \mathbb{E}[T] < \mathbb{E}[\xi_y] \cdot \mathbb{E}[T] < \lim_{\delta \to \infty} \mathbb{E}[\xi_\delta] \cdot \mathbb{E}[T] \quad &\Rightarrow \\
\lim_{\delta \to r^+} \mathbb{E}[\xi_\delta] \cdot \mathbb{E}[T] < \quad \mathbb{E}[S(T)] \quad < \lim_{\delta \to \infty} \mathbb{E}[\xi_\delta] \cdot \mathbb{E}[T] \quad &\Rightarrow \\
\lim_{\delta \to r^+} \mathbb{E}[\xi_\delta)] < \quad \frac{\mathbb{E}[S(T)]}{\mathbb{E}[T]} \quad < \lim_{\delta \to \infty} \mathbb{E}[\xi_\delta]\,. &
\end{aligned}$$

Having $\mathbb{E}[S(T)] = x$ and noting that $\mathbb{E}[T] = \bar{h}(x)$ is in fact the average number of hops over distance x, we obtain Equation (4) and the theorem follows. ■

Given a physical distance x and under a greedy forwarding algorithm, Theorem 1 bounds the expected number of hops over x. As mentioned earlier, a problem that limits the accuracy of the given bounds is the assumption on the convergence of the routing algorithm w.h.p. This assumption might be true when studying dense networks, but it is not applicable to networks of finite node density where a dead-end might be encountered. In the following, we extend the case studied in Theorem 1 to account for such possibilities as well.

B. Greedy Forwarding with Backtracking

We analyze a modified greedy forwarding algorithm which works as follows. At every stage t of the walk, the packet either makes a progress of $+\xi(t)$ towards destination with probability p, or backtracks for a random step size of $-\xi(t)$ with probability $1-p$ in the event of encountering a dead-end. Considering the underlying P.P.P., the probability p is then

$$p = 1 - \exp\big(-\rho\,|\mathcal{H}(\cdot)|\big)\,,$$

where ρ is the intensity of the P.P.P., and $|\mathcal{H}(\cdot)|$ denotes the Lebesgue measure of the hand-off region. The corresponding random walk, hence, is formalized as follows:

$$S(t) = \begin{cases} S(t-1) + \xi(t-1) & \text{with probability } p\,, \\ S(t-1) - \xi(t-1) & \text{with probability } 1-p\,, \end{cases}$$

and $S(0) = 0$. Therefore, $\mathbb{E}[S(t)] = S(t-1) + (2p-1)\,\xi(t-1)$. Consider the process $M(t) = S(t) - t\,(2p-1)\,\mathbb{E}\xi$ for $t > 0$. We first verify that $M(t)$ is a martingale.

$$\begin{aligned}
&\mathbb{E}[M(t+1) \mid M(t)] \\
&\quad = \mathbb{E}\big[S(t+1) - (t+1)(2p-1)\mathbb{E}\xi \,\big|\, S(t) - t(2p-1)\mathbb{E}\xi\big] \\
&\quad = \mathbb{E}\big[S(t) + (2p-1)\xi(t) - (t+1)(2p-1)\mathbb{E}\xi \,\big|\, \cdot\,\big] \\
&\quad = S(t) - t\,(2p-1)\,\mathbb{E}\xi = M(t)\,.
\end{aligned}$$

Define a stopping time $T = \inf\{t : S(t) \geq D\}$. By the optional stopping theorem, $\mathbb{E}[M(T)] = \mathbb{E}[M(0)] = 0$. Thus,

$$\begin{aligned}
\mathbb{E}[S(T)] &= (2p-1)\,\mathbb{E}[T]\,\mathbb{E}[\xi] \quad \Rightarrow \\
x &= (2p-1)\,\mathbb{E}[T]\,\mathbb{E}[\xi] \quad \Rightarrow \\
\mathbb{E}[T] &= \frac{x}{(2p-1)\,\mathbb{E}[\xi]}\,.
\end{aligned} \tag{7}$$

The natural constraint of $\mathbb{E}[T] > 0$ requires that $p > 1/2$ in order for Equation (7) to make sense. As $p \to \frac{1}{2}^+$, $\mathbb{E}[T]$ diverges, which is an intuitive behavior. Also, when $p = 1$, (7) simplifies to (6) which is also expected.

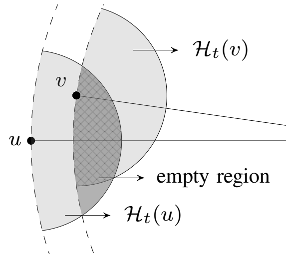

Fig. 2: Succeeding hand-off regions may overlap. Here, the darker shaded region is empty, and part of it, *i.e.*, the crosshatched area, overlaps the hand-off region of v on $v \rightsquigarrow t$.

