Average Message Delivery Time for Small-world Networks in the Continuum Limit

Hazer Inaltekin, Mung Chiang and H. Vincent Poor
Department of Electrical Engineering, Princeton University, Princeton, NJ 08544
{hinaltek, chiangm, poor}@princeton.edu

***Abstract*—Small-world networks are networks in which the graphical diameter of the network is as small as the diameter of random graphs but whose nodes are highly clustered when compared with the ones in a random graph. Examples of small-world networks abound in sociology, biology, neuroscience and physics as well as in human-made networks such as the Internet, power grids and businesses. This paper analyzes the average delivery time of messages in dense small-world networks constructed on metric-measure spaces. Iterative equations for the average message delivery time in these networks are provided for the situation in which nodes employ a simple greedy geographic routing algorithm. It is shown that two network nodes communicate with each other only through their short-range contacts, and that the average message delivery time rises linearly if the separation between them is small. On the other hand, if their separation increases, the average message delivery time rapidly saturates to a constant value and stays almost the same for all large values of their separation.**

I. Introduction and Related Work

A. Background

Many networks arising in science [1], technology [2] and society [3] exhibit complex connections by means of small numbers of long-range links, and are surprisingly closely connected despite their extent. Such networks have become known as small-world networks. The most striking characteristic of small-world networks is their small graphical diameter even though they are highly clustered. One important field of science in which small-world network structures frequently arise is the theory of social networks through which the connections among people in a social network are studied [4]. Social networks will be the major focus of this paper; however, the results obtained for the average delivery time of messages are also valid in similar small-world network models appearing in other fields of science and technology.

Social ties among people are believed to form a small-world network. As a historical note, one of the key issues in the theory of social networks has been to understand how people in a social network are linked together, and the research efforts in this direction can be traced back to the famous six-degrees of separation principle, which was first speculated by the Hungarian writer Frigyes Karinthy in [5]. Since then, it has been believed that people are connected to one another through short chains of intermediate acquaintances, whose lengths are around six. Studies such as those described in [6], [7] and [8] are experimental verifications of the small-world phenomenon in social networks.

The first striking empirical verification of Karinthy's conjecture was the series of experiments ([6] and [7]) conducted by the social psychologist Stanley Milgram in the 1960s. The results of these experiments suggested that people are connected through surprisingly few ”degrees of separation”. This observation is in line with Karinthy's claim, which was made as early as 1929. A typical realization of the experiment was as follows. Two people located in the US are randomly chosen, usually one in Nebraska and the other one in Massachusetts. The person in Nebraska is assigned to be the originator of the message whose aim is to deliver the message to the target in Massachusetts. The originator is provided with some basic information about the target such as address and occupation, and he is only allowed to send the message to others whom he knows on a first name basis. Intermediate message holders repeat the same step until it reaches the target in Massachusetts. In his work [6], Milgram reported that the lengths of the chains of the successfully delivered messages ranged between two and ten with the median value at five. Milgram's experiment forms the first experimental basis for the famous six-degrees of separation principle in social networks.

The first important conclusion of Milgram's experiment is that the length of the shortest path, regardless of the number of nodes, connecting any two nodes must be small in any graphical model aiming to characterize the structure of connections among the nodes of a social network. For this reason, random graphs [9], at first glance, may be thought to be good candidates for modeling the connections in a social network. However, they fail to capture a very important clustering property of social networks: friends of a person are also likely to be friends with one another. To capture the clustering feature of a social network as well as to maintain the small shortest-path property enjoyed by random graphs, Watts and Strogatz proposed an elegant small-world network model in [10]. In their model, they consider a regular network in which nodes are evenly spaced on a circle. They first construct the links forming the local contacts of a node, i.e., each node on the circle is connected to its k nearest neighbors. They then rewire each one of these links to any other randomly selected node in the network with some very small probability, which forms the long-range contacts of a node. They show that even if the number of long-range links is very small, there is a dramatic decrease in the diameter of the network. Therefore, they prove that it is possible to obtain small-world network

This research was supported by the National Science Foundation under Grants ANI-03-38807 and CNS-06-25637.

models without changing the clustering property manifested by social networks. Their work has attracted considerable attention, and has formed the mathematical basis of many small-world network models.

B. Related Work

The second conclusion, which is as important as the first one, of Milgram's experiment is that people operating based only on their local information are very efficient in finding the short paths connecting them. This second point forms the *algorithmic* perspectives of small-world networks, which were first investigated by Kleinberg in [12] and [13]. In his work, Kleinberg studied a more general version of the Watts-Strogatz model by considering a set of nodes located on a grid. Each node maintains some finite number of short-range and long-range contacts. He obtained bounds on the average delivery time of a message when each message holder, based on its local information, uses a simple greedy forwarding rule to deliver the message to the target node. In [14], the authors consider a variant of Kleinberg's model which is related to continuum percolation. In their model, the density of the number of connections of a node decreases with distance but there is no uniform upper bound on the number of contacts that a node may have. They obtain bounds on the average delivery time of a message which are similar to those obtained in [13]. One artifact of their model is that for any two randomly selected nodes, regardless of how close they are to each other, the probability of sharing a common neighbor is zero, which conflicts with the clustering property enjoyed by social networks.

In this work, we also focus on the algorithmic perspectives of small-world networks, and analyze the average message delivery time when nodes run a greedy geographic forwarding algorithm to deliver a message to the final destination. Unlike [12] and [13], we do not restrict ourselves to regular grid networks. We consider general metric-measure spaces as the network domain, and nodes can be located at any point inside the network domain. In our model, we also maintain the clustering property of social networks by assuming that any two nodes are said to be local contacts of each other if the separation between them is smaller than some positive real number $r > 0$. Therefore, in contrast to [14], network nodes are very likely to share common friends if they are close to one another. This network model of local neighbors is motivated by the models of wireless packet radio networks (see the protocol model proposed in [15]). Long-range contacts will be formed randomly uniformly over the network domain. Finally, we obtain an exact recursive expression for the average message delivery time in a small-world network rather than providing bounds as in [13] and [14]. We also provide a solution for this recursive equation. Our results verify the empirical observations of the experiments conducted in [6] and [7].

II. The Notion of Network Space: Its Definition and Some Fundamental Properties

We consider a general metric-measure space as the network domain on which network nodes are distributed. Before putting forward the formal definition of a network space, we first review some basic concepts from real analysis.

We call the tuple (D, ρ) a *metric space* with *metric* ρ if D is a set, and for any two points x and y of D, there corresponds a real number $\rho(x, y)$ such that

- $\rho(x, y) \geq 0$, and $\rho(x, y) = 0$ if and only if $x = y$,
- $\rho(x, y) = \rho(y, x)$ (symmetry), and
- $\rho(x, y) \leq \rho(x, z) + \rho(z, y)$ (the triangle inequality).

For any $x \in D$ and $r > 0$, open and closed balls around x with radius r are, respectively, equal to

$$B(x, r) = \{y \in D : \rho(x, y) < r\}$$

and

$$\overline{B(x, r)} = \{y \in D : \rho(x, y) \leq r\}.$$

Consider a subset U of D. It is said to be an open set if it is an arbitrary union of open balls. It is said to be a closed set if its complement is an open set. Its interior $\text{int}(U)$ is the biggest open set contained in U. Its closure $\text{cl}(U)$ is the smallest closed set containing U. Its boundary ∂U is equal to the difference between its closure and interior (i.e., $\partial U = \text{cl}(U) - \text{int}(U)$). For any given metric space (D, ρ), the topology $\mathcal{V}_D$ on D generated by the metric ρ is the collection of all open subsets of D. In this paper, we always refer to the metric generated topology when we speak of a topology on a metric space.

For a given metric space (D, ρ), we say that a subset D_0 of D is a *dense* subset of D if every non-empty open subset of D consists of infinitely many points of D_0. We say a metric space (D, ρ) is *separable* if D has a countable dense subset D_0.

We formally define the notion of network space as follows.

Definition 1: The quadruple $(D, \mathcal{S}_D, \lambda, \rho)$ is called a *topological network space*, where D is the set of points on which network nodes are distributed, $\mathcal{S}_D$ is the Borel σ-algebra of subsets of D, $\lambda : \mathcal{S}_D \to [0, \infty]$ is a measure on $\mathcal{S}_D$, and $\rho : D \times D \to [0, \infty)$ is a metric.

For the rest of the paper, the quadruple $(D, \mathcal{S}_D, \lambda, \rho)$ will be referred to simply as a network space by bearing in mind that the underlying topology on D in Definition 1 is the usual topology generated by the metric ρ. Therefore, $\mathcal{S}_D$ is the smallest σ-algebra that contains all the open subsets of D.

Definition 2: A network space $(D, \mathcal{S}_D, \lambda, \rho)$ is said to be *finite* if $\lambda(D) < \infty$. It is said to be *translation invariant* if for all $x, y \in D$ and $r > 0$, $\lambda(B(x, r)) = \lambda(B(y, r))$. And, it is said to be a *positive network space* if for all $x, y \in D$ and $r_1, r_2 \geq 0$ satisfying $r_1 + r_2 > \rho(x, y)$, $\lambda\left(B(x, r_1) \bigcap B(y, r_2)\right) > 0$.

A. Some Properties of Network Spaces

Property 1: Every positive network space contains infinitely many points. Moreover, any given open ball with positive radius also contains infinitely many points.

Proof: Consider any positive network space $(D, \mathcal{S}_D, \lambda, \rho)$. Suppose there exist only finitely many points in D. Take $x \in D$, and let

$$\epsilon_x = \inf\{\rho(x, y) : y \in D, y \neq x\}.$$

Since there are only finitely many points in D, $\epsilon_x > 0$. Now, take another point $y \in D$, which is not equal to x. Let $d = \rho(x, y)$. Set $r_1 = \frac{2\epsilon_x}{3}$ and $r_2 = d - \frac{\epsilon_x}{3}$. Then,

$$\begin{aligned}\lambda\left(B(x, r_1) \bigcap B(y, r_2)\right) &= \lambda\left(\{x\} \bigcap B(y, r_2)\right) \\ &= \lambda(\emptyset) = 0, \text{ a contradiction.}\end{aligned}$$

Similarly, if there exist $x \in D$ and $r > 0$ such that $B(x, r)$ contains only finitely many points, we can find $\epsilon_x \in (0, r)$ such that $B(x, \epsilon_x) = \{x\}$. Then, by choosing r_1 and r_2 as above for any given $y \neq x$, one can arrive at a contradiction. ■

Now, consider a topological measure space $(D, \mathcal{S}_D, \lambda)$, where the topology on D is any topology and $\mathcal{S}_D$ is the Borel σ-algebra containing all open subsets of D. Then, λ is said to be *everywhere positive* if $\lambda(U) > 0$ for every non-empty open set U.

Property 2: Let $(D, \mathcal{S}_D, \lambda, \rho)$ be a positive network space. Then, λ is everywhere positive.

Proof: Assume there exists a non-empty open subset U of D with $\lambda(U) = 0$. Then, there exists $x \in U$ and $\epsilon > 0$ such that $B(x, \epsilon) \subseteq U$. Therefore, $\lambda(B(x, \epsilon)) = 0$. Take $y \in \Omega$, and let $d = \rho(x, y)$. Let $r_1 = \frac{2\epsilon}{3}$ and $r_2 = d - \frac{\epsilon}{3}$. Then,

$$\lambda\left(B(x, r_1) \bigcap B(y, r_2)\right) \leq \lambda(B(x, r_1)) = 0,$$

which contradicts the positivity of the network space. ■

Property 3: Let $(D, \mathcal{S}_D, \lambda, \rho)$ be a positive network space. Then, $\mathrm{cl}(B(x, r)) = \overline{B(x, r)}$ for all $x \in D$ and $r > 0$.

Proof: $\overline{B(x, r)}$ is a closed set containing $B(x, r)$. Thus, $\mathrm{cl}(B(x, r)) \subseteq \overline{B(x, r)}$. Now assume there exists $y \in \overline{B(x, r)}$ that does not belong to $\mathrm{cl}(B(x, r))$. This is possible only if $\rho(x, y) = r$, and there exists $\epsilon > 0$ such that $B(x, r) \bigcap B(y, \epsilon) = \emptyset$. Thus, $\lambda(B(x, r) \bigcap B(y, \epsilon)) = 0$, which contradicts the positivity of the network space. ■

Consider a measurable space $(D, \mathcal{S}_D)$ such that all singleton sets are measurable.[1] A measure λ on this space is said to be *non-atomic* if $\lambda(\{x\}) = 0$ for all $x \in D$.

Property 4: Let $(D, \mathcal{S}_D, \lambda, \rho)$ be a positive, translation invariant and finite network space. Then, λ is a non-atomic measure.

Proof: Suppose there exists $x \in D$ such that $\lambda_x = \lambda(\{x\}) > 0$. Take a decreasing sequence $\{r_n\}_{n=1}^{\infty}$ such that $r_n \downarrow 0$. Then, for any given $y \in D$, we have

$$\begin{aligned}\lambda_x &= \lim_{n\to\infty} \lambda(B(x, r_n)) \\ &= \lim_{n\to\infty} \lambda(B(y, r_n)) \\ &= \lambda(\{y\}),\end{aligned}$$

where the first and third equalities follow from the continuity of measure from above [16], and the second one follows from the translation invariance property of λ. Therefore, λ assigns the same positive value to all points of D. Since there are infinitely many points in D by Property 1, $\lambda(D) = \infty$. This contradicts the finiteness of the network space. ■

[1]A measurable space is a pair $(D, \mathcal{S}_D)$ in which D is any set, and $\mathcal{S}_D$ is a σ-algebra of subsets of D.

