Les Cahiers du GERAD ISSN: 0711–2440

The feasibility of ultra large-scale distributed networks in symmetrical network typologies

R. Khalvandi, B. Sansò

G–2024–80

December 2024

GERAD HEC Montréal
3000, chemin de la Côte-Sainte-Catherine
Montréal (Québec) Canada H3T 2A7

Tél. : 514 340-6053
Téléc. : 514 340-5665
info@gerad.ca
www.gerad.ca

The feasibility of ultra large-scale distributed networks in symmetrical network typologies

Reza Khalvandi [a, b]

Brunilde Sansò [a, b]

[a] *Département de génie électrique, Polytechnique Montréal, Montréal, (Qc), Canada, H3T 1J4*

[b] *GERAD, Montréal (Qc), Canada, H3T 1J4*

reza.khalvandi-ilezoole@polymtl.ca
brunilde.sanso@polymtl.ca

December 2024
Les Cahiers du GERAD
G–2024–80

Abstract : This study investigates the feasibility of large-scale distributed networks. The core focus of our research is the impact of multi-hop communication on point-to-point capacity per user (C_{P2P}), which is crucial for the scalability of distributed networks. By employing mathematical analysis, we estimate C_{P2P} in a distributed network with symmetrically arranged nodes, accounting for the power-law distribution of interaction probability based on distance. Our findings reveal that the capacity bands achieved surpass existing benchmarks by approximately $\sqrt{\ln(n)}$, depending on the power-law exponent (α) value. Additionally, we present a novel stochastic analysis to determine the power-law exponent from available statistical data on social interactions. Both the mathematical analysis and real-world statistical data on the realistic value of α consistently support the feasibility of very large-scale distributed networks for applications rooted in social interactions.

Keywords : Large-scale ad hoc networks, future distributed networks, wireless network capacity, capacity estimation

1 Introduction

Distributed systems and open networks have transformed the digital landscape, enabling resource-intensive tasks on standard PCs and revolutionizing industries through open software development. Advances in distributed web and edge-cloud computing suggest benefits for local applications through distributed solutions. Web3, a decentralized architecture, contrasts with the centralized control of the traditional Web [2]. However, achieving large-scale, reliable distributed communication remains a challenge. Practical and theoretical hurdles have led to hierarchical and semi-centralized backbone communication networks. Despite these challenges, distributed networking is vital for managing increasing traffic and node numbers, promoting sustainable communication. Our research explores the potential of large-scale distributed networking to create open, infrastructure-less communication networks maintained by individuals. We explore the importance of multi-hop communication in the matter of scalability, particularly in wireless mediums due to their ad hoc and infrastructure-less nature.

A large-scale distributed wireless network could solve key communication challenges. The rise of applications like autonomous vehicles, Unmanned Aerial Vehicles (UAVs), and the Internet of Things (IoT) strains the cellular system, driving the development of 6G technology to support a connection density ten times higher than 5G [7]. Traditional cellular networks may be insufficient, leading to interest in distributed wireless networking as a cost-effective, infrastructure-free solution capable of handling significant traffic. Moreover, distributed networks can offer connectivity to millions without access [3] and serve as backup communication during natural disasters [1].

Despite the potential of distributed networking to address these issues, a robust and scalable distributed solution has yet to be developed due to practical challenges. However, technologies like Artificial Intelligence (AI), precise positioning, blockchain, and distributed processing offer promising solutions. Higher device density enhances network connectivity, making large-scale networking more reliable. Improved hardware, communication capacity, longer battery life, and distributed processing techniques help maintain an accurate network view and calculate efficient routes. Distributed reinforcement learning algorithms are powerful for decision-making in variable network states, paving the way for scalable distributed wireless networks.

Even with this optimistic perspective and renewed interest, distributed networking should still be able to meet the necessary point-to-point capacity (C_{P2P}) as if the increase in network size leads to a significant reduction in C_{P2P}, regardless of practical challenges like physical layer technology, routing, and resource allocation methods, large-scale distributed networking would be impractical. Hence, it is crucial to evaluate the network resources' capacity for expansion, and subsequently, the significance of effective and decentralized network management in resource allocation becomes evident.

In view of the above, this study assesses the feasibility of large-scale distributed wireless networking by examining the point-to-point capacity (C_{P2P}) as the number of nodes (n) approaches infinity ($n \to \infty$). Stochastic analysis is utilized due to the challenges in estimating the end-to-end capacity through simulation for large-scale distributed networks. We will demonstrate, in Section 3, that the upper bound of C_{P2P} is $\frac{\mathbf{E}(C_L)}{\mathbf{E}(d)}$, where $\mathbf{E}(C_L)$ represents the expected concurrent transmission capacity per node, and $\mathbf{E}(d)$ signifies the expected number of hops for point-to-point connections. The value of $\mathbf{E}(C_L)$ is predominantly influenced by physical layer technology, available frequency resources (W), and network topology. Although communication mediums and technologies may differ, ranging from visible light to actual cellphone communication technology, satellite communication, UAVs, or even wired communication, the reflections can be universally represented in $\mathbf{E}(C_L)$.

However, the primary limitation in a large-scale distributed network arises from the inherent characteristics of multi-hop communication, which is the core focus of our study. In such networks, transmission capacity is shared among multiple point-to-point connections. Specifically, if $\mathbf{E}(d)$ increases significantly as the network size grows, C_{P2P} decreases substantially. The value of $\mathbf{E}(d)$ depends on the interaction between nodes over distance and network size. Hence, selecting an appropriate interaction probability model over distance is critical for accurate capacity estimation.

We employ a power-law distribution for the interaction probability between nodes based on real-world data to estimate C_{P2P} in a symmetric node arrangement network topology. While acknowledging that symmetric node arrangement might not realistically reflect real-world scenarios, we have chosen this model for its simplicity as the first step to assess the feasibility of large-scale distributed networking. This decision stems from the complexities such as physical layer issues, network topology, and optimal routing associated with more detailed models. Indeed, a symmetric configuration, despite not encapsulating all the complexities of real-world wireless communications, can provide a foundational understanding of the feasibility of distributed networks. Furthermore, scalability depends on the power-law exponent (α) value [5], which requires analysis in real-world scenarios. To the best of our knowledge, prior to this work, the feasibility of large-scale distributed networking due to the lack of previous investigations into the power-law factor's significance in capacity estimation was largely unexplored, whether in symmetric or random arrangements.

The primary contribution of our work is a comprehensive analysis within a simple framework, yet capable of addressing the scalability question in distributed networking focusing on the estimating capacity bounds C_{P2P}, and deriving α based on statistical data. Hopefully, a key finding is that the statistical data suggests the feasibility of large-scale networks, encouraging us to further research.

Summarizing, the contributions of this work are:

- We employ mathematical analysis and statistical data to estimate the point-to-point capacity per user (C_{P2P}) in a distributed network with symmetrically arranged nodes. This analysis provides valuable insights into the performance of such networks.
- We demonstrate the impact of the power-law distribution of interaction probability between nodes based on their distance. We show that when the power-law exponent (α) exceeds 3, the point-to-point capacity (C_{P2P}) remains constant at $\Theta(1)$, indicating robust performance in large-scale distributed networks.
- We use stochastic analysis to extract the power-law exponent (α) from statistical data. By applying this analysis to data from various empirical studies, the research consistently demonstrates the feasibility of very large-scale distributed networks for applications based on social interactions, emphasizing the practicality of the proposed approach.

The rest of the paper is organized as follows: In Section 2, state-of-the-art related to capacity estimation is detailed. In Section 3, we examine a scenario where network nodes are symmetrically arranged with similar communication technologies, resulting in equal link capacity (C_L) between neighboring nodes. We estimate the point-to-point capacity, declared as C_{P2P}, in this scenario and compute the value of $\mathbf{E}(d)$. In Section 4, we demonstrate that the aforementioned scenario in Section 3, can be generalized to represent a power-law distribution interaction probability model where we approximated $\mathbf{E}(d)$ and consequently estimated C_{P2P} bounds of a power-law distribution model. Furthermore, in Section 5, a stochastic analysis to extract the power law exponent (α) from the data is performed. Finally, in Section 6, we integrate the results from Sections 4 and 5 where utilizing existing empirical studies to determine the feasibility of large-scale distributed networking in the real world. Finally, we resume our work, express our main conclusion, and explain the limitation of the work and potential of future works in Section 7.

2 State of the art

2.1 Capacity estimation

The seminal work by Gupta and Kumar [11] was the first to use stochastic estimation to evaluate Ad Hoc network capacity that assumes interaction probability is not dependent on the distance where $C_{P2P} = \Theta(\frac{W}{\sqrt{n}})$. Its findings indicate that practical Ad Hoc networking should be restricted to at most a few thousand wireless nodes, and cannot be scaled for larger numbers of nodes since

$\lim_{n\to\infty} C_{P2P} = 0$. However, research has shown that distance plays a crucial role in social interaction. Latané et al. [15] conducted an experiment that demonstrated a strong correlation between social interaction and distance, and they developed a power-law distribution for interaction probability based on their findings. The relationship between the number of interactions and distance follows a linear trend when plotted on a log-log scale. Similarly, Backstrom et al. [6] observed the same results in a more recent study of social networks that shows that the probability of friendship over distance has a power-law distribution.

Table 1: Parameters description.