The bounds given in Theorem 1 are expressed in terms of the expected progress the greedy forwarding algorithm makes per hop when at a limiting distance of ∞ or r from destination. In the next section, we examine the tightness of the suggested bounds in Theorem 1.

IV. Expected Progress per Hop

Aside from the size of the hand-off region, the expected progress per hop also depends on the forwarding policy of the greedy algorithm, *i.e.*, the criteria by which the next relay is chosen from within the set of potential next-hops. Several next-hop selection policies have been proposed in the context of greedy routing algorithms. A widely used policy is to always choose the next-hop with the least remaining distance (LRD) to the destination. Even though this strategy does not guarantee that the packet necessarily travels the fewest hops, it ensures the maximum possible progress towards the destination at every hop.

An issue with the LRD policy is that it violates the required condition on the independence of per-hop progresses. To clarify, observe that the hand-off regions of subsequent hops are not disjoint. For instance, in Fig. 2, the hand-off region of node u overlaps that of node v on $v \rightsquigarrow t$ in the crosshatched region. Therefore, if node v is chosen as next-hop for $u \rightsquigarrow t$ under LRD, then v cannot logically have a potential next-hop in the crosshatched region. This implies that when LRD is used as the forwarding policy, the information from the past history of the walk can affect the future decisions.

To be able to use Theorem 1, we must make sure that the adopted forwarding policy does not violate the independence of succeeding progresses as described above. One such compliant policy is random greedy forwarding (RGF) by which a current relay forwards the packet to a randomly chosen next-hop. Such a next-hop could clearly be located anywhere within its hand-off region, and its election as the next relay does not impose any restriction on the location of subsequent hops. As such, RGF satisfies the required conditions of Theorem 1.

In what follows, we quantify the expected progress per hop under RGF in 2-D space. It is worth mentioning that although RGF is not an optimal forwarding strategy, it can serve as a lower-bound for more aggressive policies such as LRD.

A. A Lower-Bound on $\mathbb{E}[\xi]$

Consider the case when the source and destination are located at a distance $r + \epsilon$ for a small positive $\epsilon \to 0$. In this case, the hand-off region for the source can be approximated by a symmetrical biconvex lens, as illustrated in the left-hand-side of Fig. 1. For the moment, assume $r = 1$ and define the boundaries of the hand-off region as follows:

$$|\omega| = \begin{cases} \sqrt{1-(1-\delta)^2} = \sqrt{2\delta - \delta^2} & \text{if } 0 \le \delta \le \frac{1}{2}\,, \\ \sqrt{1-\delta^2} & \text{if } \frac{1}{2} \le \delta \le 1\,. \end{cases}$$

Due to the symmetry of the region, the enclosed area can be calculated as

$$A(r) = 4 \cdot \int_0^{\frac{1}{2}} \int_0^{\sqrt{2\delta-\delta^2}} \mathrm{d}\omega \, \mathrm{d}\delta = 4 \cdot \int_0^{\frac{1}{2}} \sqrt{2\delta - \delta^2} \, \mathrm{d}\delta\,.$$

Because the next-hop can be located anywhere within the hand-off region with equal probability, $\mathbb{E}[\xi(r)]$ is the expected distance from the relay over the region which can be calculated as follows:

$$\mathbb{E}[\xi(r)] = \frac{2}{A(r)} \Bigg(\int_0^{\frac{1}{2}} \int_0^{\sqrt{2\delta-\delta^2}} \sqrt{\delta^2 + \omega^2} \, \mathrm{d}\omega \, \mathrm{d}\delta + \int_{\frac{1}{2}}^{1} \int_0^{\sqrt{1-\delta^2}} \sqrt{\delta^2 + \omega^2} \, \mathrm{d}\omega \, \mathrm{d}\delta \Bigg)\,.$$

Using numerical methods and noting that $\mathbb{E}[\xi(r)]$ is linear in r, for a general case, we obtain that

$$\lim_{\delta \to r^+} \mathbb{E}[\xi(\delta)] \approx 0.643\, r\,. \tag{8}$$

B. An Upper-Bound on $\mathbb{E}[\xi]$

Consider the right-hand side of Fig. 1. The boundary of the hand-off region is defined as follows:

$$|\omega| = \sqrt{1-\delta^2} \quad \text{for } 0 \le \delta \le 1\,.$$

The area of the hand-off region is clearly $A(\infty) = \pi/2$. Hence,

$$\mathbb{E}[\xi(\infty)] = \frac{2}{A(\infty)} \Bigg(\int_0^1 \int_0^{\sqrt{1-\delta^2}} \sqrt{\delta^2 + \omega^2} \, \mathrm{d}\omega \, \mathrm{d}\delta \Bigg) = \frac{2}{3}\,.$$