One easy corollary of Property 4 is that any positive, translation invariant and finite network space contains uncountably many points.

Corollary 1: Let $(D, \mathcal{S}_D, \lambda, \rho)$ be a positive, translation invariant and finite network space. If λ is a non-trivial measure (i.e., $\lambda(D) > 0$), then D consists of uncountably many points.

Definition 3: A network space $(D, \mathcal{S}_D, \lambda, \rho)$ is called a *regular* network space if it is a separable, positive, translation invariant and finite network space.

Further properties of regular network spaces such as the measure of boundary sets in these spaces and their relationship to convex metric spaces are explored in the appendix. These properties will not be needed in our proofs below, but are included for the sake of the curiosity of interested readers.

III. Network Model

To investigate the average message delivery time in small-world networks, we consider regular network spaces. Let $(D, \mathcal{S}_D, \lambda, \rho)$ be a regular network space. We assume that there exists an underlying probability space $(\Omega, \mathcal{S}_\Omega, \mathbb{P})$ on which all the random variables of interest are defined. We define node location random variables as follows.

Definition 4: A node location random variable is a measurable function

$$X : (\Omega, \mathcal{S}_\Omega, \mathbb{P}) \to (D, \mathcal{S}_D, \lambda, \rho). \tag{1}$$

We say X is uniformly distributed if $\mathbb{P}\{X \in U\} = \frac{\lambda(U)}{\lambda(D)}$ for all $U \in \mathcal{S}_D$.

We distribute n *relay* nodes randomly (from a uniform distribution) and independently over the network domain. Locations of *source* s and *target* t nodes will be assumed to be arbitrary so that we may analyze the average delivery time of a message as a function of their separation. Node location random variables are represented as follows:

$$\begin{aligned}X_s &\triangleq \text{ location of the source node,} \\ X_t &\triangleq \text{ location of the target node, and} \\ X_1, X_2, \ldots, X_n &\triangleq \text{ locations of } n \text{ relay nodes.}\end{aligned}$$

For each $n \geq 1$, we define $\mathcal{H}_n = \{X_i\}_{i=1}^n$. The random set $\mathcal{H}_n$ will be called the *node-location process*. We define the continuum limit as $\mathcal{H}_\infty = \bigcup_{n=1}^{\infty} \mathcal{H}_n$.

For a given network space $(D, \mathcal{S}_D, \lambda, \rho)$, we say that a subset D_0 of D is a dense subset of the network domain if every non-empty open subset of D consists of infinitely many points of D_0. An important topological property regarding $\mathcal{H}_\infty$ is the following.

Theorem 1: For any given regular network space $(D, \mathcal{S}_D, \lambda, \rho)$, $\mathcal{H}_\infty$ is a dense subset of the network domain with probability one (w.p.1).

Proof: Let D_0 be a countable dense subset of D. Choose any $x \in D_0$. For any small ball $B\left(x, \frac{1}{p}\right) \subseteq D$ around x with radius $\frac{1}{p}$ and p being an integer greater than or equal to 1, let $Z_n^{x,p} = \sum_{i=1}^n \xi_i$, and $\xi_i = \mathbb{1}_{\left\{X_i \in B\left(x, \frac{1}{p}\right)\right\}}$. Since, under our model, the ξ_i's are independent and identically distributed (i.i.d) random variables with finite means $\mu > 0$, we have $\frac{Z_n^{x,p}}{n} \to \mu$ w.p.1. Observe that

$$Z_\infty^{x,p} = \left|\mathcal{H}_\infty \bigcap B\left(x, \frac{1}{p}\right)\right| \geq \left|\mathcal{H}_n \bigcap B\left(x, \frac{1}{p}\right)\right| = Z_n^{x,p}$$

for all $n \geq 1$. Thus, $Z_\infty^{x,p} \geq \lim_{n\to\infty} Z_n^{x,p} = \infty$. As a result, there will be infinitely many points of $\mathcal{H}_\infty$ lying in $B\left(x, \frac{1}{p}\right)$ w.p.1. Let $\Omega_{x,p} \subseteq \Omega$ be the set on which $Z_\infty^{x,p} = \infty$. We define Ω_0 as

$$\Omega_0 = \bigcap_{x \in D_0} \bigcap_{p=1}^{\infty} \Omega_{x,p}.$$

$\mathbb{P}\{\Omega_0\} = 1$ since it is an intersection of countably many sets having probability measure one. Take an $\omega \in \Omega_0$. Now, for any $x \in D$ and $\epsilon > 0$, we can find $z \in D_0$ and $p \geq 1$ such that $B\left(z, \frac{1}{p}\right) \subseteq B(x, \epsilon)$. Therefore, $|\mathcal{H}_\infty(\omega) \bigcap B(x,\epsilon)| = \infty$ for any $x \in D$ and $\epsilon > 0$. Now, take any non-empty open subset U of D. Then, there exists $x \in U$ and $\epsilon > 0$ such that $B(x,\epsilon) \subseteq U$. Therefore, $|\mathcal{H}_\infty(\omega) \bigcap U| = \infty$. This completes the proof. ■

A. *Connectivity Properties of Nodes*

We assume that each node in the network maintains both local and long-range contacts as in other existing small-world network models. We first describe how the set of local contacts of a node is formed.

(i) Local Contacts of a Node: All nodes in the network have a uniform communication range r. Nodes lying inside this communication range of a node form its local contacts. Let us consider a particular node i with location X_i. Then, its *local-neighborhood* will be the ball

$$B(X_i, r) = \{x \in D : \rho(x, X_i) < r\}. \tag{2}$$

The set of local neighbors of node i is, on the other hand, equal to

$$Loc(X_i) = \{1 \leq j \leq n : X_j \in B(X_i, r)\}. \tag{3}$$

(ii) Long-range Contacts of a Node: In addition to its local contacts, each node maintains some number of long-range links for communication. We assume that there are two types of long-range links: **(a)** *incoming long-range links* and **(b)** *outgoing long range links*. A node can receive a message through an incoming long-range link but cannot send any messages over this link. On the other hand, a node can send a message over an outgoing long-range link but cannot receive anything. For the rest of the paper, it is assumed that each node has L, $L \geq 1$, outgoing long-range links. This model for the long-range contacts of a node is made merely for the sake of analytical simplicity. Our solution technique presented below can also be extended to other models for forming long-range contacts.

B. *Limiting Process for Forming the Outgoing Long-range Contacts*

We now put forward the rules for the limiting process for forming outgoing long-range contacts as the number of relay nodes goes to infinity.

1) *Rules for forming outgoing long-range contacts:* Source and target nodes are only allowed to choose their outgoing long-range contacts among relay nodes. Relay nodes are only allowed to choose their outgoing long-range contacts among relay nodes. A node chooses the node at the receiving end of an outgoing long-range link at random uniformly among all other nodes which do not lie in its local neighborhood. After a node chooses a long-range contact, it is not allowed to change it as new relay nodes are added to the network.
2) *Initialization and the limiting process:* We initialize the number of relay nodes to N_{init} where N_{init} is an arbitrary integer greater than 1, and uniformly distribute N_{init} relay nodes over D. We place source and target nodes at arbitrary positions in D. Nodes then choose their outgoing long-range contacts as explained above. We send the number of relay nodes to infinity by adding new relay nodes. As a new relay node is added, nodes which have not been able to pick $L \geq 1$ different outgoing long-range contacts yet are given a chance to form a long-range outgoing contact with this new relay node. Nodes which have already chosen their L outgoing long-range contacts are not allowed to change them.

Let $E_{n,i}$ be the event that node $i \in \{s,t\} \bigcup \{1, 2, \cdots, n\}$ has L outgoing long-range contacts when there are n relay nodes located in D. On this event, node i's outgoing long-range contacts will be uniformly distributed over $D - B(X_i, r)$. Therefore, the probability that node i has an outgoing long-range contact j ($i \xrightarrow{\text{LRC}} j$) in $A \subseteq D$ conditioned on X_i and $E_{n,i}$ can be written as

$$\mathbb{P}\{\exists j \text{ in } A \bigcap \mathcal{H}_n \text{ s.t } i \xrightarrow{\text{LRC}} j | X_i, E_{n,i}\} = 1 - \left(1 - \frac{\lambda\left(A - B(X_i, r)\right)}{\lambda\left(D - B(X_i, r)\right)}\right)^L.$$

We have the following theorem when there are infinitely many relay nodes located in D.

Theorem 2: For a given node $i \in \{s,t\} \bigcup \{1, 2, 3, \cdots\}$ and a given subset A of D,

$$\mathbb{P}\{\exists j \text{ in } A \bigcap \mathcal{H}_\infty \text{ s.t } i \xrightarrow{\text{LRC}} j | X_i\} = 1 - \left(1 - \frac{\lambda\left(A - B(X_i, r)\right)}{\lambda\left(D - B(X_i, r)\right)}\right)^L.$$

Proof: Let $E_{n,i}$ be defined as above, and

$$G_{n,i} = \{\exists j \text{ in } A \bigcap \mathcal{H}_n \text{ s.t } i \xrightarrow{\text{LRC}} j\}.$$

$\{E_{n,i}\}_{n \geq N_{\text{init}}}$ is an increasing sequence of events since nodes are not allowed to change their outgoing long-range contacts once they choose them. Therefore, whenever a node has L outgoing long-range contacts for some $n \geq N_{\text{init}}$, it also

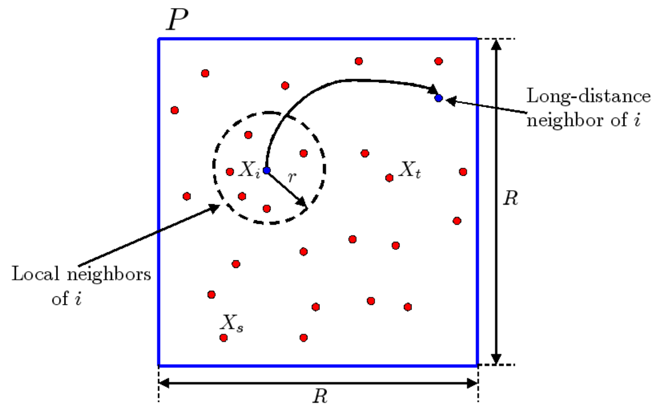

Fig. 1. A typical realization of the network.

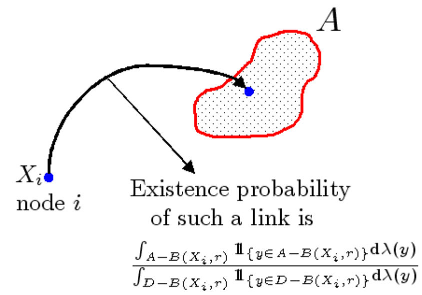

Fig. 2. An illustration of the probability with which a node chooses its outgoing long-range contacts when $L = 1$.

has L outgoing long-range contacts for all $k \geq n$. Similarly, $\{G_{n,i}\}_{n\geq N_{init}}$ is an increasing sequence of events. Let

$$E_{\infty,i} = \bigcup_{n\geq N_{\text{init}}} E_{n,i}$$

and

$$G_{\infty,i} = \bigcup_{n\geq N_{\text{init}}} G_{n,i}.$$

Since $\mathcal{H}_\infty$ is a dense subset of D, we have

$$\mathbb{P}\left(E_{\infty,i}\right) = \mathbb{P}\left(E_{\infty,i}|X_i\right) = \lim_{n\to\infty} \mathbb{P}\left(E_{n,i}|X_i\right) = 1.$$

Observe also that $G_{\infty,i} = \{\exists j \text{ in } A \bigcap \mathcal{H}_\infty \text{ s.t } i \stackrel{\text{LRC}}{\longrightarrow} j\}$. Then, by using the continuity of measure from below [16] and Bayes theorem, we have

$$\begin{aligned}
&\mathbb{P}\left(G_{\infty,i}|X_i\right)\\
&\quad= \mathbb{P}\left(G_{\infty,i}\bigcap E_{\infty,i}|X_i\right)\\
&\quad= \lim_{n\to\infty} \mathbb{P}\left(G_{n,i}\bigcap E_{n,i}|X_i\right)\\
&\quad= \lim_{n\to\infty} \mathbb{P}(G_{n,i}|X_i, E_{n,i})\mathbb{P}\left(E_{n,i}|X_i\right)\\
&\quad= \lim_{n\to\infty} \left(1-\left(1-\frac{\lambda\left(A-B(X_i,r)\right)}{\lambda\left(D-B(X_i,r)\right)}\right)^L\right)\cdot\mathbb{P}(E_{n,i}|X_i)\\
&\quad= 1-\left(1-\frac{\lambda\left(A-B(X_i,r)\right)}{\lambda\left(D-B(X_i,r)\right)}\right)^L.
\end{aligned}$$

■

A typical realization of the network, and the probability with which nodes form their outgoing long-range contacts when $L = 1$ are depicted in Fig. 1 and Fig. 2, respectively.