Describe	Parameter
n	Number of nodes in network
m	Number of levels
l	Physical link distance
C_L	Single hope transmission capacity (bps)
$\mathbf{E}(C_L)$	The expected value of the transmission rate per node (bps)
W	Available frequency (Hz)
$\mathbf{E}(d)$	Expected hops number of the point-to-point connections
C_{P2P}	Expected average point to point capacity
α	Power-law distribution parameter
β	Exponential distribution parameter
q	Number of nodes at the first level
a	Area of the first level
P^0	Interaction probability constant
P_i	Node's interaction probability of level i^{th}
N_i	Number of nodes in level i^{th}
$\mathbf{E}_i(d)$	Average path length of level i^{th}
γ	Interaction probability constant
ζ	Exponential step size
d	Source-destination distance per hop
$C(d)$	Number of contacts over distance (d)
$\mathbf{E}(C_i)$	Expected number of contacts at level i^{th}
$P(d)$	Interaction probability over distance (d)
$SINR$	The ratio of signal to interference plus noise

In fact, the probability of interaction over distance (d) can be represented mathematically as $\frac{1}{d^\alpha}$ where α is the power-law exponent, and the distance between the nodes is represented by d. Li et al. [16] utilized this power-law distribution to estimate C_{P2P}. According to their findings, when the probability of longer paths decreases fast (with $\alpha > 2$) in a one-dimensional network, C_{P2P} is $\Theta(1)$ and it remains constant regardless of the network's growth rate. This emphasizes the significance of the source-destination interaction probability model when comparing the results of Gupta et al. [11] with those of Li et al. [16]. Azimdoost et al. [5] examines the capacity of wireless networks in the context of social networks and utilizes a power-law distribution model for path length where each individual in the network has a finite number of contacts (q, where $q < \infty$ in most cases), and the distribution of these contacts follows a pattern of ($\frac{1}{d^\alpha}$). Their research reveals that when $q < \infty$ and $\alpha > 3$, the point-to-point capacity (C_{P2P}) decreases only by the order of $\Theta(\frac{W}{\log(n)})$. In a similar vein, Fu et al. [10] conducted a comparable investigation on the capacity of wireless networks incorporating social attributes, considering both unicast and multicast communication scenarios. Wei et al. [20] determined the potential capacity of three-dimensional wireless social networks by considering the advancements in aeronautical telecommunication and UAVs. The outcomes of their research indicate that when a significant concentration of social groups or contact points surpasses a threshold value ($\alpha > 4$), the wireless social network exhibits scalability. Moreover, the capacity of these networks is significantly greater compared to a wireless network that lacks social behavior. Hou et al. [14] investigated the capacity of hybrid networks that consist of both Ad Hoc and cellular transmissions. They also demonstrated that distance plays a crucial role in social interaction, as a general rule.

In capacity estimation, the distribution of the node's position is also a significant factor, particularly concerning physical layer issues. For instance, in an optimally placed scenario where nodes are

symmetrically arranged, Gupta and Kumar [11] demonstrated that $\mathbf{E}(C_L)$ is $\Theta(W)$, where W is total available frequency resources. On the other hand, when the nodes are randomly distributed throughout the network area, $\mathbf{E}(C_L)$ is $\Theta(\frac{W}{\sqrt{\ln(n)}})$ [11]. In Alfano et al. [4], nodes were positioned based on a shot-noise Cox process, allowing for the modeling of clustering behavior in large-scale systems. However, they did not differentiate the distance effect in interaction probability. He et al. [12] estimated the transmission capacity of vehicular Ad Hoc network (VANET) where the car-following model was introduced to describe the distribution of vehicles and a transmission capacity upper bound in a large-scale environment was calculated. Zhou et al. [21] conducted a study to estimate the asymptotic capacity and delay of large social-aware mobile Ad Hoc wireless networks (MANETs). They focused on a constrained mobility model and a rank-based social model. The researchers examined the upper bound of capacity in such networks using the protocol interference model. Qin et al. [17] estimated the transport capacity of full-duplex Ad Hoc networks by limiting communication pairs to a maximum distance, and they included the distance factor in interaction probability but with a uniform probability distribution within a designated communication radius. In the study by Hou et al. [13], the effect of beam-forming technology on improving capacity was examined, particularly its role in reducing interference. Although their work mainly covers the physical layer, not the application layer, beamforming can significantly increase C_L in an Ad Hoc network and correspondingly C_{P2P}. Currently, 3-D networks are becoming more prevalent, and Wang et al. [19] examined capacity in a network model similar to that of [11], but with the addition of multi-beam directional antenna technology, which was compared to omnidirectional transmission technology. Cheng et al. [8] proposed a 3D-MANET cell-gridded network model based on Zipf's law, which can simulate various node distribution scenarios in 3D space by modifying a specific parameter. The study deduced and analyzed the packet delivery rate, network capacity, and delay performance of the 3D-MANET.

2.2 Empirical work on social tie

Despite the extensive research examining critical factors influencing distributed wireless network capacity, such as issues at the wireless physical layer, the randomness of node distribution, emergent technologies, and interaction probabilities, the feasibility of very large-scale distributed networking from a capacity standpoint remains an open question. Consequently, this study not only computes C_{P2P} bounds premised on the power law distribution but also scrutinizes the real-world instances of the α parameter. We introduce a mathematical formulation for deriving α from statistical data, accompanied by empirical analyses to ascertain this parameter, akin to the inquiries by Small et al. [18], which delved into the correlation between interaction probability and distance through empirical data. In Section 6, we integrate three pertinent studies [6,9,15] for the mathematical extraction of the power law exponent in Section 5. Leveraging real-world data, we initiate a discourse on the scalability of distributed networks.

3 Case study

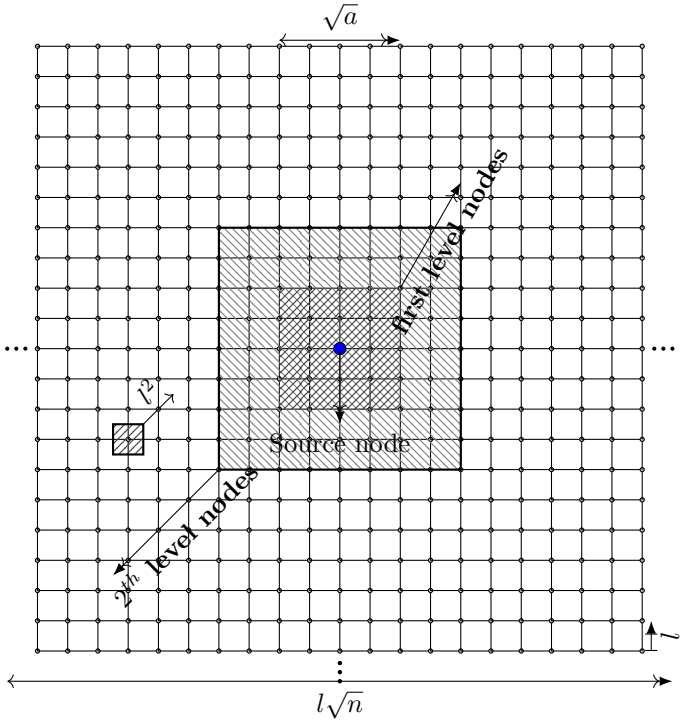
The initial phase involves a stochastic analysis of available point-to-point capacity (C_{P2P}). In a distributed network, any link capacity (C_L) is divided between different point-to-point connections. The aggregate transmission rate across all concurrent links can be represented by $\sum_{i=1}^{n} C_L^i$, where for each node i, $C_L^i \geq 0$. Considering $\mathbf{E}(C_L)$ to denote the average transmission capacity for each node, the overall expected transmission rate can thus be expressed as $\mathbf{E}(\sum_{i=1}^{n} C_L^i) = \sum_{i=1}^{n} \mathbf{E}(C_L) = n\mathbf{E}(C_L)$. On the other hand, if any node j can transmit data on a point-to-point connection with the average rate C_{P2P} and path length d_j, then total concurrent transmission resources used by all point-to-point connections is $\sum_{j=1}^{n} C_{P2P} d_j$ where $\mathbf{E}(\sum_{j=1}^{n} C_{P2P} d_j) = C_{P2P}\mathbf{E}(\sum_{j=1}^{n} d_j) = C_{P2P}\mathbf{E}(\sum_{j=1}^{n} \mathbf{E}(d)) = nC_{P2P}\mathbf{E}(d)$ where the expected number of hops of point-to-point connections is $\mathbf{E}(d)$. This value should be equal to or lower than the expected value of the total transmission rate in the network

so that:

$$nC_{P2P}\mathbf{E}(d) \leq n\mathbf{E}(C_L) \Rightarrow C_{P2P} \leq \frac{\mathbf{E}(C_L)}{\mathbf{E}(d)}\,.$$

Therefore, the upper bound of the C_{P2P} is $\frac{\mathbf{E}(C_L)}{\mathbf{E}(d)}$. In our model, nodes are arranged in a symmetric topology, positioned at the vertices of a square lattice, as shown in Figure 1. They create a distributed network where every node has the same transmission capacity (C_L) with its neighboring nodes, and therefore $\mathbf{E}(C_L) = C_L$. The physical link distance between a node and its neighboring node is l. By assuming a square around each node with the area of l^2 the total network area is covered without overlap. Therefore, node density is $\frac{1}{l^2}$ and the total network area is almost $l^2 n$, given that the total number of nodes within the network is n. In this scenario, as all links have the same link length and resources, consequently the optimal routing is the physically shortest path [11], thus, the problem of capacity estimation is reduced to estimate the expected number of hops ($\mathbf{E}(d)$).

Figure 1: The lattice structure of the assumed distributed network.