By analogy to the previous case, we obtain that

$$\lim_{\delta \to \infty} \mathbb{E}[\xi(\delta)] \approx 0.667\, r\,. \tag{9}$$

From Theorem 1 and Equations (8) and (9), we obtain that, under a routing with RGF policy, the average hop count over any given distance $x \gg r$ is bounded as

$$1.50 \left(\frac{x}{r}\right) < \bar{h}(x) < 1.56 \left(\frac{x}{r}\right)\,. \tag{10}$$

Note that x/r is the theoretical lower-bound on the number of hops under any routing scheme, which, of course, can almost never be attained on a RGG.

V. Analysis of Scalability

Leveraging the mathematical models developed in previous sections, we now examine the scalability conditions of random networks under proximity-based social models. The following theorem identifies a large family of social relationships that let a communication network scale.

Theorem 2. *Under a proximity-based social model identified by the power-law p.d.f.* $f_\alpha(\cdot)$*, an extended or dense* RGG *exhibits scalability if* $\alpha > 3$.

Proof: From Definition 4, the required condition on scalability is to maintain $\mathbb{E}[\mathcal{L}_\alpha] < \infty$ when the network size goes to infinity. From Equation (10) and for a range-adjusted distance measure x we have $\bar{h}(x) < c'x$ for some constant $c' > 0$. Substituting this into Equation (3) and expanding yields

$$\begin{aligned}\mathbb{E}[\mathscr{L}_\alpha] &< \int_1^{\frac{d}{r}} c\,(rx)^{1-\alpha} \cdot c'\,x\,\mathrm{d}x \\ &= c\,c'\,r^{1-\alpha} \int_1^{\frac{d}{r}} x^{\,2-\alpha}\,\mathrm{d}x \quad (\text{assuming } \alpha > 2, \alpha \neq 3) \\ &= c' \cdot \left(\frac{2-\alpha}{3-\alpha}\right) \cdot \left(\frac{d^{\,(3-\alpha)} - r^{\,(3-\alpha)}}{r\,d^{\,(2-\alpha)} - r^{\,(3-\alpha)}}\right). \end{aligned} \tag{11}$$

Examining the convergence conditions for (11), in the following, we prove the result stated by the theorem for the extended and dense models of networks separately.

a) The case for extended networks: Santi and Blough [12] derive the critical transmission range for extended networks as $r(d) = \Theta(\sqrt{\log d})$, where d is the diameter of the network. We are interested in the scaling behavior of the network at the limiting condition of $d \to \infty$. Owing to the fact that d grows much faster than $r(d)$ in case of an extended network, we can write

$$\begin{aligned}\lim_{d\to\infty} \mathbb{E}[\mathscr{L}_\alpha] &< \lim_{d\to\infty} c'\left(\frac{\alpha-2}{\alpha-3}\right)\left(\frac{d^3\,r(d)^{\,\alpha} \quad - r(d)^3\,d^\alpha}{d^2\,r(d)^{(\alpha+1)} - r(d)^3\,d^\alpha}\right) \\ &= c'\left(\frac{\alpha-2}{\alpha-3}\right) < \infty \quad \text{providing that } \alpha > 3\,. \end{aligned}$$

When $\alpha < 3$, the big fraction on the right-hand side above diverges as $d/r(d)$. Therefore, the foregoing limit is finite only if $\alpha > 3$.

b) The case for dense networks: The critical transmission range to ensure connectivity in dense graphs is derived by Gupta and Kumar [13] as $r(n) = \Theta\Big(\sqrt{\frac{\log n}{n}}\Big)$. Similar to the case for extended networks, we write

$$\lim_{n\to\infty} \mathbb{E}[\mathscr{L}_\alpha] < \lim_{n\to\infty} c'\left(\frac{\alpha-2}{\alpha-3}\right)\left(\frac{d^3\,r(n)^{\,\alpha} \quad - r(n)^3\,d^\alpha}{d^2\,r(n)^{(\alpha+1)} - r(n)^3\,d^\alpha}\right).$$

Because $r(n) \xrightarrow{n\to\infty} 0$, the rightmost term becomes an indeterminate form. Assuming $\alpha > 3$ and by applying l'Hopital's rule to the rightmost term 3 times, it is easy to verify that it tends to 1 as $n \to \infty$. Therefore,

$$\lim_{n\to\infty} \mathbb{E}[\mathscr{L}_\alpha] < c'\left(\frac{\alpha-2}{\alpha-3}\right) < \infty \quad \text{providing that } \alpha > 3\,.$$

The divergence of the rightmost term when $\alpha < 3$ is clear noting that $d/r(n)$ would be the dominant term therein.