C. δ-Greedy Geographic Forwarding Rule

Nodes use a simple δ-greedy geographic forwarding (δ-GGF) rule to deliver their messages to their intended target nodes. The only global information needed to deliver a message is the location of its final destination, which can be encoded inside the message by the message originator.

According to this rule, any message holder having distance $d \in [kr, k+1))$ to X_t for some $k \geq 1$ chooses one of its local contacts lying in $B(X_t, kr)$ and providing a forward progress in the direction of the target node at least $r - \delta$ for some $\delta > 0$ as a next hop local contact. After a node chooses its next hop local contact, it is not allowed to change it as new relay nodes are added to the network even if there are some other local contacts providing forward progress larger than the chosen one. This rule can be interpreted as that nodes have *bounded greediness* while choosing their next hop local contacts. Fixing the local contact achieving a minimum desired forward progress in the direction of the target node makes δ-GGF a *path convergent* forwarding rule which will be analyzed in greater detail in Section IV.

If the current message holder has an outgoing long-range contact providing forward progress in the direction of the target node larger than the progress provided by the local contact chosen as described above, this long-range contact is preferred over the local contact.

IV. Message Trajectory and Average Message Delivery Time

A. Message Trajectory

In accordance with the search algorithm, a message wanders around the network hop by hop until it reaches the target. When there are n relay nodes, we denote the resulting random trajectory of the message by

$$M_m^{(n)} \triangleq \text{ location of the message at the } m\text{th hop.}$$

$M_0^{(n)}$ is always set to X_s. We define the message delivery time τ_n under $\mathcal{H}_n$ to be equal to the first time the message reaches the target node. Therefore, τ_n can be written as

$$\tau_n \triangleq \inf\{m \geq 0 : M_m^{(n)} = X_t\}.$$

Note that $\tau_n = \infty$ if the forwarding rule cannot find a path from the source node to the target node. We assume that the message stays at the target node once it reaches it. Therefore, $M_m^{(n)} = X_t$ for all $m \geq \tau_n$.

τ_n also depends on the locations of source and target nodes since the message starts its journey at X_s, and finishes it at X_t. For example, $\tau_n = 0$ for all $n \geq m$ if $X_s = X_t$ since there is no need for message forwarding when source and target nodes are located at the same position. If $X_s \neq X_t$, a message should be forwarded at least once in order to be delivered to the target node. Thus, $\tau_n \geq 1$ when $X_s \neq X_t$. For the rest of the paper, we will not explicitly write τ_n as a function of X_s and X_t by keeping this fact in mind.

Definition 5: We say a forwarding rule is *loop-free* if it does not form any loop while delivering the message to the target node.

Definition 6: We say a forwarding rule is *time convergent* if for any given $X_s \in D$ and $X_t \in D$, τ_n converges to a limiting real-valued random variable w.p.1 as n goes to infinity under the rules in III-B for forming long-range contacts.

For example, the shortest path routing is a time convergent forwarding rule since the message delivery time monotonically decreases as new relay nodes are added to the network domain. For time convergent forwarding rules, we let

$$\lim_{n\to\infty} \tau_n = \tau$$

be the message delivery time under $\mathcal{H}_\infty$.

Definition 7: We say a forwarding rule is *path convergent* if for any given $X_s \in D$ and $X_t \in D$, the path $\{M_m^{(n)}\}_{m=0}^{\tau_n}$ that the message takes to travel from X_s to X_t converges to a limiting path as n goes to infinity under the rules in III-B for forming long-range contacts.

For path convergent forwarding rules, we let

$$\lim_{n\to\infty} M_m^{(n)} = M_m$$

be the location of the message at the mth, $m \geq 0$, hop under $\mathcal{H}_\infty$. Therefore, $\{M_m\}_{m\geq 0}$ is the trajectory of the message under $\mathcal{H}_\infty$. We have the following theorem establishing the relationship between time convergent and path convergent forwarding rules.

Theorem 3: Consider a path convergent forwarding rule F. If, for any given $X_s \in D$ and $X_t \in D$, the message delivery time τ_n under F satisfies

$$\liminf_{n\to\infty} \tau_n = \tau < \infty \text{ w.p.1}$$

and F delivers the message to the target node at the next hop whenever it enters $B(X_t, r)$, then F is a time convergent forwarding rule with the limiting message delivery time τ.

Proof: This is trivial if $X_s = X_t$ since there is no need for message forwarding. Therefore, take any X_s and X_t with $X_s \neq X_t$ in D, and take any $\omega \in \{\tau < \infty\}$. Set

$$m_\star = \liminf_{n\to\infty} \tau_n(\omega).$$

Consider the subsequence $\{\tau_{n_1}(\omega)\}_{n_1 \geq N_{\text{init}}}$ of $\{\tau_n(\omega)\}_{n=N_{\text{init}}}^\infty$ such that

$$\lim_{n_1\to\infty} \tau_{n_1}(\omega) = m_\star.$$

Observe that there exists a constant N large enough such that $\tau_{n_1}(\omega) = m_\star$ for all $n_1 \geq N$ since $\tau_{n_1}(\omega)$ is a finite integer for all n_1 large enough. Therefore, $M_{m_\star}^{(n_1)}(\omega) = X_t$ for all $n_1 \geq N$.

Since the target node cannot have any incoming long-range links according to the rules explained in III-B for forming long-range contacts, the message must first enter the ball $B(X_t, r)$ before the final delivery of the message to the target node. Therefore, $M_{m_\star-1}^{(n_1)}(\omega)$ converges to $M_{m_\star-1}(\omega) \in B(X_t, r)$ as n_1 goes to infinity. Since F is a path convergent rule, we also have $M_{m_\star-1}^{(n)}(\omega)$ converges to $M_{k_\star-1}(\omega)$ as n goes to infinity. Since $B(X_t, r)$ is an open set, we must have $M_{m_\star-1}^{(n)}(\omega) \in B(X_t, r)$ for all n large enough. Since F delivers the message to the target node at the next hop once it enters $B(X_t, r)$, $\tau_n(\omega) = m_\star = \tau(\omega)$ for all n large enough. Therefore, τ_n converges to τ on the event $\{\tau < \infty\}$, which has probability one. ■

Let $\mathcal{F}$ denote the collection of all *time convergent* and *loop-free* forwarding rules. We have the following theorem establishing the relationship between $\mathbb{E}[\tau_n]$ and $\mathbb{E}[\tau]$ for this type of rules.

Theorem 4: Fix $X_s \in D$ and $X_t \in D$. Then, for any forwarding rule in $\mathcal{F}$ with $\mathbb{P}\{\tau_n > B\} = o(n^{-1})$ for some finite $B > 0$,

$$\lim_{n\to\infty} \mathbb{E}[\tau_n \mathbb{1}_{\{\tau_n < \infty\}}] = \mathbb{E}[\tau]. \tag{4}$$

Proof: Since the forwarding rule is loop-free, we have

$$\tau_n \mathbb{1}_{\{\tau_n < \infty\}} = \tau_n \mathbb{1}_{\{\tau_n \leq B\}} + \tau_n \mathbb{1}_{\{B < \tau_n \leq n+1\}}.$$

Since τ_n converges to τ w.p.1, we have $\mathbb{P}\{\tau > B\} = 0$. Therefore, by using bounded convergence theorem, we have

$$\lim_{n\to\infty} \mathbb{E}[\tau_n \mathbb{1}_{\{\tau_n \leq B\}}] = \mathbb{E}[\tau].$$

By using the condition on the decay rate of the tail of the distribution of τ_n, we have

$$\begin{aligned} 0 \leq \lim_{n\to\infty} \mathbb{E}[\tau_n \mathbb{1}_{\{B < \tau_n \leq n+1\}}] &\leq \lim_{n\to\infty} (n+1)\mathbb{P}\{\tau_n > B\} \\ &= 0. \end{aligned}$$

As a result, $\lim_{n\to\infty} \mathbb{E}[\tau_n \mathbb{1}_{\{\tau_n < \infty\}}] = \mathbb{E}[\tau]$. ■

We now show that Theorem 4 holds for δ-GGF. To this end, we need to show a series of lemmas.

Lemma 1: For any regular network space $(D, \mathcal{S}_D, \lambda, \rho)$, δ-GGF is both a path convergent and a time convergent forwarding rule for any $\delta > 0$.

Proof: Let $d = \rho(X_s, X_t)$ be the distance between source and target nodes. It is easy to see that the lemma holds for $d \in [0, r)$. Therefore, let us focus on $d \in [kr, (k+1)r)$ for $k \geq 1$. Since $(D, \mathcal{S}_D, \lambda, \rho)$ is a regular network space, $\mathcal{H}_\infty$ is a dense subset of D w.p.1. Let $\Omega_0 \subseteq \Omega$ be the set on which $\mathcal{H}_\infty$ is a dense subset of D, and take any $\omega \in \Omega_0$. Source node will able to pick L of its long-range contacts for all n large enough due to $\mathcal{H}_\infty(\omega)$ being dense in D. Let $\{i_j(s,\omega)\}_{j=1}^L$ be the indexes of these long-range contacts.

Now, assume $d - r + \delta \leq kr$. Since $\mathcal{H}_\infty(\omega)$ is dense subset of D, $B(X_t, d-r+\delta) \bigcap B(X_s, r)$ will eventually contain a relay node. Since such a relay node achieves a forward progress at least $r - \delta$ in the direction of the target node and lies in $B(X_t, kr)$, the source node will be able to pick a local contact when n is large enough. Similarly for $d - r + \delta > kr$, the source node will eventually be able to pick a local contact lying in $B(X_t, kr) \bigcap B(X_s, r)$ and achieving a forward progress at least $r - \delta$ in the direction of the target node. Let $i_{L+1}(s,\omega)$ be the index of the local contact chosen by the source node. After determining all of the possible next hop contacts, the source node chooses the one achieving the largest forward progress among $\{i_j(s,\omega)\}_{j=1}^{L+1}$. Let

$$i_1^\star(\omega) = \arg \min_{i \in \{i_j(s,\omega)\}_{j=1}^{L+1}} \rho(X_i(\omega), X_t).$$

Hence, $M_1^{(n)}(\omega)$ is equal to $X_{i_1^\star(\omega)}(\omega)$ for all n large enough, and therefore converges to $X_{i_1^\star(\omega)}(\omega)$. Inductively define

$$i_m^\star(\omega) = \arg \min_{i \in \{i_j(i_{m-1}^\star(\omega),\omega)\}_{j=1}^{L+1}} \rho(X_i(\omega), X_t)$$

for $m \geq 2$. Starting from the relay node $i_{m-1}^\star(\omega)$ and repeating the same steps above for $i_{m-1}^\star(\omega)$, we have that $M_m^{(n)}(\omega)$ is equal to $X_{i_m^\star(\omega)}(\omega)$ for all n large enough. Therefore, we conclude that $M_m^{(n)}$ converges for all $m \geq 1$ as n goes to infinity on the event Ω_0. As a result, δ-GGF is a path convergent rule for any $\delta > 0$.

To see why it is also time convergent, first observe that for any given $X_s \in D$ and $X_t \in D$,

$$\liminf_{n\to\infty} \tau_n = \tau < \infty \text{ w.p.1}$$

since it achieves a forward progress at least $r - \delta$ in the direction of the target node at each hop. Secondly, observe that δ-GGF rule delivers the message to the target node at the next hop whenever the message enters $B(X_t, r)$. By appealing to Theorem 3, we conclude the proof. ■

We write $f_1(k) \sim f_2(k)$ as $k \to \infty$ f $\lim_{k\to\infty} \frac{f_1(k)}{f_2(k)} = 1$.

Lemma 2: Let $a_k \to 0$ and $b_k \to \infty$ as $k \to \infty$. Then, $(1+a_k)^{b_k} \sim \exp(a_k b_k)$ if and only if $(a_k)^2 b_k \to 0$ as $k \to \infty$.

Proof: $\Rightarrow$: If $(1 + a_k)^{b_k} \sim \exp(a_k b_k)$, we have

$$\begin{aligned} \lim_{k\to\infty} \frac{\exp\left(b_k \log\left(1+a_k\right)\right)}{\exp\left(a_k b_k\right)} &= 1 \\ \Rightarrow \lim_{k\to\infty} \exp\left(a_k b_k \left(\frac{\log\left(1+a_k\right)}{a_k} - 1\right)\right) &= 1 \\ \Rightarrow \lim_{k\to\infty} a_k b_k \left(\frac{\log\left(1+a_k\right)}{a_k} - 1\right) &= 0 \end{aligned}$$

Since a_k goes to zero as k goes to infinity, we can write the power series expansion of $\log(1 + a_k)$ as

$$\log(1+a_k) = a_k - \frac{{a_k}^2}{2} + \frac{{a_k}^3}{3} - \frac{{a_k}^4}{4} + \cdots.$$

Thus, we have

$$\frac{\log(1+a_k)}{a_k} - 1 = a_k \sum_{l=0}^{\infty} \frac{1}{l+2}(-1)^{l+1} a_k^l.$$

As a result, we need to have

$$\lim_{k\to\infty} b_k a_k^2 \sum_{l=0}^{\infty} \frac{1}{l+2}(-1)^{l+1} a_k^l = 0.$$

Since a_k goes to zero as k goes to infinity, we have, for sufficiently large k, $\left|\frac{1}{l+2}(-1)^{l+1} a_k^l\right| \leq 0.5^l$. Then, by Lebesgue dominated convergence theorem, we obtain

$$\begin{aligned} \lim_{k\to\infty} \sum_{l=0}^{\infty} \frac{1}{l+2}(-1)^{l+1} a_k^l &= \sum_{l=0}^{\infty} \lim_{k\to\infty} \frac{1}{l+2}(-1)^{l+1} a_k^l \\ &= \sum_{l=0}^{\infty} \frac{1}{l+2}(-1)^{l+1} \mathbb{1}_{\{l=0\}} \\ &= -\frac{1}{2}. \end{aligned}$$

Therefore, we conclude that $b_k a_k^2$ converges to zero as k goes to infinity.