To estimate $\mathbf{E}(d)$, we utilized an exponentially decreasing distance-dependent distribution in which nearby nodes have the highest probability of communicating with each other, and the interaction probability decreases exponentially by a factor of β across several levels as the source-destination distance increases. Therefore, the total network area is divided into multiple nested levels around the source node. To illustrate this, let us assume that the source node is at the center of the network, as shown in Figure 1. In this scheme, we map a generic square of area a m^2 around the source node. Nodes inside this square are grouped in the first level, and any node inside this square has an interaction probability of $P_1 = P^0$ with the source node. In the second level, where nodes are located between squares of side lengths $\sqrt{a}$ and $2\sqrt{a}$, the node's interaction probability with the source node is $P_2 = \beta P^0$. For nodes in the i^{th} level ($i \geq 2$), that are between the squares of side length $2^{i-2}\sqrt{a}$ and $2^{i-1}\sqrt{a}$, $P_i = \beta^{i-1}P^0$. Therefore, the interaction probability per level ($i \geq 1$) is given by:

$$P_i = P^0\beta^{i-1}\,. \tag{1}$$

Here, we compute the number of nodes in each level. The node density is $\frac{1}{l^2}$, and we can estimate that the total number of nodes within the initial square, as well as the first level, that is approximately $\frac{a}{l^2}$. However, the exact number of nodes depends on whether the initial square vertices intersect with any of the nodes in the network or not. In fact, the number of nodes can range between $\frac{a}{l^2}$ to $(\frac{\sqrt{a}}{l}+1)^2$. Nonetheless, if $\frac{\sqrt{a}}{l} \gg 1$, we can approximate the number of nodes as:

$$q = \frac{a}{l^2}\,. \tag{2}$$

Therefore, if the first square covers q nodes, the second square, which has an area four times as large, covers $4q$ nodes. Moreover, nodes in the second level are surrounded by two squares of side lengths of $\sqrt{a}$ and $2\sqrt{a}$, and therefore the number of nodes in the second level is $N_2 = 4q - q = 3q$. For outer levels, the number of nodes increases by a factor of four each time the level is increased. Thus, the number of nodes in the i^{th} level ($i \geq 2$) is $N_i = 3q2^{2(i-2)}$. The number of nodes in the first level except the source node is $q-1$. However, as we considered $\frac{\sqrt{a}}{l} \gg 1$ that means also $q \gg 1$ (Equation (2)), we assume $q - 1 \approx q$. Therefore, the number of nodes in each level N_i is given by Equation (3).

$$N_i = \begin{cases} q & i = 1 \\ (2^2-1)q2^{2(i-2)} = 3q2^{2(i-2)} & i \geq 2\,. \end{cases} \tag{3}$$

Now, If we keep increasing levels to cover the whole network area, the network side length ($l\sqrt{n}$) is approximately equal to the square side length that encompasses nodes in the last level ($\sqrt{a}2^m$) as well as all network nodes. Hence, the number of levels (m) is calculated as:

$$\sqrt{a}2^m = l\sqrt{n} \rightarrow m = \log_2(l\sqrt{\frac{n}{a}}) = \log_2(\sqrt{\frac{n}{q}})\,. \tag{4}$$

Before estimating $\mathbf{E}(d)$, we need to determine P^0. When a source node selects a destination from all other nodes, all of the $n-1$ nodes have a non-zero probability of p_j to be chosen as the destination so that $\sum_{j=1}^{n-1} p_j = 1$. As all nodes at the same level have an equal probability of being chosen, $\sum_{j=1}^{n-1} p_j = 1$ can be reformulated as Equation (5) that is the summation of the interaction probability times the number of nodes over any level.

$$\sum_{i=1}^{m} N_i P_i = 1\,. \tag{5}$$

We determine P^0 in two scenarios: when the source node is located in the center and in the corner of the network (which represent the best and worst case scenarios, respectively see Figure 2). In both the center and corner arrangements, there are $N_1 = q$ nodes with a probability of $P_1 = P^0$ and $N_i = 3q2^{2(i-2)}$ nodes in the i^{th} level with a probability of $P_i = \beta^{i-1}P^0$ ($2 \leq i \leq m$) could be chosen as the destination. Therefore, by using the value of P_i and N_i (Equations (1) and (3)) in the Equation (5), Equation (6) is achieved as

$$qP^0 + 3q\sum_{i=2}^{m} P^0\beta^{i-1}4^{i-2} = 1\,. \tag{6}$$

Here, $\sum_{i=2}^{m} P^0\beta^{i-1}4^{i-2}$ could be written as $\sum_{i=0}^{m-2} P^0\beta(4\beta)^i$. For the case of the $4\beta = 1$, $\sum_{i=0}^{m-2} P^0\beta(4\beta)^i = \sum_{i=0}^{m-2} P^0\beta = (m-1)P^0\beta$. Otherwise, $\sum_{i=0}^{m-2} P^0\beta(4\beta)^i$ is a geometric series summation so that:

$$\sum_{i=0}^{m-2} P^0\beta(4\beta)^i = \frac{P^0\beta((4\beta)^{m-1}-1)}{4\beta-1}\,.$$

Consequently, If $m \gg 1$:

$$\sum_{i=0}^{m-2} P^0\beta(4\beta)^i = \begin{cases} \frac{P^0\beta}{1-4\beta} & 4\beta < 1 \\ (m-1)P^0\beta & 4\beta = 1 \\ \frac{P^0\beta(4\beta)^{m-1}}{4\beta-1} & 4\beta > 1\,. \end{cases}$$

Now, by using Equation (4), $m-1 = \log_2(\sqrt{\frac{n}{q}}) - 1$. It gives us:

$$\sum_{i=0}^{m-2} P^0\beta(4\beta)^i = \begin{cases} \frac{P^0\beta}{1-4\beta} & \beta < \frac{1}{4} \\ (\log_2(\sqrt{\frac{n}{q}}) - 1)P^0\beta & \beta = \frac{1}{4} \\ \frac{P^0\beta(4\beta)^{\log_2(\sqrt{\frac{n}{q}})-1}}{4\beta-1} & \beta > \frac{1}{4}\,. \end{cases}$$

By substituting $\sum_{i=0}^{m-2} P^0\beta(4\beta)^i$ values form the above equation in Equation (6), for the case $\beta < \frac{1}{4}$,

$$qP^0 + 3q\frac{P^0\beta}{1-4\beta} = 1 \Rightarrow P^0 = \frac{1}{q(1+\frac{3\beta}{1-4\beta})}\,.$$

If $\beta = \frac{1}{4}$, $qP^0 + 3q(\log_2(\sqrt{\frac{n}{q}}) - 1)P^0\beta = 1$, and then:

$$P^0 = \frac{1}{q(1-3\beta+3\beta\log_2(\sqrt{\frac{n}{q}}))}\,.$$

If $\beta > \frac{1}{4}$, $qP^0(1 + \frac{3\beta(4\beta)^{\log_2(\sqrt{\frac{n}{q}})-1}}{4\beta-1}) = 1$. Since $1 \ll \frac{3\beta(4\beta)^{\log_2(\sqrt{\frac{n}{q}})-1}}{4\beta-1}$, we can approximate P^0 as $P^0 = \frac{4(4\beta-1)}{3q(4\beta^{\log_2(\sqrt{\frac{n}{q}})})}$. Finally, since $a^{\log(b)} = b^{\log(a)}$, $P^0 = \frac{4(4\beta-1)}{3q(\sqrt{\frac{n}{q}}^{\log_2(4\beta)})}$. The general form of the P^0 is expressed in Equation (7) as

$$P^0 = \begin{cases} \frac{1}{q(1+\frac{3\beta}{1-4\beta})} & \beta < \frac{1}{4} \\ \frac{1}{q(1-3\beta+3\beta\log_2(\sqrt{\frac{n}{q}}))} & \beta = \frac{1}{4} \\ \frac{4(4\beta-1)}{3q(\sqrt{\frac{n}{q}}^{\log_2(4\beta)})} & \beta > \frac{1}{4}\,. \end{cases} \tag{7}$$

Here, we start calculating $\mathbf{E}(d)$ that depends on the position of the source node in the network. Hence, we considered two boundary situations where the maximum and minimum of $\mathbf{E}(d)$ occurs. In the best-case scenario, the source node is in the center of the network, and in the worst case at the corner. Figure 2 shows these scenarios. First of all, as all nodes in the same level have an equal interaction probability to communicate with the source node, and therefore $\mathbf{E}(d)$ can be calculated per level so that if $\mathbf{E}_i(d)$ is average path length per level,

$$\mathbf{E}(d) = \sum_{i=1}^{m} P_i N_i \mathbf{E}_i(d)\,. \tag{8}$$

Now, we calculate $\mathbf{E}(d)$ level by level starting with the average path length for the best-case scenario. The average path length is the summation of the distance of all nodes from the source node divided by the number of nodes. Nodes in the first level can be divided into 8 symmetric isosceles triangles as shown in Figure 3. Hence, by considering one of these isosceles triangles, the average path length can be calculated for the first level. Any triangle include $\frac{q}{8}$ nodes, and $\frac{\sqrt{q}}{2}$ nodes on its equal length sides. Nodes in the triangle can be separated into $\frac{\sqrt{q}}{2}$ layers ($\{i \mid 1 \le i \le \frac{\sqrt{q}}{2}\}$) where there are $i+1$ nodes in

the layer i^{th}, see Figure 3. The node j shortest distance from center node in the set of $\{j \mid 0 \leq j \leq i\}$, is $(i+j)$. All nodes distance summation in layer i^{th} is $\sum_{j=1}^{i}(i+j)$. All i layers nodes distances summation that covers all nodes in the triangle is $\sum_{i=1}^{\frac{\sqrt{q}}{2}}\sum_{j=1}^{i}(i+j) = \sum_{i=1}^{\frac{\sqrt{q}}{2}}(i^2+\sum_{j=1}^{i}j) \approx \sum_{i=1}^{\frac{\sqrt{q}}{2}}\frac{3i^2}{2}$ as $\sum_{j=1}^{i}j = \frac{i^2+i}{2} \approx \frac{i^2}{2}$. $\mathbf{E}_1(d)$ is the cumulative node distance divided by the number of nodes in the first level so that:

$$\mathbf{E}_1(d) = \frac{8}{q}\sum_{i=1}^{\frac{\sqrt{q}}{2}}\sum_{j=1}^{i}(i+j) \approx \frac{8}{q}\sum_{i=1}^{\frac{\sqrt{q}}{2}}\frac{3i^2}{2}\,.$$