Hence, $\alpha > 3$ is the necessary condition for scalability in the extended and dense models of RGG's. ■

A similar (though more heuristic) line of study was previously published by Li *et al.* [14] and reported $\alpha > 2$ as the scalability threshold. To clarify, we underline that this apparent inconsistency, in fact, stems from a slight difference in the construction of the models. More precisely, we have considered the effect of greater multiplicity of social contacts at farther geographical distances, while this observation was absent in the previous work of Li *et al.*

VI. Conclusions and Outlook

We investigated how geographical diversity of social interactions can affect the scalability of communication networks. Particularly, we identified a threshold on the spatial diversity of social interactions beyond which the majority of inter-node communications become statistically concentrated within a finite neighborhood around nodes. We showed that this phenomenon enables the underlying communication graph to scale as the number of nodes in the network increases.

The modeling framework we have introduced is very general and the results we have derived apply to the both dense and extended network models. The framework can be used to derive more realistic bounds on the throughput capacity of composite networks as compared against the pessimistic results of Gupta and Kumar [1] which neglects the importance of social relationships.

Acknowledgments

The authors gratefully thank John Musacchio for useful discussions and the anonymous reviewers for their constructive comments on an earlier version of this manuscript. This research was sponsored in part by the Jack Baskin Chair of Computer Engineering at UCSC.

References

[1] P. Gupta and P. Kumar, "The Capacity of Wireless Networks," *IEEE Trans. Inf. Theory*, vol. 46, no. 2, pp. 388–404, Mar. 2000.

[2] M. Grossglauser and D. Tse, "Mobility Increases the Capacity of Ad Hoc Wireless Networks," *IEEE/ACM Trans. Netw.*, vol. 10, no. 4, pp. 477–486, Aug. 2002.

[3] S. Kulkarni and P. Viswanath, "A Deterministic Approach to Throughput Scaling in Wireless Networks," *IEEE Trans. Inf. Theory*, vol. 50, no. 6, pp. 1041–1049, Jun. 2004.

[4] Z. Wang, H. Sadjadpour, and J. Garcia-Luna-Aceves, "A Unifying Perspective on the Capacity of Wireless Ad hoc Networks," in *Proc. IEEE INFOCOM*, 2008, pp. 753–761.

[5] D. Watts and S. Strogatz, "Collective Dynamics of 'Small-World' Networks," *Nature*, vol. 393, no. 6684, pp. 440–442, Jun. 1998.

[6] J. Kleinberg, "The Small-World Phenomenon: An Algorithm Perspective," in *Proc. ACM STOC*, May 2000, pp. 163–170.

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

[8] R. Lambiotte, V. Blondel, C. de Kerchove, E. Huens, C. Prieur, Z. Smoreda, and P. Van Dooren, "Geographical Dispersal of Mobile Communication Networks," *Physica A: Statistical Mechanics and its Applications*, vol. 387, no. 21, pp. 5317–5325, Sep. 2008.

[9] J. Goldenberg and M. Levy, "Distance is not Dead: Social Interaction and Geographical Distance in the Internet Era," Oct. 2009, arXiv:0906.3202v2 [cs.CY].

[10] J. Illenberger, K. Nagel, and G. Flötteröd, "The Role of Spatial Interaction in Social Networks," *Networks and Spatial Economics*, vol. 13, no. 3, pp. 255–282, Sep. 2013.

[11] B. Karp and H.-T. Kung, "GPSR: Greedy Perimeter Stateless Routing for Wireless Networks," in *Proc. ACM MobiCom*, Aug. 2000, pp. 243–254.

[12] P. Santi and D. Blough, "The Critical Transmitting Range for Connectivity in Sparse Wireless Ad hoc Networks," *IEEE Trans. Mobile Comput.*, vol. 2, no. 1, pp. 25–39, Jan. 2003.

[13] P. Gupta and P. Kumar, "Critical Power for Asymptotic Connectivity," in *Proc. IEEE CDC*, Dec. 1998, pp. 1106–1110.

[14] J. Li, C. Blake, D. De Couto, H. Lee, and R. Morris, "Capacity of Ad Hoc Wireless Networks," in *Proc. ACM MobiCom*, Jul. 2001, pp. 61–69.