$\Leftarrow$: Start by observing that

$$\begin{aligned} \frac{(1+a_k)^{b_k}}{\exp(a_k b_k)} &= \frac{\exp\left(b_k \log\left(1+a_k\right)\right)}{\exp\left(a_k b_k\right)} \\ &= \exp\left(a_k^2 b_k \sum_{l=0}^{\infty} \frac{1}{l+2}(-1)^{l+1} a_k^l\right) \end{aligned}$$

for all k large enough. We have showed that

$$\lim_{k\to\infty} \sum_{l=0}^{\infty} \frac{1}{l+2}(-1)^{l+1} a_k^l = -\frac{1}{2}.$$

Since $a_k^2 b_k$ goes to zero as k goes to infinity, by using the continuity property of $\exp(\cdot)$, we have

$$\begin{aligned} &\lim_{k\to\infty} \frac{(1+a_k)^{b_k}}{\exp(a_k b_k)} \\ &\quad = \exp\left(\lim_{k\to\infty} a_k^2 b_k \lim_{k\to\infty} \sum_{l=0}^{\infty} \frac{1}{k+2}(-1)^{l+1} a_k^l\right) \\ &\quad = 1. \end{aligned}$$

■

We will use Voronoi tessellations (see [17]) to show the tail of the distribution of the message delivery time decays to zero exponentially fast as the number of relay nodes goes to infinity when nodes use δ-GGF rule to deliver messages to the target node.

Definition 8: Let $(D, \mathcal{S}_D, \lambda, \rho)$ be a regular network space. Then, for any given finite set of points $\{x_i\}_{i=1}^p$ of D, the Voronoi tessellation $\mathcal{V}$ of D generated by $\{x_i\}_{i=1}^p$ is the collection of cells $\{V(x_i)\}_{i=1}^p$ such that

$$V(x_i) = \{x \in D : \rho(x, x_i) = \min_{1\leq j\leq p} \rho(x, x_j)\}.$$

The following lemma shows that we can find a Voronoi tessellation of the network space such that Voronoi cells are neither too fat nor too thin. The proof of Lemma 3 is adapted from [15] for regular network spaces.

Lemma 3: Let $(D, \mathcal{S}_D, \lambda, \rho)$ be a regular network space. Then, for any given $\epsilon > 0$, there is a Voronoi tessellation $\mathcal{V} = \{V(x_i)\}_{i=1}^{p^\star}$ of D such that

$$B(x_i, \epsilon) \subseteq V(x_i) \subseteq B(x_i, 2\epsilon)$$

for all $i \in \{1, 2, \cdots, p^\star\}$.

Proof: Let $\{x_i\}_{i=1}^p$ be a collection of points in D such that $\rho(x_i, x_j) \geq 2\epsilon$ for all $i, j \in \{1, 2, \cdots, p\}$ with $i \neq j$.

Assume there exists $x \in D$ such that $B(x, \epsilon) \bigcap B(x_i, \epsilon) = \emptyset$ for all $i \in \{1, 2, \cdots, p\}$. Since $(D, \mathcal{S}_D, \lambda, \rho)$ is a regular network space, this condition means that $\rho(x, x_i) \geq 2\epsilon$ for all $i \in \{1, 2, \cdots, p\}$. Therefore, we can add x to this collection of points by letting $x_{p+1} = x$.

Observe that for any such collection of points, we have $B(x_i, \epsilon) \bigcap B(x_j, \epsilon) = \emptyset$ for all $i, j \in \{1, 2, \cdots, p\}$ with $i \neq j$. Thus, by using the regularity of the network space, we have

$$\begin{aligned} \lambda\left(\bigcup_{i=1}^p B(x_i, \epsilon)\right) &= p \cdot \lambda\left(B\left(x_1, \epsilon\right)\right) \\ &\leq \lambda(D) < \infty, \end{aligned}$$

which implies any such collection can only contain finitely many points. Therefore, we must stop adding new points to this collection at some finite $p^\star$. Let $\mathcal{V} = \{V(x_i)\}_{i=1}^{p^\star}$ be the Voronoi tessellation of D generated by $\{x_i\}_{i=1}^{p^\star}$.

We first show that $B(x_i, \epsilon) \subseteq V(x_i)$. Suppose this is not correct. Then, there exists a $y \in B(x_i, \epsilon)$ such that $y \in V(x_j)$ for some $x_j \neq x_i$. Thus, $\rho(x_j, y) \leq \rho(x_i, y) < \epsilon$. Then,

$$\begin{aligned}\rho(x_i, x_j) &\leq \rho(x_i, y) + \rho(y, x_j) \\ &< 2\epsilon,\end{aligned}$$

which contradicts how we construct the set $\{x_i\}_{i=1}^{p^\star}$. Secondly, we show that $V(x_i) \subseteq B(x_i, 2\epsilon)$. Suppose this is not correct. Then, there exists $y \in V(x_i)$ such that $\rho(x_i, y) \geq 2\epsilon$. Since $\rho(x_i, y) = \min_{1 \leq j \leq p^\star} \rho(x_j, y)$, $\rho(x_j, y) \geq 2\epsilon$ for all $j \in \{1, 2, \cdots, p^\star\}$. Therefore, y can be added to this collection of points, which contradicts how we choose $p^\star$. ■

Lemma 4: Let $(D, \mathcal{S}_D, \lambda, \rho)$ be a regular network space. Then, for any given source-target separation $d = \rho(X_s, X_t)$ and $\delta > 0$, there exists a constant $B > 0$ such that τ_n under δ-GGF satisfies

$$\mathbb{P}\{\tau_n > B\} \leq \exp(-c \cdot f(n)), \tag{5}$$

where $c > 0$ is a constant independent of d and δ, and $f(n) \sim n^{2\epsilon}$ for some $\epsilon \in \left(0, \frac{1}{4}\right)$.

Proof: Choose an ϵ belonging to $\left(0, \frac{1}{4}\right)$, and a sequence $\{\epsilon_n\}_{n \geq N_{\text{init}}}$ of positive real numbers decreasing to 0 and satisfying

$$\lambda\left(B(x, \epsilon_n)\right) \geq \lambda(D) \left\lceil n^{\frac{1}{2}-\epsilon} \right\rceil^{-2}, \tag{6}$$

for some $x \in D$. By regularity of the network domain, (6) holds for all points $x \in D$.

When there are n relay nodes in D, consider a Voronoi tessellation $\mathcal{V}_n = \left\{V\left(x_i^{(n)}\right)\right\}_{i=1}^{p_n^\star}$ of D such that

$$B\left(x_i^{(n)}, \epsilon_n\right) \subseteq V\left(x_i^{(n)}\right) \subseteq B\left(x_i^{(n)}, 2\epsilon_n\right)$$

for all $i \in \{1, 2, \cdots, p_n^\star\}$. For any given $\epsilon_n > 0$, such a Voronoi tessellation exists by Lemma 3. Since relay nodes are uniformly distributed over D, the probability that relay node j belongs to Voronoi cell $V\left(x_i^{(n)}\right)$ satisfies

$$\begin{aligned}\mathbb{P}\left\{X_j \in V\left(x_i^{(n)}\right)\right\} &= \frac{\lambda\left(V\left(x_i^{(n)}\right)\right)}{\lambda(D)} \\ &\geq \frac{\lambda\left(B\left(x_i^{(n)}, \epsilon_n\right)\right)}{\lambda(D)} \\ &\geq \left\lceil n^{\frac{1}{2}-\epsilon} \right\rceil^{-2}.\end{aligned}$$

Therefore,

$$\begin{aligned}&\mathbb{P}\left\{X_j \notin V\left(x_i^{(n)}\right), \forall j \in \{1, 2, \cdots, n\}\right\} \\ &\qquad \leq \left(1 - \left\lceil n^{\frac{1}{2}-\epsilon} \right\rceil^{-2}\right)^n.\end{aligned}$$

Let $E_n = \{\omega \in \Omega : \exists i \in \{1, 2, \cdots, p_n^\star\} \text{ s.t. } X_j(\omega) \notin V\left(x_i^{(n)}\right), \forall j \in \{1, 2, \cdots, n\}\}$. By using the union bound and Lemma 2, we have

$$\begin{aligned}\mathbb{P}(E_n) &\leq p_n^\star \cdot \left(1 - \left\lceil n^{\frac{1}{2}-\epsilon} \right\rceil^{-2}\right)^n \\ &\leq p_n^\star \cdot \exp\left(-0.25 n^{2\epsilon}\right).\end{aligned}$$

Since

$$\begin{aligned}\lambda(D) \geq \lambda\left(\bigcup_{i=1}^{p_n^\star} B\left(x_i^{(n)}, \epsilon_n\right)\right) &= p_n^\star \cdot \lambda\left(B\left(x_1^{(n)}, \epsilon_n\right)\right) \\ &\geq p_n^\star \cdot \left\lceil n^{\frac{1}{2}-\epsilon} \right\rceil^{-2} \cdot \lambda(D)\end{aligned}$$

we have $p_n^\star \leq \left\lceil n^{\frac{1}{2}-\epsilon} \right\rceil^2$. Thus,

$$\begin{aligned}\mathbb{P}(E_n) &\leq 2\left(n^{1-2\epsilon}\right) \exp\left(-0.25 n^{2\epsilon}\right) \\ &= \exp\left(-c \cdot f(n)\right),\end{aligned}$$

where $f(n) = n^{2\epsilon} - 4(1-2\epsilon)\log(n) - 4\log(2)$ and $c = 0.25$.

Now, we will show that all open sets $U \subseteq D$ will eventually contain a Voronoi cell for all n large enough. For any open set U, there exists $y \in U$ and $\epsilon > 0$ such that $B(y, \epsilon) \subseteq U$. Choose the Voronoi cell $V\left(x_{i_y}^{(n)}\right)$, $i_y \in \{1, 2, \cdots, p_n^\star\}$, containing y. Then, $\rho\left(y, x_{i_y}^{(n)}\right) \leq 2\epsilon_n$. As a result, $x_{i_y}^{(n)} \in B(y, \epsilon)$ for all n large enough. Consider now $B\left(x_{i_y}^{(n)}, 2\epsilon_n\right)$. For any $z \in B\left(x_{i_y}^{(n)}, 2\epsilon_n\right)$,

$$\begin{aligned}\rho(y, z) &\leq \rho\left(y, x_{i_y}^{(n)}\right) + \rho\left(x_{i_y}^{(n)}, z\right) \\ &\leq 4\epsilon_n.\end{aligned}$$

Therefore, whenever $\epsilon_n < \frac{\epsilon}{4}$, we have that Voronoi cell containing y lies in $B(y, \epsilon)$, and U contains a Voronoi cell for all n large enough.

Therefore, on the complement E_n' of the event E_n, the message can make forward progress at least $r - \delta$ amount towards destination at each hop for all n large enough. Let $B = \left\lfloor \frac{d}{r-\delta} + 1 \right\rfloor$. Then, $E_n' \subseteq \{\tau_n \leq B\}$ for all n large enough. Thus,

$$\begin{aligned}\mathbb{P}\{\tau_n > B\} &= 1 - \mathbb{P}\{\tau_n \leq B\} \\ &\leq 1 - \mathbb{P}\left(E_n'\right) \\ &\leq \exp\left(-c \cdot f(n)\right),\end{aligned} \tag{7}$$

for all n large enough. ■

We finally have the next theorem relating the average message delivery time for dense networks with finitely many nodes to the average message delivery time in the continuum limit when nodes use δ-GGF rule to deliver messages to the target node.

Theorem 5: Let $(D, \mathcal{S}_D, \lambda, \rho)$ be a regular network space. Then, for any given $X_s \in D$ and $X_t \in D$, if nodes use the δ-GGF rule to deliver messages to the target node, τ_n converges to a limiting random variable τ and

$$\lim_{n \to \infty} \mathbb{E}\left[\tau_n \mathbb{1}_{\{\tau_n < \infty\}}\right] = \mathbb{E}\left[\tau\right].$$

Since the average message delivery time can be made arbitrarily close to the average message delivery time in the continuum limit by increasing the number of relay nodes when nodes employ δ-GGF rule to deliver messages to the target node, we will focus on the continuum limit and the δ-GGF rule for the rest of the paper.

B. Average Message Delivery Time without Long-range Contacts

We first consider the message delivery time for regular network spaces when nodes do not have long range contacts.