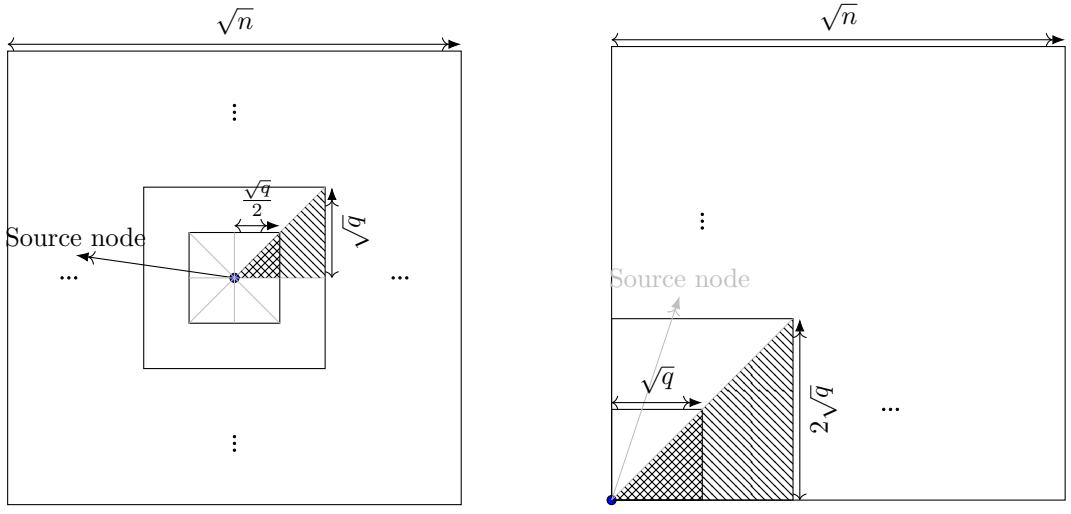
We also have $\sum_{i=1}^{n} i^2 = \frac{n(n+1)(2n+1)}{6}$. If we substitute $\sqrt{q}$ by $2k$, then $\mathbf{E}_1(d)$ could be written as $\mathbf{E}_1(d) = \frac{8}{4k^2}\sum_{i=1}^{k}\frac{3i^2}{2} = \frac{k(k+1)(2k+1)}{2k^2}$ where $\lim_{k\gg 1}\frac{k(k+1)(2k+1)}{2k^2} = k$. Finally, when $k \gg 1$ or $q \gg 1$, $\mathbf{E}_1(d) = \frac{\sqrt{q}}{2}$. In the second level, there are $3q$ nodes and 8 partial triangles (partial triangle side covers nodes between $\frac{\sqrt{q}}{2}$ and $\sqrt{q}$ hops from the source node, Figure 2). Hence, $\mathbf{E}_2(d)$ is

$$\mathbf{E}_2(d) = \frac{8}{3q}\sum_{i=\frac{\sqrt{q}}{2}}^{\sqrt{q}}\sum_{j=1}^{i}(i+j) \approx \frac{8}{6q}((2\frac{\sqrt{q}}{2})^3 - (\frac{\sqrt{q}}{2})^3) = \frac{7\sqrt{q}}{6}\,.$$

Because of geometric similarity, the average distance will be doubled each time we increase the level number. Hence, in the $i^{th}(i \geq 2)$ level, $\mathbf{E}_i(d) = \frac{7\sqrt{q}}{6}2^{i-2}$, and average path length of all levels is

$$\mathbf{E}_i(d) = \begin{cases} \frac{\sqrt{q}}{2} & i = 1 \\ \frac{(2^3-1)\sqrt{q}}{6}2^{i-2} & i \geq 2\,. \end{cases} \tag{9}$$

Figure 2: Boundary scenarios for expected path values.



(a) Best case scenario **(b) Worst case scenario**

In the worst-case scenario, the source node is assumed in the lower left edge. Therefore, the first square is divided into two triangles with $\sqrt{q}$ nodes in its side length as shown in Figure 2 and the expected path calculated as below:

$$\mathbf{E}_1(d) = \frac{2}{q}\sum_{i=1}^{\sqrt{q}}\sum_{j=1}^{i}(i+j) \approx \frac{1}{q}(2\frac{\sqrt{q}}{2})^3 = \sqrt{q}\,.$$

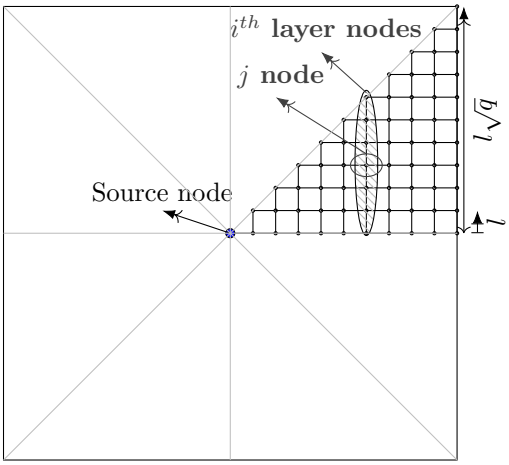
In the second level, its square is split into 2 partial triangles (partial triangle side cover nodes between $\sqrt{q}$ and $2\sqrt{q}$ hops from the source node). Thus, the average path value is

$$\mathbf{E}_2(d) = \frac{2}{3q}\sum_{i=\sqrt{q}}^{2\sqrt{q}}\sum_{j=1}^{i}(i+j) \approx \frac{1}{3q}((2\sqrt{q})^3 - (\sqrt{q})^3) = \frac{7\sqrt{q}}{3}\,.$$

Correspondingly, the i^{th} level average path is $\frac{7\sqrt{q}}{3}2^{i-2}$, and therefore

$$\mathbf{E}_i(d) = \begin{cases} \sqrt{q} & i = 1 \\ \frac{(2^3-1)\sqrt{q}}{3}2^{i-2} & i \geq 2\,. \end{cases} \tag{10}$$

Figure 3: Isosceles triangles of initial square and node arrangement inside each triangle.



The overall expected path ($\mathbf{E}(d)$) is a weighted average between all levels (Equation (8)). Moreover, the number of nodes (N_k) and interaction probability (P_k) in each level for both of the best case and worst case scenarios is equal and only worst case $\mathbf{E}_i(d)$ is twice best case $\mathbf{E}_i(d)$ (Equations (9) and (10)). Therefore, the $\mathbf{E}(d)$ for the best-case is half of the worst case $\mathbf{E}(d)$, and regardless of source node position in the network, $\mathbf{E}(d)$ is not lower than half of the worst case expected path. To avoid further complexity and without loss of generality, we consider the worst-case scenario as the overall expected path to achieve an upper band.

To finalize $\mathbf{E}(d)$ estimation, by using the value of P_k, N_k, and $\mathbf{E}_i(d)$ form Equations (1), (3) and (10) and putting them in Equation (8) we have:

$$\begin{aligned}\mathbf{E}(d) =& qP^0\sqrt{q} + \sum_{i=2}^{m} P^0\frac{7}{3}\sqrt{q}3q\beta^{i-1}2^{i-2}4^{i-2} \\ =& qP^0\sqrt{q}(1 + 7\sum_{i=0}^{m-2}\beta(8\beta)^i)\,.\end{aligned}$$

$\sum_{i=0}^{m-2}\beta(8\beta)^i$ is a geometric series so that if $m \gg 1$, $\sum_{i=0}^{m-2}\beta(8\beta)^i = \begin{cases} \frac{\beta}{1-8\beta} & \beta < \frac{1}{8} \\ (\log_2(\sqrt{\frac{n}{q}}) - 1)\beta & \beta = \frac{1}{8} \\ \frac{(\log_2(\sqrt{\frac{n}{q}}))^{\log_2(8\beta)}}{8(8\beta-1)} & \beta > \frac{1}{8}\,. \end{cases}$

By simplifying the above equation where $m \gg 1$ with the same method that we calculated P^0, and by using P^0 values of the Equation (7),

$$\mathbf{E}(d) = \begin{cases} \frac{\sqrt{q}(1+\frac{7\beta}{1-8\beta})}{(1+\frac{3\beta}{1-4\beta})} & \beta < \frac{1}{8} \\ \frac{\sqrt{q}(1-7\beta+7\beta\log_2(\sqrt{\frac{n}{q}}))}{(1+\frac{3\beta}{1-4\beta})} & \frac{1}{8} \\ \frac{7\sqrt{q}(\sqrt{\frac{n}{q}}^{\log_2(8\beta)})}{8(8\beta-1)(1+\frac{3\beta}{1-4\beta})} & \frac{1}{8} < \beta < \frac{1}{4} \\ \frac{7\sqrt{q}(\sqrt{\frac{n}{q}}^{\log_2(8\beta)})}{8(8\beta-1)(1-3\beta+3\beta\log_2(\sqrt{\frac{n}{q}}))} & \beta = \frac{1}{4} \\ \frac{(4\beta-1)7\sqrt{q}(\sqrt{\frac{n}{q}}^{\log_2(8\beta)})}{(8\beta-1)6(\sqrt{\frac{n}{q}}^{\log_2(4\beta)})} & \beta > \frac{1}{4}\,. \end{cases} \tag{11}$$

If $\beta = \frac{1}{4}$, $\sqrt{q}(\sqrt{\frac{n}{q}}^{\log_2(8\beta)}) = \sqrt{q}(\sqrt{\frac{n}{q}}) = \sqrt{n}$. Moreover, if $\beta > \frac{1}{4}$, then $\frac{\sqrt{q}(\sqrt{\frac{n}{q}}^{\log_2(8\beta)})}{(\sqrt{\frac{n}{q}}^{\log_2(4\beta)})} = \sqrt{q}(\sqrt{\frac{n}{q}}^{\log_2(\frac{8\beta}{4\beta})})$ $= \sqrt{q}\sqrt{\frac{n}{q}} = \sqrt{n}$. Therefore, these different cases are shown as

$$\mathbf{E}(d) \text{ is } \begin{cases} \Theta(1) & \beta < \frac{1}{8} \\ \Theta(\log_2(n)) & \beta = \frac{1}{8} \\ \Theta(n^{\log_2(2\sqrt{2\beta})}) & \frac{1}{8} < \beta < \frac{1}{4} \\ \Theta(\sqrt{n}) & \beta \geq \frac{1}{4}\,. \end{cases} \tag{12}$$

4 Power-law distribution capacity estimation

In this section, we aim to expand our scenario in Section 3 to estimate $\mathbf{E}(d)$ for a power-law interaction probability distribution. At first, we prove that any interaction probability that is exclusively dependent on source-destination distance should be at least $\Theta(\frac{1}{d^2})$ in Theorem 4.1. Then, we explain that for a power-law distribution, $\mathbf{E}(d)$ can be estimated by using our scenario in Section 3.

Theorem 4.1. In any interaction probability including power-law distribution, that is exclusively distance dependent ($p_i = P(d)$) where n nodes are uniformly distributed so that $n \to \infty$, $P(d)$ should fall faster than $\Theta(\frac{1}{d^2})$, so that $\lim_{d\to\infty} d^2 P(d) = 0$.

Proof. In Appendix A.1, we showed that the necessary condition on the interaction probability over distance is that $\sum^{\infty} P(d)d$ converges. $P(d)$ is considered as $\Theta(\frac{P'(d)}{d^\alpha})$ where $\lim_{d\to\infty} P'(d) = c < \infty$ and $\alpha = \{\max \alpha' \mid P(d) \text{ is } \Theta(\frac{P'(d)}{d^{\alpha'}})\}$. In order to obtain a finite sum for the series $\sum_1^{l\sqrt{n}} P'(d)d^{(1-\alpha)}$, we use Cauchy condensation test that says if $\{a_n\}$ is a positive monotone decreasing sequence, then $\sum_{n=1}^{\infty} a_n$ converges if and only if $\sum_{k=1}^{\infty} 2^k a_{2^k}$ converges. By applying it on $\sum_1^{\sqrt{n}} P'(d)d^{(1-\alpha)}$, we have $\sum P'(2^d)2^{d(2-\alpha)}$. If $\alpha > 2$ and $P'(2^d)$ is bounded, the series will converge while it diverges when $\alpha < 2$ regardless of $P'(2^d)$. When $\alpha = 2$, $\sum P'(2^d)$ converges only if $\sum P'(d)$ is a series such as $\sum_n \frac{1}{\log d \cdot \log\log d \cdots \log^{\circ(k-1)} d \cdot (\log^{\circ k} d)^\lambda}$ where $\lambda > 1$. In other word, to satisfy $\sum_{i=1}^{n-1} P_i = 1$, $P(d)$ should fall faster than $\Theta(\frac{1}{d^2})$ so that $\lim_{d\to\infty} d^2 P(d) = 0$. □

Based on Theorem 4.1 any interaction probability, including power-law distribution, that is exclusively distance dependent should be at least $\Theta(\frac{1}{d^2})$.

Here, we want to estimate $\mathbf{E}(d)$ in a symmetric node arrangement scenario such as the one presented in Section 3, where interaction probability follows a power law distribution ($P(d) = \frac{\gamma}{d^\alpha}$) and γ is an interaction probability constant. All nodes at the same distance from a source node have the same interaction probability with the source node. Hence, to estimate $\mathbf{E}(d)$, we can group all nodes in the distance $d = i$ from the source node, calculate $\mathbf{E}(d)$ separately for any group, and then sum up over

all groups. However, when $i \gg 1$, there is no sensible difference between the interaction probability of nodes in the distance $d = i$ and $d = i+1$ as $\frac{\gamma}{i^\alpha} \approx \frac{\gamma}{(i+1)^\alpha}$. Therefore, we change step size from $1, 2, ..., i$ to $1, \zeta^2, ..., \zeta^i$ ($\zeta > 1$), and nodes are grouped in multiple nested levels. If, as same as Section 3, first level nodes are included in a square of area a, and all nodes between distance $\sqrt{a}\zeta^{i-1}$ and $\sqrt{a}\zeta^i$ have the same interaction probability. Here, we approximate the interaction probability of the level i^{th} with the value of $P(d)$ at distance $d = \sqrt{a}\zeta^{i-1}$, that is $\frac{\gamma}{\sqrt{a}^\alpha \zeta^{\alpha(i-1)}}$. We can consider different reference points or an average of $P(d)$ from several points at any level. However, it should not change the ratio of interaction probability between consecutive levels. In this situation, regardless of the selected point (or points), our results show it does not affect the value of $\mathbf{E}(d)$.

In this scenario, the ratio of interaction probability between two consecutive levels i and $i+1$ is $\frac{\gamma}{\sqrt{a}^\alpha \zeta^{i\alpha}} \frac{\sqrt{a}^\alpha \zeta^{(i-1)\alpha}}{\gamma} = \frac{1}{\zeta^\alpha}$. Therefore, by considering $\beta = \zeta^{-\alpha}$ and $P^0 = \frac{\gamma}{a^{\frac{\alpha}{2}}}$, we can write that the interaction probability for nodes in the level i is $P_i = P^0 \beta^{i-1}$. In fact, it can be a generalized form of the aforementioned scenario, Section 3, if step size changes from 2 to ζ, $P^0 = \frac{\gamma}{a^{\frac{\alpha}{2}}}$, $\beta = \zeta^{-\alpha}$ and $\sqrt{q} = \frac{\sqrt{a}}{l}$. Therefore, in the same manner as Equation (1), P_i is expressed as

$$P_i = \frac{\gamma}{a^{\frac{\alpha}{2}}} \zeta^{-(i-1)(\alpha)} . \tag{13}$$

The initial level contains $N_1 = q = \frac{a}{l^2}$ nodes. With the change of step size from 2 to the general parameter ζ, the number of nodes in two consecutive levels is related by $N_{i+1} = N_i \zeta^2$ for $i \geq 2$. Hence, the number of nodes in the second level is $N_2 = \frac{a}{l^2}\zeta^2 - \frac{a}{l^2} = \frac{a}{l^2}(\zeta^2 - 1)$, and the expression for N_i is derived using the formula presented in Equation (14) as

$$N_i = \begin{cases} \frac{a}{l^2} & i = 1 \\ \frac{a}{l^2}(\zeta^2 - 1)\zeta^{2(i-2)} & i \geq 2 . \end{cases} \tag{14}$$

The average distance in the first level is $\mathbf{E}_1(d) = \sqrt{q} = \frac{\sqrt{a}}{l}$. In the case of two consecutive levels, due to geometric similarity, we can express $\mathbf{E}_{i+1}(d) = \mathbf{E}_i(d)\zeta$ for $i \geq 2$.

Now, for the second level, we calculate $\mathbf{E}_2(d)$ approximation as:

$$\begin{aligned} \mathbf{E}_2(d) &= \frac{2}{N_2} \sum_{i=\sqrt{q}}^{\zeta\sqrt{q}} \sum_{j=1}^{i} (i+j) \approx \frac{1}{(\zeta^2-1)q}((\zeta\sqrt{q})^3 - (\sqrt{q})^3) \\ &= \frac{(\zeta^3-1)\sqrt{q}}{(\zeta^2-1)} = \frac{(\zeta^3-1)\sqrt{a}}{(\zeta^2-1)l} . \end{aligned}$$

Thus, $\mathbf{E}_i(d)$ can also be determined by Equation (15), which corresponds to the Equation (8):

$$\mathbf{E}_i(d) = \begin{cases} \frac{\sqrt{a}}{l} & i = 1 \\ \frac{\sqrt{a}}{l} \frac{(\zeta^3-1)}{(\zeta^2-1)} \zeta^{i-2} & i \geq 2 . \end{cases} \tag{15}$$

The number of levels (m) is also given by

$$\sqrt{a}\zeta^m = l\sqrt{n} \rightarrow m = \log_\zeta(l\sqrt{\frac{n}{a}}) . \tag{16}$$

According to the Equation (5), $\sum_{i=1}^{m} N_i P_i = 1$. By substituting the values of N_i and P_i as given in Equation (13) and Equation (14), respectively, we can express:

$$\frac{a}{l^2} \frac{\gamma}{a^{\frac{\alpha}{2}}} + (\zeta^2 - 1)\frac{a}{l^2} \sum_{i=2}^{m} \frac{\gamma}{a^{\frac{\alpha}{2}}} \zeta^{-\alpha} \zeta^{-\alpha(i-2)} \zeta^{2(i-2)} = 1 .$$

This equality could be simplified as $\frac{a}{l^2}\frac{\gamma}{a^{\frac{\alpha}{2}}}(1+(\zeta^2-1)\zeta^{-\alpha}\sum_{i=2}^{m}\zeta^{(2-\alpha)(i-2)})=1$.