Theorem 6: For any given $X_s \in D$ and $X_t \in D$, τ, with probability one, is equal to

$$\tau = \begin{cases} 0 & \text{if } X_s = X_t, \\ \left\lfloor \frac{\rho(X_s, X_t)}{r} \right\rfloor + 1 & \text{if } d(X_s, X_t) > 0 \end{cases} \tag{8}$$

under δ-GGF with no long-range contacts.

Proof: We prove this theorem by induction. If $\rho(X_s, X_t) = 0$, then there is no need for message forwarding since source and target nodes are located at the same position.

If $0 < \rho(X_s, X_t) < r$, then the message is delivered to the destination in one hop since the target node is in the local communication range of the source node. So, (8) holds for $\rho(X_s, X_t) \in (0, r)$.

If $r \leq \rho(X_s, X_t) < 2r$, then the message is delivered to the destination in two hops. Note that $\lambda\left(B(X_t, r) \bigcap B(X_s, r)\right) > 0$ for this range of source-target separation since the network space is regular. Hence, $B(X_t, r) \bigcap B(X_s, r)$ is a non-empty open subset of the network domain, and the source node has local contacts lying in $B(X_t, r)$ and achieving forward progress in the direction of the target node at least $r - \delta$ w.p.1 by Theorem 1. Therefore, in the first hop, the message enters the ball $B(X_t, r)$ by means of one of the local contacts of the source node lying in $B(X_t, r)$, and in the second hop it is delivered to the destination. As a result, (8) also holds when $\rho(X_s, X_t) \in [r, 2r)$.

Suppose now (8) is correct for $\rho(X_s, X_t) \in [kr, (k+1)r)$ for $k \geq 2$. Then, $\tau = k + 1$ for this range of source-target separation. We now show that it also holds for $\rho(X_s, X_t) \in [(k+1)r, (k+2)r)$. By regularity of the network space,

$$\lambda\left(B(X_t, (k+1)r) \bigcap B(X_s, r)\right) > 0.$$

Therefore, the source node has local contacts lying in $B(X_t, (k+1)r)$ and achieving forward progress at least $r - \delta$ w.p.1 by Theorem 1. Since the local communication range of the source node is r, all of its local contacts lying in $B(X_t, (k+1)r)$ are located in $B(X_t, (k+1)r) - B(X_t, kr)$. Therefore,

$$\rho(M_1, X_t) \in [kr, (k+1)r)$$

after the first hop. By the induction hypothesis, it takes $k+1$ hops to reach the destination after the first hop. Therefore, $\tau = k+2$ if $\rho(X_s, X_t) \in [(k+1)r, (k+2)r)$. This completes the proof. ∎

C. Piece-wise Constant Behavior and Spherical Symmetry Property of $\mathbb{E}[\tau]$ in Small-world Networks

We will now analyze some basic properties of the average message delivery time in small-world networks for regular network spaces. We first show its piece-wise constant behavior with jumps at exactly integer multiples of nodes' local communication radius r. This behavior is merely due to the communication model that we employ for forming local contacts. We assume that two nodes are local contacts of each other, and thereby are able to communicate with each other if the separation between them is smaller than r. On the other hand, if the separation between them is greater than $r + \epsilon$ for any $\epsilon \geq 0$, they are not local contacts of each other, and therefore are not able to communicate with each other unless there is a long-range link connecting them. This discontinuity in the rule for forming local contacts, in turn, leads to discontinuities in the average message delivery time in small-world networks when it is viewed as a function of the source-target separation.

Theorem 7: Consider a small-world network constructed on a regular network space $(D, \mathcal{S}_D, \rho, \lambda)$, and containing nodes that have local communication range r and uniformly distributed long-range contacts on D. If nodes employ the δ-GGF rule to relay messages for some $\delta > 0$, then for any $k \geq 2$, $\mathbb{E}[\tau]$ stays the same for all $\rho(X_s, X_t) \in [(k-1)r, kr)$.

Proof: We prove this proposition by induction. First note that the message is delivered to the target node in one hop if $\rho(X_s, X_t) \in (0, r)$ since source and target nodes are local contacts of each other. We then show that the proposition is correct for $k = 2$. $\rho(X_s, X_t) \in [r, 2r)$ when $k = 2$, and $B(X_s, r) \bigcap B(X_t, r)$ is a non-empty open subset of the network domain for this range of source-target separation by regularity of the network space. As a result, the source node has some local contacts lying in $B(X_t, r)$ and achieving forward progress at least $r - \delta$ by Theorem 1. Therefore, the message first enters the ball $B(X_t, r)$ through either a short-range contact or a long-range contact of the source on the first hop, and then is delivered to the target node on the second hop. Hence, the message delivery time is equal to 2 under δ-GGF when $\rho(X_s, X_t) \in [r, 2r)$, which also proves that $\mathbb{E}[\tau]$ is constant for this range of source-target separation.

Now, assume that the proposition holds for any $k \in \{2, 3, \cdots, N\}$, where $N \geq 3$. For any $k \in \{2, 3, \cdots, N\}$, let g_k be the corresponding constant value when $\rho(X_s, X_t) \in [(k-1)r, kr)$. Set g_1 to 1 since it takes one hop to deliver the packet to the target when $\rho(X_s, X_t) \in (0, r)$. We now show that the theorem continues to hold for $N + 1$. To this end, we divide the ball $B(X_t, Nr)$ into N concentric discs centered at X_t as follows.

$$B(X_t, Nr) = \bigcup_{k=1}^{N} C(X_t, (k-1)r, kr),$$

where $C(X_t, r_1, r_2) = B(X_t, r_2) - B(X_t, r_1)$ for any $r_2 > r_1$. By regularity of the network space and Theorem 1, the source node has a local contact lying in $C(X_t, (N-1)r, Nr)$ and achieving forward progress at least $r - \delta$ when $\rho(X_s, X_t) \in [Nr, (N+1)r)$. This observation combined with the fact that long-range contacts are formed uniformly over the network domain implies that the probability of M_1 lying in a concentric disc $C(X_t, (k-1)r, kr)$ for a given $k \in \{1, 2, \cdots, N\}$ is the same for all source nodes with $\rho(X_s, X_t) \in [Nr, (N+1)r)$. By the induction hypothesis, the average message delivery time after the first hop depends only on in which concentric disc M_1 lies in but not its exact position. Therefore, we obtain another constant value which is the same for all $\rho(X_s, X_t) \in [Nr, (N+1)r)$ after averaging g_k's by using the corresponding probabilities. ■

An important corollary of this proposition is the spherical symmetry property of $\mathbb{E}[\tau]$ for small-world networks. That is, for any given target node, consider the boundary $\partial B(X_t, d)$ of the ball centered at X_t with radius d. Then, wherever the source node starts transmitting its message on $\partial B(X_t, d)$, it takes, on the average, the same amount of time for this message to reach the target node.

Corollary 2: For any given $d > 0$, $\mathbb{E}[\tau]$ is the same for all $X_s \in \partial B(X_t, d)$.

Therefore, the average value of τ is only a function of the separation between the source-target pair, and there is no ambiguity in the following definition.

Definition 9: $g(d) \triangleq \mathbb{E}^d[\tau]$ is the average delivery time of a message to the final destination X_t starting from the source node X_s when $\rho(X_s, X_t) = d$.

D. Average Message Delivery Time for Small-world Networks Constructed on the Plane When $L = 1$

We will first calculate the average message delivery time for small-world networks constructed on the plane. Then, we will extend the same analysis to regular network spaces.

To this end, we consider an R by R square, and uniformly distribute relay nodes over this square. Let P denote this R by R square. Therefore, the network space for this case is $(P, \mathcal{S}_P, \lambda, \rho)$, where $\mathcal{S}_P$ is the Borel σ-algebra on P, λ is the Lebesgue measure on $\mathcal{S}_P$ and ρ is the usual Euclidean distance.

Let $d = \rho(X_s, X_t)$ be the distance between source and target nodes. Note that $(P, \mathcal{S}_P, \lambda, \rho)$ is not a regular network space since the translation invariance property does not hold due to edge effects. Therefore, we choose X_t such that $B(X_t, d+r) \subset P$ in order to eliminate possible edge effects that can occur. This choice, in turn, necessitates R to be greater than $2\cdot(d+r)$. When we extend the analysis presented below to small-world networks constructed on regular network spaces, we will not run into the edge-effect problem.

The δ-GGF rule induces a probability distribution, let us denote this distribution by μ, on the set of all possible paths between source and target nodes. For each such path, there is an associated message delivery time. Therefore, the average delivery time of a message from the source node to the target node can be obtained by integrating the message delivery time for all paths with respect to the probability distribution induced on the path space by the δ-GGF rule:

$$\mathbb{E}[\tau] = \int_{\mathcal{P}:\ \mathcal{P} \text{ is a path between } X_s \text{ and } X_t} \tau(\mathcal{P}) \mathrm{d}\mu(\mathcal{P}).$$

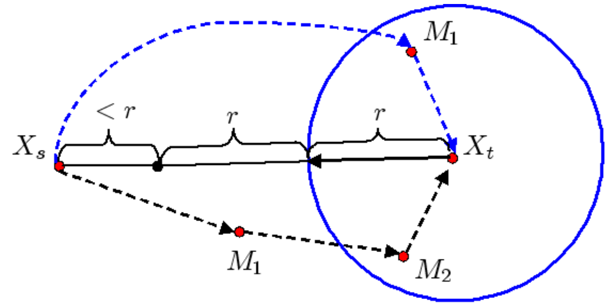

Fig. 3. Exemplary trajectories of the message when $2r \leq \rho(X_s, X_t) < 3r$.

However, such an approach is prohibitive since calculating the distribution induced by the δ-GGF rule on all possible paths between source and destination nodes is intractable. Therefore, we employ a Markov chain based first step analysis to obtain a recursive equation for the average delivery time of a message in small world networks.

To this end, let us first calculate $g(d)$ for several ranges of d before going into a more general solution.

- **(a)** $d = 0$: Then, $g(d) = 0$. There is no need to transmit the message.
- **(b)** $0 < d < r$: Then, $g(d) = 1$. We reach the destination in exactly one hop.
- **(c)** $r \leq d < 2r$: Then, $g(d) = 2$. The source can reach the destination in exactly two hops. Note that the source node has a contact lying in the ball $B(X_t, r)$ and achieving forward progress $r - \delta$ w.p.1 since $\mathcal{H}_\infty$ is dense in P. Therefore, on the first hop, the message enters $B(X_t, r)$ either through a long-range contact or a local contact. On the second hop, it is delivered to X_t.
- **(d)** $2r \leq d < 3r$: Here, $g(d)$ is a little bit tricker to obtain. The key is to analyze where the message can be located after the first hop. It can be delivered to X_t in two hops if it makes its way through $B(X_t, r)$ on the first hop. Otherwise, it first enters the disc $C(X_t, r, 2r) = B(X_t, 2r) - B(X_t, r)$, and then it is delivered to the target in two hops thereafter. As a result, the second path takes three hops to deliver the message to the final destination. Exemplary trajectories of the message are shown in Fig. 3.

These observations can be formally expressed as follows.

$$\begin{aligned} g(d) &= 1 + \mathbb{E}\left[\mathbb{E}^{\rho(M_1, X_t)}[\tau]\right] \\ &= 1 + \mathbb{E}\Big[\mathbb{E}^{\rho(M_1, X_t)}[\tau] \cdot \mathbb{1}_{\{0<\rho(M_1,X_t)<r\}} \\ &\qquad + \mathbb{E}^{\rho(M_1, X_t)}[\tau] \cdot \mathbb{1}_{\{r\leq\rho(M_1,X_t)<2r\}}\Big] \\ &= 1 + \mathbb{P}\{0 < \rho(M_1, X_t) < r\} \\ &\qquad + 2 \cdot \mathbb{P}\{r \leq \rho(M_1, X_t) < 2r\}. \end{aligned} \tag{9}$$

By letting

$$\alpha = \mathbb{P}\{0 < \rho(M_1, X_t) < r\} = \frac{\pi r^2}{R^2 - \pi r^2}, \tag{10}$$

we obtain

$$g(d) = 1 + \alpha + 2(1 - \alpha) \quad \text{for } 2r \leq d < 3r. \tag{11}$$

General calculations for any d lying in $\left[k \cdot r, (k+1) \cdot r\right)$, $k \geq 2$, are also in this spirit. We first look at where the message can be located at the first hop, then analyze the average message delivery time from this point on. This analysis leads to a recursive solution for $g(d)$ for any value of d. To this end, by using the piece-wise constant behavior of $g(d)$, we let

$$\begin{aligned} g(d) &= g_0 \text{ when } d = 0, \\ g(d) &= g_1 \text{ when } 0 < d < r, \\ g(d) &= g_k \text{ when } (k-1)r \leq d < kr,\ 2 \leq k \leq \lfloor \tfrac{R}{2r} - 1 \rfloor, \\ g(d) &= g_{\lfloor \frac{R}{2r} - 1 \rfloor + 1} \text{ when } \lfloor \tfrac{R}{2r} - 1 \rfloor r \leq d \leq \tfrac{R}{2} - r. \end{aligned}$$

Consider g_{k+1} when $2 \leq k \leq \lfloor \frac{R}{2r} - 1 \rfloor$. Then,

$$\begin{aligned} g_{k+1} &= 1 + \mathbb{E}\left[\mathbb{E}^{\rho(M_1, X_t)}\left[\tau\right]\right] \\ &= 1 + \mathbb{E}\left[\sum_{i=1}^{k} \mathbb{E}^{\rho(M_1, X_t)}\left[\tau\right] \mathbb{1}_{\{(i-1)r \leq \rho(M_1, X_t) < ir\}}\right] \\ &= 1 + \mathbb{E}\left[\sum_{i=1}^{k} g_i \mathbb{1}_{\{(i-1)r \leq \rho(M_1, X_t) < ir\}}\right] \\ &= 1 + \left(1 - \alpha(k-1)^2\right) g_k + \alpha \sum_{i=1}^{k-1} (2i-1) g_i. \end{aligned}$$

Note that in the third equality above, there is no harm in writing the limits of the indicator function as $0 \leq \rho(M_1, X_t) < r$ when $i = 1$ since $\mathbb{P}\{\rho(M_1, X_t) = 0\} = 0$. To obtain a second order non-constant coefficient linear recursive equation, we subtract g_k from g_{k+1}.

$$g_{k+1} - g_k = (g_k - g_{k-1}) \cdot \left(1 - \alpha(k-1)^2\right). \quad (12)$$

Let $u_k = g_{k+1} - g_k$ and $\beta_k = 1 - \alpha(k-1)^2$. Then, u_k satisfies the following first order linear recursive equation.

$$u_k = \beta_k \cdot u_{k-1} \text{ for } k \geq 1, \quad (13)$$

with the initial condition $u_0 = 1$. Observe that $u_k = \prod_{i=1}^{k} \beta_i$ for $k \geq 1$. Then, the solution for (12) is obtained as

$$\begin{aligned} g_{k+1} &= u_k + u_{k-1} + \cdots + u_0 \\ &= 1 + \sum_{j=1}^{k} \prod_{i=1}^{j} \beta_i \text{ for } k \geq 1. \quad (14) \end{aligned}$$

The following theorem summarizes these findings.

Theorem 8: Consider a small-world network constructed on $(P, \mathcal{S}_P, \lambda, \rho)$, and containing nodes that have local communication range r and one uniformly distributed long-range outgoing contact on P. For $i \geq 1$, let $\beta_i = 1 - \alpha(i-1)^2$, where α is defined as in (10). If nodes employ the δ-GGF rule to relay messages, then the average message delivery time, for $1 \leq k \leq \lfloor \frac{R}{2r} - 1 \rfloor$, is given by

$$g_{k+1} = 1 + \sum_{j=1}^{k} \prod_{i=1}^{j} \beta_i, \quad (15)$$

where $g_k = g(d)$ when $(k-1)r \leq d < kr$, $g_{\lfloor \frac{R}{2r} - 1 \rfloor + 1} = g(d)$ when $\lfloor \frac{R}{2r} - 1 \rfloor r \leq d \leq \frac{R}{2} - r$, and $g_0 = 0$ and $g_1 = 1$ are the initial conditions.

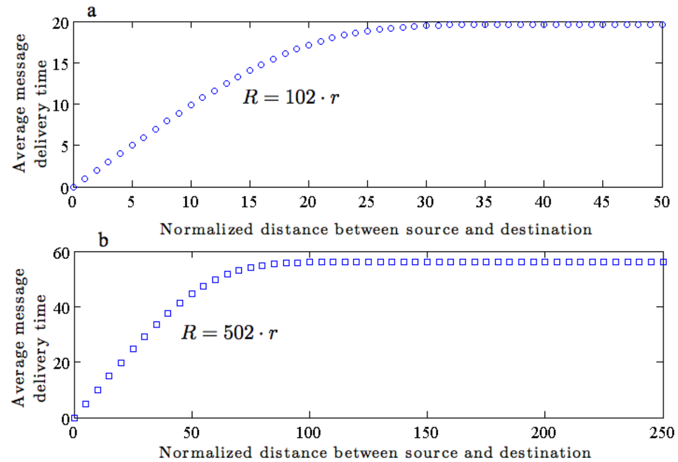

Fig. 4. Change of average message delivery time as a function of normalized source-destination separation. a, $R = 102 \cdot r$. b, $R = 502 \cdot r$. In both figures, the horizontal axis is normalized with respect to r.

In Fig 4-a and Fig. 4-b , we plot the change of average message delivery time with respect to the separation between source and target nodes for two different network sizes. In both figures, the horizontal axis is normalized with respect to the local communication range r of nodes. Fig. 4 reveals that the average message delivery time increases linearly for small values of separation between a source and target pair. On the other hand, it quickly converges to a constant, and remains essentially the same at this constant value for a broad range of values of source-target separations. This means that messages are first forwarded by means of long-range contacts until they enter a certain range of the target. From this point on, they are delivered to the target through local contacts. This is also in line with the observation of Travers and Milgram [7]: "*Chains which converge on the target principally by using geographic information reach his hometown or surrounding areas readily, but once there often circulate before entering targets circle of acquaintances.*".

When there are L outgoing long-range links for $L \geq 1$, the next theorem gives us the average message delivery time for small-world networks constructed on the plane. Its derivation will be given for the more general case of regular network spaces in the next section.

Theorem 9: Consider a small-world network constructed on $(P, \mathcal{S}_P, \lambda, \rho)$, and containing nodes that have local communication range r and L, $L \geq 1$, uniformly distributed long-range outgoing contacts on P. For $i \geq 1$, let $\beta_i = 1 - \alpha(i-1)^2$, where α is defined as in (10). If nodes employ the δ-GGF rule to relay messages, then the average message delivery time, for $1 \leq k \leq \lfloor \frac{R}{2r} - 1 \rfloor$, is given by

$$g_{k+1} = 1 + \sum_{j=1}^{k} \left(\prod_{i=1}^{j} \beta_i\right)^L, \quad (16)$$

where $g_k = g(d)$ when $(k-1)r \leq d < kr$, $g_{\lfloor \frac{R}{2r} - 1 \rfloor + 1} = g(d)$ when $\lfloor \frac{R}{2r} - 1 \rfloor r \leq d \leq \frac{R}{2} - r$, and $g_0 = 0$ and $g_1 = 1$ are the initial conditions.

E. Average Message Delivery Time for Small-world Networks Constructed on Regular Network Spaces

1) Calculations for One Outgoing Long-range Link: The calculations for obtaining the average message delivery time for small-world networks constructed on general regular network spaces are similar to the calculations presented above. We first analyze where the message may lie after the first message forwarding, and obtain a recursive expression for the average message delivery time.

Let $(D, \mathcal{S}_D, \lambda, \rho)$ be a regular network space. For any given $x \in D$ and $i \geq 1$, we define

$$b_i(r) = \lambda(B(x, ir)) \tag{17}$$

and

$$\begin{aligned} c_i(r) &= \lambda\left(C(x, (i-1)r, ir)\right) \\ &= \lambda\{y \in D : (i-1)r \leq \rho(x,y) < ir\}. \end{aligned} \tag{18}$$

Not that $b_i(r)$ and $c_i(r)$ do not depend on the x around which they are centered because of translation invariance property of regular network spaces. Note also that $b_0(r) = 0$ since $B(x, 0) = \emptyset$.

Definition 10: The diameter of a given network space $(D, \mathcal{S}_D, \lambda, \rho)$ is equal to

$$R = \sup_{x,y \in D} \rho(x,y). \tag{19}$$

Let R be the diameter of $(D, \mathcal{S}_D, \lambda, \rho)$. For $0 \leq d < 2r$, it is easy to see that $g(0) = 0$, $g(d) = 1$ for $0 < d < r$ and $g(d) = 2$ for $r \leq d < 2r$. For other values of d, we define g_k's iteratively as in planar small world networks: $g(0) = g_0$, $g(d) = g_1$ for $0 < d < r$ and $g(d) = g_k$ for $(k-1)r \leq d < kr$ and $2 \leq k \leq \lfloor \frac{R}{r} \rfloor$. We also set $g_{\lfloor \frac{R}{r} \rfloor + 1} = g(d)$ for $\lfloor \frac{R}{r} \rfloor r \leq d \leq R$.

Let us now calculate g_{k+1} by focusing on $d \in [kr, (k+1)r)$. Then,

$$\begin{aligned} g_{k+1} &= 1 + \mathbb{E}\left[\mathbb{E}^{\rho(M_1, X_t)}\left[\tau\right]\right] \\ &= 1 + \mathbb{E}\left[\sum_{i=1}^{k} g_i \mathbb{1}_{\{(i-1)r \leq \rho(M_1, X_t) < ir\}}\right] \\ &= 1 + \sum_{i=1}^{k} g_i \mathbb{P}\left\{(i-1)r \leq \rho(M_1, X_t) < ir\right\}. \end{aligned}$$

By the non-atomicity property of the measure λ, $\mathbb{P}\{\rho(M_1, X_t) = 0\} = 0$. Therefore, there is no harm in writing the limits of the indicator function as $(i-1)r \leq \rho(M_1, X_t) < ir$ in the second equality above when $i = 1$. For $1 \leq i \leq k-1$, let

$$\begin{aligned} \alpha_i &= \mathbb{P}\left\{(i-1)r \leq \rho(M_1, X_t) < ir\right\} \\ &= \frac{c_i(r)}{\lambda(D) - b_1(r)}. \end{aligned} \tag{20}$$

For $i = k$, let

$$\begin{aligned} \beta_k &= \mathbb{P}\left\{(k-1)r \leq \rho(M_1, X_t) < kr\right\} \\ &= 1 - \frac{b_{k-1}(r)}{\lambda(D) - b_1(r)}. \end{aligned}$$

Then,

$$g_{k+1} = 1 + \beta_k g_k + \sum_{i=1}^{k-1} \alpha_i g_i.$$

Let $u_k = g_{k+1} - g_k$. Then, $u_k = \beta_k g_k - g_{k-1}(\beta_{k-1} - \alpha_{k-1})$. It is easy to show that

$$\begin{aligned} \beta_{k-1} - \alpha_{k-1} &= 1 - \frac{b_{k-1}(r)}{\lambda(D) - b_1(r)} \\ &= \beta_k. \end{aligned}$$

As a result, $u_k = \beta_k u_{k-1}$ for $k \geq 1$ with the initial condition $u_0 = 1$. Then, the solution for g_k is obtained as

$$g_{k+1} = 1 + \sum_{j=1}^{k} \prod_{i=1}^{j} \beta_i \text{ for } k \geq 1.$$

Theorem 10: Consider a small-world network constructed on a regular network space $(D, \mathcal{S}_D, \lambda, \rho)$ with diameter R, and containing nodes that have local communication range r and one uniformly distributed long-range contact on D. For $i \geq 1$, let

$$\beta_i = 1 - \frac{b_{i-1}(r)}{\lambda(D) - b_1(r)}, \tag{21}$$

where $b_i(r)$ is defined as in (17). If nodes employ the δ-GGF rule to relay messages, then the average message delivery time, for $1 \leq k \leq \lfloor \frac{R}{r} \rfloor$, is given by

$$g_{k+1} = 1 + \sum_{j=1}^{k} \prod_{i=1}^{j} \beta_i,$$

where $g_k = g(d)$ when $(k-1)r \leq d < kr$, $g_{\lfloor \frac{R}{r} \rfloor + 1} = g(d)$ when $\lfloor \frac{R}{r} \rfloor r \leq d \leq R$, and $g_0 = 0$ and $g_1 = 1$ are the initial conditions.

2) Calculations for L Outgoing Long-range Links: We now analyze the average message delivery time when nodes have $L \geq 1$ outgoing long-range links. g_k's for $k \geq 0$ are defined similarly. It continues to hold that $g_0 = 0$ and $g_1 = 1$, and it is easy to see that $g_2 = 2$. Let us now assume that $kr \leq d = \rho(X_s, X_t) < (k+1)r$ and $k \geq 2$. Then, the probability that one of the given long-range contacts of the source lies in $C(X_t, (i-1)r, ir)$ is equal to α_i for $1 \leq i \leq k-1$, where α_i is defined as in (20). In addition, the probability that one of the given long-range contacts of the source node does not lie in $B(X_t, ir)$ is equal to β_{i+1} for $1 \leq i \leq k-1$, where β_i is defined as in (21).