As $\zeta>1$, when $\alpha>2$ and $m\ggg 1$, then $\sum_{i=2}^{m}\zeta^{(2-\alpha)(i-2)}=\frac{1}{1-\zeta^{(2-\alpha)}}$. Hence, $\frac{\gamma}{a^{\frac{\alpha}{2}}}$ is given by

$$\frac{\gamma}{a^{\frac{\alpha}{2}}}=\frac{l^2}{a(1+\frac{(\zeta^2-1)\zeta^{-\alpha}}{1-\zeta^{(2-\alpha)}})}\,. \tag{17}$$

Based on the Equation (8), $\mathbf{E}(d)$ is $\sum_{i=1}^{m}P_iN_i\mathbf{E}_i(d)$. By substituting the values of P_i, N_i, and $\mathbf{E}_i(d)$ as stated in Equation (13), Equation (14), and Equation (15) respectively, we can express $\mathbf{E}(d)$ as follows:

$$\mathbf{E}(d)=(\frac{\sqrt{a}}{l})^3\frac{\gamma}{a^{\frac{\alpha}{2}}}+\sum_{2}^{m}\frac{\gamma}{a^{\frac{\alpha}{2}}}(\zeta^3-1)(\frac{\sqrt{a}}{l})^3\zeta^{-\alpha}\zeta^{-\alpha(i-2)}\zeta^{3(i-2)}\,.$$

By using Equation (17) where $\alpha>2$:

$$\mathbf{E}(d)=\frac{\sqrt{a}}{l(1+\frac{(\zeta^2-1)\zeta^{-\alpha}}{1-\zeta^{(2-\alpha)}})}(1+(\zeta^3-1)\zeta^{-\alpha}\sum_{i=2}^{m}\zeta^{(3-\alpha)(i-2)})\,.$$

Again if $m\ggg 1$ and by using Equation (16):

$$\sum_{i=2}^{m}\zeta^{(3-\alpha)(i-2)}=\begin{cases}\frac{1}{1-\zeta^{(3-\alpha)}} & \alpha>3\\ (\log_\zeta(l\sqrt{\frac{n}{a}})-1) & \alpha=3\\ \frac{\zeta^{(3-\alpha)(\log_\zeta(l\sqrt{\frac{n}{a}})-1)}}{\zeta^{3-\alpha}-1} & 2<\alpha<3\,.\end{cases}$$

For the case of $2<\alpha<3$, as $\frac{\zeta^{(3-\alpha)(\log_\zeta(l\sqrt{\frac{n}{a}})-1)}}{\zeta^{3-\alpha}-1}=\frac{\zeta^{(3-\alpha)\log_\zeta(l\sqrt{\frac{n}{a}})}}{(\zeta^{(3-\alpha)})(\zeta^{3-\alpha}-1)}$, and $\zeta^{(3-\alpha)\log_\zeta(l\sqrt{\frac{n}{a}})}=(l\sqrt{\frac{n}{a}})^{\log_\zeta(\zeta^{(3-\alpha)})}=(\frac{l^2n}{a})^{(\frac{3-\alpha}{2})}$, then:

$$\frac{\zeta^{(3-\alpha)(\log_\zeta(l\sqrt{\frac{n}{a}})-1)}}{\zeta^{3-\alpha}-1}=\frac{(\frac{l^2n}{a})^{(\frac{3-\alpha}{2})}}{(\zeta^{(3-\alpha)})(\zeta^{3-\alpha}-1)}\,.$$

Therefore, $\mathbf{E}(d)$ can be estimated for power-law distribution in the same way as Equation (11) where $\alpha>2$,

$$\mathbf{E}(d)=\begin{cases}\frac{\sqrt{a}(1+\frac{(\zeta^3-1)\zeta^{-\alpha}}{1-\zeta^{(3-\alpha)}})}{l(1+\frac{(\zeta^2-1)\zeta^{-\alpha}}{1-\zeta^{(2-\alpha)}})} & \alpha>3\\ \frac{\sqrt{a}(\zeta^{-3}+(1-\zeta^{-3})\log_\zeta(l\sqrt{\frac{n}{a}}))}{l(1+\frac{(\zeta^2-1)\zeta^{-3}}{1-\zeta^{-1}})} & \alpha=3\\ \frac{\sqrt{a}(\zeta^3-1)\zeta^{-\alpha}(\frac{l^2n}{a})^{(\frac{3-\alpha}{2})}}{l(\zeta^{(3-\alpha)})(\zeta^{(3-\alpha)}-1)(1+\frac{(\zeta^2-1)\zeta^{-\alpha}}{1-\zeta^{(2-\alpha)}})} & 2<\alpha<3\,.\end{cases} \tag{18}$$

For the case $2<\alpha<3$, $\mathbf{E}(d)$ can be further simplified as

$$\mathbf{E}(d)=\frac{(\frac{\sqrt{a}}{l})^{\alpha-2}(\zeta^3-1)\zeta^{-\alpha}n^{(\frac{3-\alpha}{2})}}{(\zeta^{(3-\alpha)})(\zeta^{(3-\alpha)}-1)(1+\frac{(\zeta^2-1)\zeta^{-\alpha}}{1-\zeta^{(2-\alpha)}})}\,.$$

If we reduce the step size as much as possible ($\zeta\to 1$), $\lim_{\zeta\to 1}\mathbf{E}(d)$ can be written as

$$\mathbf{E}(d)=\begin{cases}\frac{\sqrt{a}(\alpha-2)}{l(\alpha-3)} & \alpha>3\\ \frac{\sqrt{a}(1+3\ln(l\sqrt{\frac{n}{a}}))}{3l} & \alpha=3\\ \frac{(\frac{\sqrt{a}}{l})^{\alpha-2}3(\alpha-2)n^{(\frac{3-\alpha}{2})}}{\alpha(3-\alpha)} & 2<\alpha<3\,.\end{cases} \tag{19}$$

$$\mathbf{E}(d) \text{ is } \begin{cases} \Theta(1) & \alpha > 3 \\ \Theta(\ln(n)) & \alpha = 3 \\ \Theta(n^{\frac{(3-\alpha)}{2}}) & 2 < \alpha < 3 \end{cases} \tag{20}$$

Since $\mathbf{E}(C_{P2P}) = \frac{\mathbf{E}(C_L)}{\mathbf{E}(d)}$, and by using Equation (19), capacity $\mathbf{E}(C_{P2P})$ is given by

$$\mathbf{E}(C_{P2P}) = \begin{cases} \frac{C_L l(\alpha-3)}{\sqrt{a}(\alpha-2)} & \alpha > 3 \\ \frac{C_L 3l}{\sqrt{a}(1+3\ln(l\sqrt{\frac{n}{a}}))} & \alpha = 3 \\ \frac{C_L \alpha(3-\alpha)}{(\frac{\sqrt{a}}{l})^{\alpha-2} 3(\alpha-2) n^{(\frac{3-\alpha}{2})}} & 2 < \alpha < 3\,. \end{cases} \tag{21}$$

If nodes are optimally placed in a symmetric arrangement $C_L \propto W$ [11], the order of point-to-point capacity (C_{P2P}) when the number of nodes goes to infinity ($n \to \infty$), is given by

$$C_{P2P} \text{ is } \begin{cases} \Theta(W) & \alpha > 3 \\ \Theta(\frac{W}{\ln(n)}) & \alpha = 3 \\ \Theta(\frac{W}{n^{\frac{(3-\alpha)}{2}}}) & 2 < \alpha < 3\,. \end{cases} \tag{22}$$

When comparing our findings to prior research in this area, the study that bears the greatest resemblance to ours is Azimdoost et al [5] that presented the following capacity outcome:

$$\begin{cases} \Theta(\frac{n-q-1}{n^2 r^{\alpha-1}(n)}) & 2 \le \alpha \le 3, q < \infty \\ \Theta(\frac{n-q-1}{n^2 r^2(n)}) & \alpha \ge 3, q < \infty\,. \end{cases}$$

By incorporating the function $r(n) = \sqrt{\frac{\ln(n)}{n}}$, as proposed in the work of [11], to guarantee network connectivity, and assuming a contact number of $q = 1$, the capacity ranges presented in [5] are

$$\begin{cases} \Theta(\frac{1}{n^{\frac{(3-\alpha)}{2}} \ln(n)^{\frac{(\alpha-1)}{2}}}) & 2 \le \alpha \le 3 \\ \Theta(\frac{1}{\ln(n)}) & \alpha > 3\,. \end{cases}$$

Since the scenario described in Azimdoost et al.'s work [5] may not necessarily be symmetric, it is important to consider this factor when making comparisons. Gupta and Kumar [11] showed that the achievable capacity bounds for symmetric node configurations can be at least $\sqrt{\ln(n)}$ times higher than the capacity bounds for random configurations. To ensure a fair comparison, we normalize their capacity bounds by multiplying them by the factor of $\sqrt{\ln(n)}$. Therefore, we have the following normalized capacity bounds as

$$\begin{cases} \Theta(\frac{1}{n^{\frac{(3-\alpha)}{2}} \ln(n)^{\frac{\alpha}{2}-1}}) & 2 \le \alpha \le 3 \\ \Theta(\frac{1}{\sqrt{\ln(n)}}) & \alpha > 3\,. \end{cases} \tag{23}$$

Consequently, our capacity bounds, presented in Equation (22), outperform the results presented in [5]. Specifically, for $2 < \alpha < 3$, our capacity bounds of $\ln(n)^{\frac{\alpha}{2}-1}$ are superior, and for $\alpha > 3$, our capacity band is $\sqrt{\ln(n)}$ times better than the bound in [5].

5 Power law exponent estimation

The ultimate objective of capacity estimation is to determine if large-scale distributed networking is feasible. Since our capacity ranges are influenced by power-law exponent (α), the answer varies accordingly. Hence, to address the question, it is essential to know the actual value of α, and the first step is the estimate α based on statistical data. Therefore, we propose a method to extract the power-law distribution parameter (α) from statistical data.

Again, we consider Section 4 and assume that the interaction probability is changed in the multiple levels where our steps are $1, \zeta^2, ..., \zeta^i$. Here, the product of the interaction probability and the number of nodes in each level could be defined as the expected number of contacts for the source node in that level ($\mathbf{E}(C_i)$) so that:

$$\mathbf{E}(C_i) = P_i N_i \,. \tag{24}$$

By substituting P_i and N_i from Equations (13) and (14) when ($i \geq 2$), we can write:

$$\begin{aligned}\mathbf{E}(C_i) &= \frac{\gamma}{a^{\frac{\alpha}{2}}}\zeta^{-(i-1)(\alpha)}(\zeta^2-1)\frac{a}{l^2}\zeta^{2(i-2)} \\ &= \frac{\gamma}{a^{\frac{\alpha}{2}}}\zeta^{-\alpha}(\zeta^2-1)\frac{a}{l^2}\zeta^{-(i-2)(\alpha)}\zeta^{2(i-2)} \\ &= \frac{\gamma}{a^{\frac{\alpha}{2}}}\zeta^{-\alpha}(\zeta^2-1)\frac{a}{l^2}\zeta^{(2-\alpha)(i-2)} \,.\end{aligned}$$

We can define the average distance of each level from the source node as $\bar{d}_i = \frac{d^i_{max}+d^i_{min}}{2} = \frac{\sqrt{a}\zeta^{i-1}+\sqrt{a}\zeta^i}{2} = \sqrt{a}\zeta\zeta^{i-2}\frac{1+\zeta}{2}$. Therefore,

$$\zeta^{i-2} = \frac{2\bar{d}_i}{\sqrt{a}\zeta(1+\zeta)} \,. \tag{25}$$

By substituting ζ^{i-2} from Equation (25), we can write $\mathbf{E}(C_i)$ as

$$\mathbf{E}(C_i) = \frac{\gamma}{a^{\frac{\alpha}{2}}}\zeta^{-\alpha}(\zeta^2-1)\frac{a}{l^2}\left(\frac{2\bar{d}_i}{\sqrt{a}\zeta(1+\zeta)}\right)^{(2-\alpha)} \,.$$

Based on Theorem 4.1, for any power-law distribution interaction probability when network size goes to infinity, we should have $\alpha > 2$. Hence, by using the value of $\frac{\gamma}{\sqrt{a}}$ from Equation (17) where $\alpha > 2$,

$$\mathbf{E}(C_i) = \frac{1}{(1+\frac{(\zeta^2-1)\zeta^{-\alpha}}{1-\zeta^{(2-\alpha)}})}\zeta^{-\alpha}(\zeta^2-1)\left(\frac{2\bar{d}_i}{\sqrt{a}\zeta(1+\zeta)}\right)^{(2-\alpha)} \,.$$

Therefore, if $\alpha > 2$ and $k \geq 2$, $\mathbf{E}(C_i)$ is

$$\mathbf{E}(C_i) = \frac{2^{2-\alpha}\sqrt{a}^{\alpha-2}(1-\zeta^{-2})}{(1+\frac{(\zeta^2-1)\zeta^{-\alpha}}{1-\zeta^{(2-\alpha)}})(1+\zeta)^{2-\alpha}}{\bar{d}_i}^{2-\alpha} \,. \tag{26}$$

If we reduce the step size as much as possible ($\zeta \to 1$), $\bar{d}_i \approx d_i$ and $\lim_{\zeta\to 1}\mathbf{E}(C_i)$ can be written as Equation (27):

$$\mathbf{E}(C_i) = \frac{2(\alpha-2)\sqrt{a}^{\alpha-2}}{\alpha}{d_i}^{2-\alpha} \,. \tag{27}$$

By taking the logarithm on both sides of Equation (27), Equation (28) is achieved as

$$\log(C(d)) = -(\alpha-2)\log(d) + C_0 \,. \tag{28}$$

Hence, if we define b as the slope of the $(C(d), d)$ curve in the log-log scale, then $\alpha = -b + 2$.

6 Feasibility of large-scale ad hoc networks

The capacity bounds in our study (Equation (22)) vary depending on the power-law exponent (α). For instance, in Equation (22), when $\alpha \to 2$, the point-to-point capacity (C_{P2P}) is $\Theta(\frac{1}{\sqrt{n}})$. On the other hand, when $\alpha > 3$, regardless of the network size, C_{P2P} remains constant at $\Theta(1)$. Hence, determining the actual or estimated value of α is crucial in assessing the scalability of distributed networking.

As there is a lack of sufficient research specifically focused on estimating α, we devised a formula to extract α from statistical data. Now, we can extract the power-law exponent (α) from various case studies and empirical studies on interaction probability over distance and delve into a discussion about the feasibility of large-scale distributed networking based on realistic data.

Latane [15] utilized a power-law distribution model to describe social connections as a function of distance. This study analyzed empirical data from hundreds of individuals across different geographic areas spanning 1000 miles. For all different analyses in Table 2, the logarithmic plot of the number of interactions against distance showed a slope of very close to -1. For these experiments, the value of α calculated using Equation (28) is almost 3.

Authors of [15], expressed the social impact theory that suggests a consistent natural relationship between interaction probability and distance that remains unaffected by technological advancements so that the number of interactions over distance in log-log scale has the slope line of -1. Furthermore, the social impact also expresses that this natural relationship is not affected by technological advancements over time. The result of a newer analysis conducted by Backstrom et al. (2010) [6] also supports social impact theory where they examined nearly 3 million Facebook users and 30 million friendship connections across thousands of miles. The results of their study also exhibited a power-law distribution, with a slope of -1.05. This indicates that the power-law parameter, α, has a value of 3.05. Moreover, based on data from the more recent study by Daraganova et al. (2012) [9] involving hundreds of individuals in Australia, the power-law distribution parameter could be calculated within the range of $[2.969, 2.996]$.

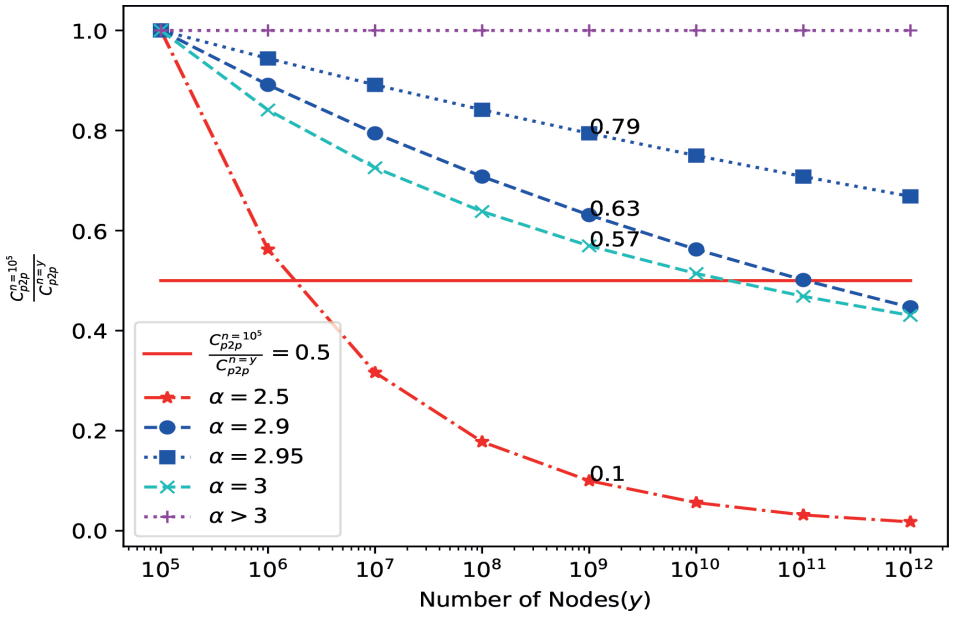
Table 2: Empirical works social tie.

Research Work	Case Study	Slope Line (log-log scale)	α (Equation (28))
Latane [15]	Memorable Interaction over distance (US)[fig1]	-1.01	3.01
	Memorable Interaction over distance (China)[fig2]	-1.05	3.05
	Memorable Interaction over distance (Social sociologists)[fig3]	-0.93	2.93
Backstrom [6]	Friendship probability over distance (Facebook)[fig4]	-1.05	3.05
Daraganova [9]	community network interaction probability in Australia[table 4(λ column)]	-0.969	2.969
		-0.995	2.995
		-0.972	2.972
		-0.985	2.985
		-0.995	2.995

These homogeneous results (Table 2) are justified by [15] in the social impact theory that is explained by Zipf's principle, which suggests that individuals aim to maximize social interaction while minimizing energy expenditure. However, the homo-phony of α values so that the range of α in all cases of empirical data from [6, 9, 15] is in $[2.9, 3.1]$ can be justified also from capacity analysis viewpoint. In fact, by considering the power-law interaction probability model for real-world communication, the analysis of expected hop count ($\mathbf{E}(d)$) in distributed networks (Section 4) can be approached similarly to calculating the path length for actual communication in the physical world. Therefore, point-to-point available capacity per user follows a similar trend with Equation (22), where we interpret W as a communication resource such as street width. In this situation, if $\alpha > 3$, regardless of network size, distributed networking is feasible, and if $\alpha \in [2.9, 3]$, C_{P2P} decreases slowly. In Figure 4, a base network size is considered ($n = 10^5$) and the normalized capacity is defined as the ratio of C_{P2P} of a given network size $n = y$ to $C_{P2P}^{n=10^5}$. As an example, as shown as Figure 4, the normalized capacity when network size is $1billion$ nodes, $(\frac{C_{P2P}^{n=10^9}}{C_{P2P}^{n=10^5}})$ reduces only by factor of $0.79, 0.63, 0.57$, for $\alpha = 2.95$, $\alpha = 2.9$, and $\alpha = 3$ respectively. While, if α deviates from these domain, for example if $\alpha = 2.5$, $(\frac{C_{P2P}^{n=10^9}}{C_{P2P}^{n=10^5}})$ decreases significantly when expanding network size. In other words, to achieve feasible large-scale

communication where the interaction probability follows a power-law distribution, the parameter (α) should not be significantly lower than 3 in the real world.

Figure 4: The normalized capacity size $\left(\frac{C_{P2P}^{n=y}}{C_{P2P}^{n=10^5}}\right)$ for different cases of α.



At first glance, it may seem surprising. However, if we compare two cities with populations of 10^5 and 10^7 where the social interaction pattern is $\propto \frac{1}{d^2}$ so that $\mathbf{E}(d)$ is $\Theta(\sqrt{n})$ the average street width in the larger city should be $\sqrt{\frac{10^7}{10^5}} = 10$ times wider than the smaller city to provide the same transportation capacity. Nevertheless, in reality, we know that the street widths in both cities are more or less the same. In other words, the value of α (social tie) is regulated by nature so that large-scale communication can be possible for any individual from a capacity viewpoint.