Therefore, for $1 \leq i \leq k-1$, we have

$$\begin{aligned} &\mathbb{P}\left\{(i-1)r \leq \rho(M_1, X_t) < ir\right\} \\ &\qquad = \sum_{j=1}^{L} \binom{L}{j} \alpha_i^j \beta_{i+1}^{L-j} \\ &\qquad = \beta_{i+1}^L \left(\left(1 + \frac{\alpha_i}{\beta_{i+1}}\right)^L - 1\right) \\ &\qquad = (\beta_{i+1} + \alpha_i)^L - \beta_{i+1}^L. \end{aligned} \tag{22}$$

Note that

$$\begin{aligned}\beta_{i+1}+\alpha_i &= 1-\frac{b_i(r)-c_i(r)}{\lambda(D)-b_1(r)}\\ &= 1-\frac{b_{i-1}(r)}{\lambda(D)-b_1(r)}\\ &= \beta_i.\end{aligned}$$

Therefore,

$$\mathbb{P}\left\{(i-1)r\le\rho(M_1,X_t)<ir\right\}=\beta_i^L-\beta_{i+1}^L.$$

We also find that

$$\mathbb{P}\left\{(k-1)r\le\rho(M_1,X_t)<kr\right\}=\beta_k^L.$$

As a result, we obtain the following recursion for g_k:

$$g_{k+1}=1+g_k\beta_k^L+\sum_{i=1}^{k-1}\left(\beta_i^L-\beta_{i+1}^L\right)g_i.$$

Let $u_k=g_{k+1}-g_k$ with the initial condition $u_0=1$ as before. Then, $u_k=\beta_k^L u_{k-1}$, which leads to:

$$u_k=\prod_{i=1}^{k}\beta_i^L.$$

Finally, the solution for g_{k+1}, when there are L outgoing long-range links, is obtained as

$$g_{k+1}=1+\sum_{j=1}^{k}\left(\prod_{i=1}^{j}\beta_i\right)^L. \qquad (23)$$

These findings are summarized formally in the following theorem.

Theorem 11: Consider a small-world network constructed on a regular network space $(D,\mathcal{S}_D,\lambda,\rho)$ with diameter R, and containing nodes that have local communication range r and $L\ge 1$ uniformly distributed long-range contacts on D. For $i\ge 1$, let

$$\beta_i=1-\frac{b_{i-1}(r)}{\lambda(D)-b_1(r)},$$

where $b_i(r)$ is defined as in (17). If nodes employ the δ-GGF rule to relay messages, then the average message delivery time, for $1\le k\le\lfloor\frac{R}{r}\rfloor$, is given by

$$g_{k+1}=1+\sum_{j=1}^{k}\left(\prod_{i=1}^{j}\beta_i\right)^L,$$

where $g_k=g(d)$ when $(k-1)r\le d<kr$, $g_{\lfloor\frac{R}{r}\rfloor+1}=g(d)$ when $\lfloor\frac{R}{r}\rfloor r\le d\le R$, and $g_0=0$ and $g_1=1$ are the initial conditions.

F. Small-world Networks Constructed on the Sphere Surface S^2

As an application of Theorem 10, we first study the average delivery time of a message for small-world networks constructed on the surface of a sphere with radius R and embedded in $\mathbb{R}^3$. Let $(S^2,\mathcal{S}_{S^2},\lambda,\rho)$ be the network space, where S^2 is the network domain (i.e., the surface of a sphere)

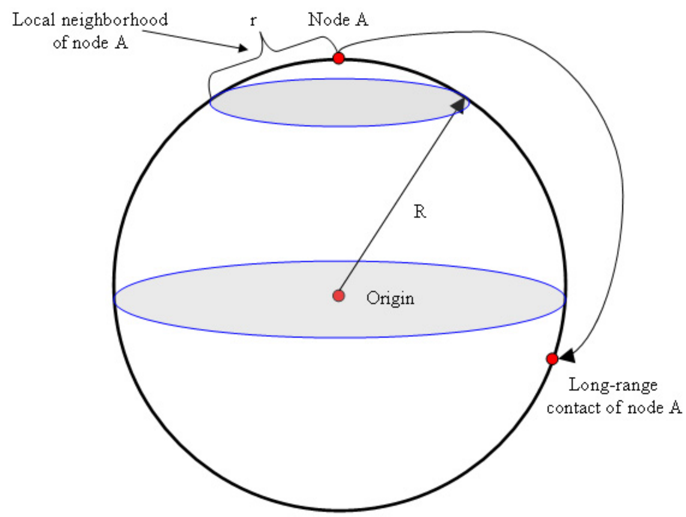

Fig. 5. An illustration of the surface of a sphere and a disk centered around a generic network node A with radius r on this surface. Short-range and long-range contacts of A are also shown in the figure.

$\mathcal{S}_{S^2}$ is the Borel σ-algebra of subsets of S^2, and λ is Lebesgue measure on $\mathcal{S}_{S^2}$. The distance metric ρ for the sphere surface is different than the usual Euclidean distance in that the distance between any two points on S^2 is measured in terms of the smaller arc length of the great circle connecting them. Therefore, the maximum separation between any two points can be at most πR, which is also the diameter of the network domain. Without loss of generality, we assume that the target node is located at one of the poles of the sphere, which is always possible by reorienting the coordinate axes.

We start our analysis by calculating the area of balls on S^2. In Fig. 5, we show the surface of a sphere and a ball centered around a generic network node A with radius r on this surface. A long-range contact of the node A on this surface is also illustrated in Fig. 5. The differential area element for S^2 can be written as

$$\mathrm{d}x=R^2\sin(\phi)\mathrm{d}\phi\mathrm{d}\theta.$$

On the disk with radius r, ϕ changes from 0 to $\frac{r}{R}$, and θ changes from 0 to 2π. Then, $b_1(r)$ is calculated as

$$\begin{aligned}b_1(r) &= \int_0^{2\pi}\int_0^{\frac{r}{R}}R^2\sin(\phi)\mathrm{d}\phi\mathrm{d}\theta\\ &= 2\pi R^2\left(1-\cos\left(\frac{r}{R}\right)\right).\end{aligned}$$

Similarly one can also obtain

$$b_i(r)=2\pi R^2\left(1-\cos\left(\frac{ir}{R}\right)\right) \qquad (24)$$

for any $i\ge 0$.

Let us define β_i for $i\ge 1$ as in (21). Then,

$$\begin{aligned}\beta_i &= 1-\frac{b_{i-1}(r)}{4\pi R^2-b_1(r)}\\ &= \frac{\cos(\theta)+\cos\left((i-1)\theta\right)}{1+\cos(\theta)},\end{aligned}$$

where $\theta = \frac{r}{R}$. We call θ *the clustering coefficient* of the network. Small values of θ correspond to networks that are weakly clustered. In such networks, it is very unlikely that distant nodes share a common contact. On the other hand, large values of θ signify that the network is strongly clustered. In these networks, most of the nodes either lie in another's local neighborhoods, or are connected to one another through a common contact.

As a result, we have the following theorem for the average message delivery time for small-world networks constructed on the surface of a sphere.

Theorem 12: Consider a small-world network constructed on the surface S^2 of a three dimensional sphere with radius R, and containing nodes that have local communication range r and L, $L \geq 1$, uniformly distributed long-range contacts on S^2. Let $(S^2, \mathcal{S}_{S^2}, \lambda, \rho)$ be the corresponding network space. For $i \geq 1$, let

$$\beta_i = \frac{\cos(\theta) + \cos\left((i-1)\theta\right)}{1 + \cos(\theta)},$$

where $\theta = \frac{r}{R}$. If nodes employ the δ-GGF rule to relay messages, then the average message delivery time, for $1 \leq k \leq \lfloor \frac{\pi R}{r} \rfloor$, is given by:

$$g_{k+1} = 1 + \sum_{j=1}^{k} \left(\prod_{i=1}^{j} \beta_i \right)^L,$$

where $g_k = g(d)$ when $(k-1)r \leq d < kr$, $g_{\lfloor \frac{\pi R}{r} \rfloor + 1} = g(d)$ when $\lfloor \frac{\pi R}{r} \rfloor r \leq d \leq \pi R$, and $g_0 = 0$ and $g_1 = 1$ are the initial conditions.

In Fig. 6-a and Fig. 6-b, we plot the change of average message delivery time with respect to the separation between source and target nodes for two different values of network clustering coefficient when $L = 1$. In each figure, the horizontal axis is normalized with respect to the local communication range of network nodes. We observe a similar behavior as in small-world networks constructed on the plane for the average message delivery time. The average message delivery time first increases linearly as the separation between source and target nodes is small. It then converges to a constant value, and remains constant at this value for a broad range of source-target separation.

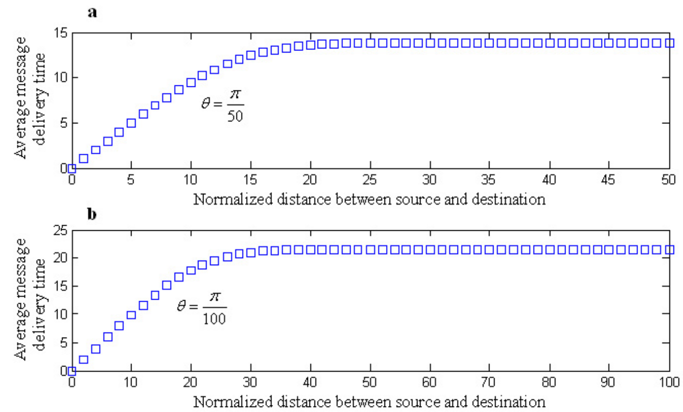

Fig. 6. Change of average message delivery time as a function of normalized source-destination separation for small-world networks constructed on the surface of a sphere. a, $\theta = \frac{\pi}{50}$. b, $\theta = \frac{\pi}{100}$. In both figures, the horizontal axis is normalized with respect to r.

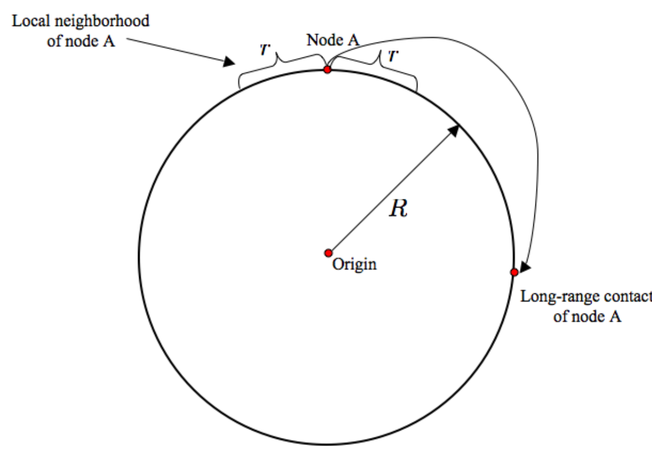

Fig. 7. An illustration of a generic node A and its short-range and long-range contacts for a small-world network constructed on the circle.

G. Small-world Networks Constructed on the Circle S^1

As another application of Theorem 10, we now study the average message delivery time for small-world networks constructed on the circle with radius R and embedded in $\mathbb{R}^2$. Let $(S^1, \mathcal{S}_{S^1}, \lambda, \rho)$ be the network space, where S^1 is the network domain, $\mathcal{S}_{S^1}$ is the Borel σ-algebra of subsets of S^1 and λ is Lebesgue measure on S^1. The distance metric ρ for the circle is also different from the usual Euclidean distance in that the distance between any two points on S^1 is measured in terms of the length of the smaller arc connecting them. Therefore, the maximum separation between any two points on S^1 is equal to πR, which is also the network diameter.

In Fig. 7, we show a generic network node A and its short-range and long-range neighbors for a small-world network constructed on the circle. For $i \geq 0$, $b_i(r)$ is equal to $2ir$ for S^1. Thus, for $i \geq 1$, we obtain

$$\beta_i = 1 - \frac{b_{i-1}(r)}{2\pi R - b_1(r)} = \frac{\pi - i\theta}{\pi - \theta}, \quad (25)$$

where $\theta = \frac{r}{R}$.

As a result, we have the following theorem for the average message delivery time for small-world networks constructed on the circle.

Theorem 13: Consider a small-world network constructed on the circle S^1 with radius R, embedded in $\mathbb{R}^2$ and containing nodes that have local communication range r and L, $L \geq 1$, uniformly distributed long-range contacts on S^1. Let $(S^1, \mathcal{S}_{S^1}, \lambda, \rho)$ be the corresponding network space. For $i \geq 1$, let

$$\beta_i = \frac{\pi - i\theta}{\pi - \theta},$$

where $\theta = \frac{r}{R}$. If nodes employ the δ-GGF rule to relay messages, then the average message delivery time, for $1 \leq$

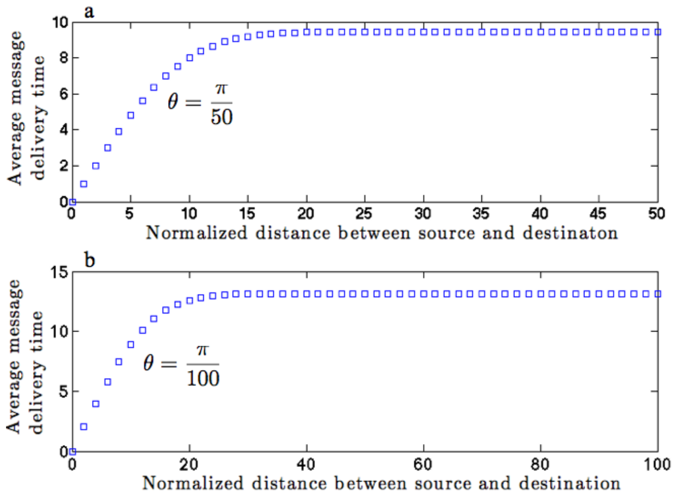

Fig. 8. Change of average message delivery time as a function of normalized source-destination separation for small-world networks constructed on the circle. a, $\theta = \frac{\pi}{50}$. b, $\theta = \frac{\pi}{100}$. In both figures, the horizontal axis is normalized with respect to r.