7 Limitations and future work

In this paper, we thoroughly examined the feasibility of large-scale distributed networking for a symmetric node arrangement. Our capacity bounds surpassed state-of-the-art results in the field. We further extracted the power law exponent from statistical data and found a consistent pattern, with ($\alpha \approx 3$) across various empirical analyses. The main conclusion is that very large-scale distributed networking could be feasible, as capacity is not a barrier to expansion. Notably, statistical data on actual traffic patterns for social interaction applications support this view, indicating promising potential for very large-scale distributed networking.

However, more extensive empirical analysis is needed to gain a deeper understanding of point-to-point traffic patterns and origin-destination communications. Additionally, exploring the strong correlation between statistical data and stochastic capacity analysis could be an interesting topic for future research. Furthermore, in practice, symmetric node distribution is atypical but simplifies network routing and capacity analysis, allowing us to probe network feasibility. Without such simplifications, the potential of distributed networks might be underestimated. Our study thus prioritizes multi-hop communication's impact on network expansion and hints at promising prospects for large-scale Ad Hoc networks, warranting deeper capacity analysis and empirical studies on social ties. However, our study represents an initial step toward addressing the feasibility issue. Our future work involves a numerical analysis of more comprehensive scenarios taken into account wireless layers issues, and evaluating

wireless network performance metrics such as average capacity per user, delay, power consumption, and spectral efficiency. Additionally, we plan to study the impact of random node arrangement on network capacity, reflecting more authentic conditions, and addressing core challenges like routing and resource allocation in these environments.

A Appendix A: Exclusively distance-dependent distribution

A.1 Exclusively distance-dependent distribution

If node i^{th} from the $n-1$ nodes wants to communicate with the source node with probability p_i, the summation of all nodes' interaction probability with the source node is 1 ($\sum_{i=1}^{n-1} p_i = 1$). As all nodes in a given distance from the source node have equal interaction probability, we can sum up the number of nodes times interaction probability at any distance ($\sum_{i=1}^{l\sqrt{n}} P(d)N(d) = 1$, so that $N(d)$ is the number of nodes between two circles of radiuses $l(d+.5)$ and $l(d-.5)$. When nodes are uniformly distributed with node density $\rho = \frac{1}{l^2}$, the number of nodes in a given distance ($N(d)$) from the source node has a linear relation with the distance, as shown in Figure 5. Hence,

$$N(d) = \rho 2\pi(l^2(d+0.5)^2 - (d-0.5)^2) = 2\pi\frac{l^2}{l^2}d = 2\pi d\,.$$

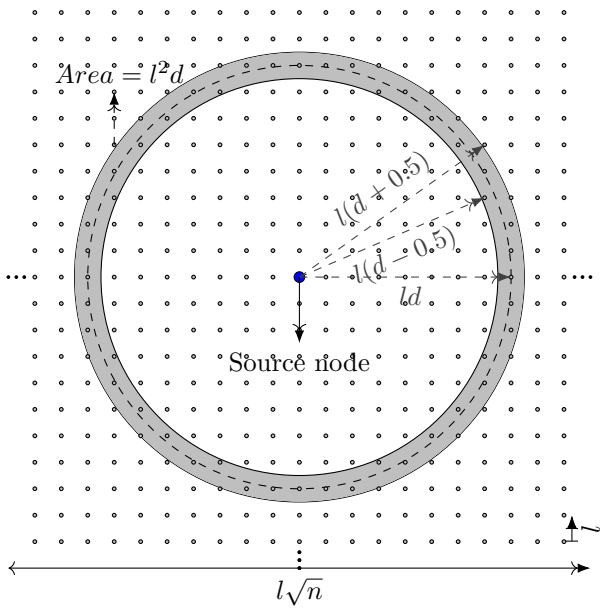
Furthermore, As the network diameter is $l\sqrt{n}$, the max distance is proportional to $l\sqrt{n}$. Therefore, to cover all nodes in the network, we should have:

$$\sum P(d)N(d) = 1 \mid d \in \{l, 2l, \ldots, l\sqrt{n}\}\,.$$

It means $\sum_{l}^{l\sqrt{n}} 2\pi P(d)d = 1$ where $l\sqrt{n}$ is the network diameter that is proportional to the average link length (l) and $\sqrt{n}$.

When $n \to \infty$, the necessary condition for $\sum_{l}^{l\sqrt{n}} 2\pi P(d)d = 1$ is that $\sum^{\infty} P(d)d$ converges.

Figure 5: Number of nodes in the distance d.



References

[1] Explainer: Why is it taking so long to re-establish power after the ice storm?, 2022. URL: https://montrealgazette.com/news/local-news/quebec-ice-storm-explainer-how-did-we-get-here-and-what-happens-next/.

[2] How to migrate a web2 application to web3?, 2023. URL: https://www.leewayhertz.com/how-to-migrate-a-web2-application-to-web3/.

[3] IEEE connecting unconnected, 2023. URL: https://ctu.ieee.org/.

[4] Giusi Alfano, Michele Garetto, Emilio Leonardi, and Valentina Martina. Capacity scaling of wireless networks with inhomogeneous node density: Lower bounds. IEEE/ACM Transactions on Networking, 18(5):1624–1636, 2010. doi:10.1109/TNET.2010.2048719.

[5] Bita Azimdoost, Hamid R Sadjadpour, and JJ Garcia-Luna-Aceves. Capacity of wireless networks with social behavior. IEEE Transactions on Wireless Communications, 12(1):60–69, 2012. doi:10.1109/GLOCOM.2016.7841651.

[6] Lars Backstrom, Eric Sun, and Cameron Marlow. Find me if you can: improving geographical prediction with social and spatial proximity. In Proceedings of the 19th international conference on World wide web, pages 61–70, 2010. doi:10.1145/1772690.1772698.

[7] Mohammed Banafaa, Ibraheem Shayea, Jafri Din, Marwan Hadri Azmi, Abdulaziz Alashbi, Yousef Ibrahim Daradkeh, and Abdulraqeb Alhammadi. 6g mobile communication technology: Requirements, targets, applications, challenges, advantages, and opportunities. Alexandria Engineering Journal, 2022.

[8] Zairan Cheng and Ying Liu. Performance study and optimization of 3d-manet: A new analytical perspective based on zipf's law. Wireless Communications and Mobile Computing, 2022, 2022. doi:10.1155/2022/9904257.

[9] Galina Daraganova, Pip Pattison, Johan Koskinen, Bill Mitchell, Anthea Bill, Martin Watts, and Scott Baum. Networks and geography: Modelling community network structures as the outcome of both spatial and network processes. Social networks, 34(1):6–17, 2012. doi:10.1016/j.socnet.2010.12.001.

[10] Luoyi Fu, Wentao Huang, Xiaoying Gan, Feng Yang, and Xinbing Wang. Capacity of wireless networks with social characteristics. IEEE Transactions on Wireless Communications, 15(2):1505–1516, 2015. doi:10.1109/TWC.2015.2491278.

[11] Piyush Gupta and Panganmala R Kumar. The capacity of wireless networks. IEEE Transactions on information theory, 46(2):388–404, 2000. doi:10.1109/18.825799.

[12] Xinxin he, Weisen Shi, and Tao Luo. Transmission capacity analysis for vehicular ad hoc networks. IEEE Access, 6:30333–30341, 2018. doi:10.1109/ACCESS.2018.2843333.

[13] Jiancao Hou and Mohammad Shikh-Bahaei. Transmission capacity of full-duplex mimo ad-hoc network with limited self-interference cancellation. In 2018 IEEE Global Communications Conference (GLOBECOM), pages 1–7. IEEE, 2018. doi:10.1109/GLOCOM.2018.8647531.

[14] Ronghui Hou, Yu Cheng, Jiandong Li, Min Sheng, and King-Shan Lui. Capacity analysis of hybrid wireless networks with long-range social contacts behavior. In 2015 IEEE Conference on Computer Communications (INFOCOM), pages 2209–2217. IEEE, 2015. doi:10.1109/INFOCOM.2015.7218607.

[15] Bibb Latané, James H Liu, Andrzej Nowak, Michael Bonevento, and Long Zheng. Distance matters: Physical space and social impact. Personality and Social Psychology Bulletin, 21(8):795–805, 1995. doi:10.1177/0146167295218002.

[16] Jinyang Li, Charles Blake, Douglas SJ De Couto, Hu Imm Lee, and Robert Morris. Capacity of ad hoc wireless networks. In Proceedings of the 7th annual international conference on Mobile computing and networking, pages 61–69, 2001. doi:10.1145/381677.381684.

[17] Dongrun Qin and Zhi Ding. Transport capacity analysis of wireless in-band full duplex ad hoc networks. IEEE Transactions on Communications, 65(3):1303–1318, 2016. doi:10.1109/TCOMM.2016.2640278.

[18] Mario L Small and Laura Adler. The role of space in the formation of social ties. Annual Review of Sociology, 45:111–132, 2019. doi:10.1146/annurev-soc-073018-022707.

[19] Yuhua Wang, Laixian Peng, Renhui Xu, and Shuai Cheng. The upper bound of capacity for multi-mode networking with multi-beam antenna array. In 2021 7th International Conference on Computer and Communications (ICCC), pages 2143–2148. IEEE, 2021. doi:10.1109/ICCC54389.2021.9674633.

[20] Zhiqing Wei, Huici Wu, Xin Yuan, Sai Huang, and Zhiyong Feng. Achievable capacity scaling laws of three-dimensional wireless social networks. IEEE Transactions on Vehicular Technology, 67(3):2671–2685, 2018. doi:10.1109/TVT.2017.2772797.

[21] Yang Zhou, Yan Shi, and Shanzhi Chen. Capacity and delay analysis for large social-aware mobile ad hoc wireless networks. Applied Sciences, 10(5):1719, 2020. doi:10.3390/app10051719.