$k \leq \lfloor \frac{\pi R}{r} \rfloor$, is given by:

$$g_{k+1} = 1 + \sum_{j=1}^{k} \left(\prod_{i=1}^{j} \beta_i \right)^L ,$$

where $g_k = g(d)$ when $(k-1)r \leq d < kr$, $g_{\lfloor \frac{\pi R}{r} \rfloor + 1} = g(d)$ when $\lfloor \frac{\pi R}{r} \rfloor r \leq d \leq \pi R$, and $g_0 = 0$ and $g_1 = 1$ are the initial conditions.

In Fig. 8-a and Fig. 8-b, we plot the change of average message delivery time with respect to the separation between source and target nodes for two different values of network clustering coefficient when $L = 1$. In both figures, the horizontal axis is normalized with respect to local communication range of network nodes. We observe a similar behavior as in the two previous cases. The average message delivery time first increases linearly as the separation between source and target nodes is small. It then converges to a constant value, and remains constant at this value for a broad range of source-target separation.

V. Conclusions

Small-world networks arise in many disciplines of science including biology, neuroscience, sociology and computer science. In this work, we have focused on the average delivery time of messages to a final destination in dense small-world networks when nodes use a local geographic forwarding rule to relay messages.

Existing work on small-world networks only provides bounds on this average message delivery time. On the other hand, in this paper, we have presented a technique based on the first-step analysis for calculating an exact formula of the message delivery time in small-world networks constructed on general measure-metric spaces. We then have applied our solution for the general measure-metric spaces to small-world networks constructed on the surface of a three dimensional sphere and on the surface of a circle. We have also obtained the average message delivery time for small-world networks constructed on the plane when source and target nodes are located sufficiently close to the center of the network domain so that no edge effects occur.

In all of the three cases, our analytical expressions are in line with the observation of Travers and Milgram in their famous experiment conducted in [7]. Messages are first forwarded along the long-range links when they are far away from the target node. Once they are in close proximity of the target node, they are delivered by means of local contacts.

Appendix A
Measure of Boundary Sets

In this appendix, we explore the measure of boundary sets of open balls in regular network spaces. To put forward this property formally, we define the following sets for any given $x \in D$ and $n \in \mathbb{N}$.

$$Z(x) = \{r > 0 : \lambda(\partial B(x,r)) > 0\},$$

and

$$Z_n(x) = \{r > 0 : n \cdot r \in Z(x)\}.$$

Theorem 14: Let $(D, \mathcal{S}_D, \lambda, \rho)$ be a regular network space, and

$$Z = \bigcup_{n=1}^{\infty} \bigcup_{x \in D} Z_n(x).$$

Then, Z contains at most countably many points.

Proof: We first show that $Z(x)$ contains at most countably many points for all $x \in D$. Suppose there exists x in D such that $Z(x)$ contains uncountably many points. Consider the conjugate set $Z^*(x)$ of $Z(x)$, which is defined as

$$Z^*(x) = \{\lambda(\partial B(x,r)) : r \in Z(x)\}. \quad (26)$$

By Property 3, we know that the boundary of an open ball centered at x with radius r is equal to

$$\partial B(x,r) = \{y \in D : \rho(x,y) = r\}.$$

For any $\lambda^* \in Z^*(x)$, there can be only finitely many distinct $r \in Z(x)$ such that $\lambda(\partial B(x,r)) = \lambda^*$. Otherwise, there exists an infinite countable collection $\{r_n\}_{n=1}^{\infty}$ of distinct radii such that $\lambda(\partial B(x,r_n)) = \lambda^*$ for all $n \geq 1$. Then,

$$\lambda\left(\bigcup_{n=1}^{\infty} \partial B(x,r_n)\right) = \sum_{n=1}^{\infty} \lambda(\partial B(x,r_n)) = \sum_{n=1}^{\infty} \lambda^* = \infty,$$

which contradicts the finiteness of the network space. Therefore, $Z^*(x)$ also contains uncountably many points. Let

$$\overline{\lambda^*} = \sup\{\lambda^* : \lambda^* \in Z^*(x)\}$$

and

$$\underline{\lambda^*} = \inf\{\lambda^* : \lambda^* \in Z^*(x)\}.$$

If $\underline{\lambda^*} > 0$, we can find an infinite countable collection $\{r_n\}_{n=1}^{\infty}$ of distinct radii such that $\lambda(\partial B(x,r_n)) \geq \underline{\lambda^*} > 0$ for all

$n \geq 1$. Thus, $\lambda\left(\bigcup_{n=1}^{\infty} \partial B(x, r_n)\right) = \infty$, which contradicts the finiteness of the network space. Assume then $\underline{\lambda^*} = 0$. Let $\epsilon = \overline{\lambda^*} - \underline{\lambda^*}$ and

$$A_k^* = Z^*(x) \bigcap \left[\underline{\lambda^*} + \frac{\epsilon}{k+1}, \overline{\lambda^*}\right].$$

Then, $Z^*(x) = \bigcup_{k=1}^{\infty} A_k^*$. Since this union contains uncountably many points, there must exist at least one K large enough such that A_K^* contains uncountably many points. Consider the conjugate of A_K^*:

$$A_K = \left\{r > 0 : \lambda\left(\partial B(x, r)\right) \in A_K^*\right\}.$$

Then, A_K contains uncountably many distinct points. Take a countably infinite collection $\{r_n\}_{n=1}^{\infty}$ of distinct radii in A_K. Then,

$$\begin{aligned} \lambda\left(\bigcup_{n=1}^{\infty} \partial B(x, r_n)\right) &= \sum_{n=1}^{\infty} \lambda\left(\partial B(x, r_n)\right) \\ &\geq \sum_{n=1}^{\infty} \left(\underline{\lambda^*} + \frac{\epsilon}{K+1}\right) = \infty, \end{aligned}$$

contradicts the finiteness of the network space. Thus, $Z(x)$ contains at most countably many points for all $x \in D$.

Secondly, we show that $Z(x)$ is the same for all $x \in D$. To this end, take any two distinct x and y in D, and take $r \in Z(x)$. Then, for any decreasing sequence $\{\epsilon_n\}_{n=1}^{\infty}$ with $\lim_{n\to\infty} \epsilon_n = 0$, we have

$$\begin{aligned} &\lambda\left(\partial B(x, r)\right) \\ &\quad = \lambda\left(\bigcap_{n=1}^{\infty} \left(B(x, r+\epsilon_n) - B(x, r-\epsilon_n)\right)\right) \\ &\quad \overset{(a)}{=} \lim_{n\to\infty} \lambda\left(B(x, r+\epsilon_n) - B(x, r-\epsilon_n)\right) \\ &\quad = \lim_{n\to\infty} \left(\lambda\left(B(x, r+\epsilon_n)\right) - \lambda\left(B(x, r-\epsilon_n)\right)\right) \\ &\quad \overset{(b)}{=} \lim_{n\to\infty} \left(\lambda\left(B(y, r+\epsilon_n)\right) - \lambda\left(B(y, r-\epsilon_n)\right)\right) \\ &\quad \overset{(c)}{=} \lambda\left(\bigcap_{n=1}^{\infty} \left(B(y, r+\epsilon_n) - B(y, r-\epsilon_n)\right)\right) \\ &\quad = \lambda(\partial B(y, r)), \end{aligned}$$

where (a) and (c) follow from the continuity of measure from above and (b) follows from the translation invariance property of the network space. Thus, $r \in Z(y)$. Similarly, any $r \in Z(y)$ also belongs to $Z(x)$. As a result, $Z(x)$ is the same for all $x \in D$. This further implies that for any given $n \geq 1$, $Z_n(x)$ is the same for all $x \in D$.

Finally, we show that for any given $x \in D$, $Z_n(x)$ is a countable set for all $n \geq 1$. The correspondence $f : Z_n(x) \to Z(x)$ such that $f(r) = n \cdot r$ is a bijection. Therefore, $Z_n(x)$ is a countable set for all $x \in D$ and $n \geq 1$.

Combining all of these observations, we conclude that

$$\begin{aligned} Z &= \bigcup_{n=1}^{\infty} \bigcup_{x\in D} Z_n(x) \\ &= \bigcup_{n=1}^{\infty} Z_n(x), \end{aligned} \tag{27}$$

which is a countable set. ■

Appendix B
Relation to Convex Metric Spaces

In this appendix, we investigate some fundamental relations between convex metric spaces and positive network spaces. A metric space (D, ρ) is called a convex metric space if for any given two distinct points x and y in D, there exists another point z in D different from x and y such that

$$\rho(x, y) = \rho(x, z) + \rho(z, y). \tag{28}$$

Any z satisfying (28) is said to be between x and y. We say a metric space (D, ρ) is loosely convex if for any given two distinct points x and y in D and $\epsilon > 0$, there exists z in D different from x and y such that

$$\rho(x, z) + \rho(z, y) - \epsilon \leq \rho(x, y). \tag{29}$$

Theorem 15: Let $(D, \mathcal{S}_D, \lambda, \rho)$ be a positive network space. Then, (D, ρ), when viewed as a metric space, is a loosely convex metric space. Furthermore, if (D, ρ) is a compact metric space, then it is a convex metric space.

Proof: For any given two distinct points x and y in D, and $\epsilon > 0$, choose $r_1 > 0$ and $r_2 > 0$ satisfying

$$r_1 + r_2 \in \left(\rho(x, y), \rho(x, y) + \epsilon\right).$$

By positivity of the network space, the intersection of $B(x, r_1)$ and $B(y, r_2)$ is not the empty set. Take a $z \in B(x, r_1) \bigcap B(y, r_2)$. Then,

$$\rho(x, z) + \rho(z, y) \leq r_1 + r_2 \leq \rho(x, y) + \epsilon,$$

proving the loose convexity of (D, ρ).

Now assume that (D, ρ) is a compact metric space. Take any positive and decreasing sequence $\{\epsilon_n\}_{n=1}^{\infty}$ such that $\epsilon_n \downarrow 0$ as $n \to \infty$. Let

$$A_n = \overline{B(x, \rho(x, y) - r)} \bigcap \overline{B(y, r + \epsilon_n)}.$$

By positivity of the network space, $A_n \neq \emptyset$ for all n. Note that A_n is a closed set for all n. Therefore, A_n is a compact set by compactness of (D, ρ). Observe also that

$$A_1 \supseteq A_2 \supseteq A_3 \supseteq \cdots.$$

Let $A = \bigcap_{n=1}^{\infty} A_n$. By the Cantor intersection theorem for compact sets, $A \neq \emptyset$. Let $x_0 \in A$. Then, $\rho(x_0, x) \leq \rho(x, y) - r$. It can be also shown that $\rho(x_0, y) \leq r$ by noting $\rho(x_0, y) \leq r + \epsilon_n$ for all n, and taking the limit as n goes to infinity.

We now claim that $\rho(x, x_0) = \rho(x, y) - r$ and $\rho(y, x_0) = r$. Otherwise,

$$\begin{aligned} \rho(x, y) &\leq \rho(x, x_0) + \rho(x_0, y) \text{ (triangle inequality)} \\ &< \rho(x, y) - r + r = \rho(x, y), \end{aligned}$$

which is a contradiction. ■

REFERENCES

[1] R. J. Williams and N. D. Martinez, "Simple rules yield complex food webs," *Nature*, vol. 404, no. 6774, pp. 180-183, 2000.

[2] R. Albert, H. Jeong and A.-L. Barabási, "Diameter of the World Wide Web," *Nature*, vol. 401, pp. 130-131, 1999.

[3] M. E. J. Newman, "The structure of scientific collaboration networks," *Proc. Natl. Acad. Sci. USA*, vol. 98, no. 2, pp. 404-409, 2001.

[4] J. Stiller, D. Nettle and R. I. M. Dunbar, "The small world of Shakespeare's plays," *Human Nature*, vol. 14, no. 4, pp. 397-408, 2003.

[5] F. Karinthy, *Chains - Everything is Different*, Atheneum Press, Budapest, 1929.

[6] S. Milgram, "The small world problem," *Psychology Today*, vol. 1, no.1, pp. 61-67, 1967.

[7] J. Travers and S. Milgram, "An experimental study of the small world problem," *Sociometry*, vol.32, no. 4, pp. 425-443, 1969.

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

[9] B. Bollobás, *Random Graphs*, Academic Press, London, 1985.

[10] D. Watts and S. H. Strogatz, "Collective dynamics of small-world networks,", *Nature*, vol. 393, no. 6684, pp. 440-442, 1998.

[11] S. H. Strogatz, "Exploring complex networks," *Nature*, vol. 410, no. 6825, pp. 268-276, 2001.

[12] J. M. Kleinberg, "Navigation in the small world," *Nature*, vol. 406, no. 6798, pp. 845, 2000.

[13] J. M. Kleinberg, "The small world phenomenon: an algorithmic perspective," *Proceedings of the 32nd Annual ACM Symposium on the Theory of Computing*, New York, 2000.

[14] M. Franceschetti and R. Meester, "Navigation in small-world networks, a scale-free continuum model," *Journal of Applied Probability*, vol. 43, no. 4, pp. 1173-1180, December 2006.

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

[16] R. G. Bartle, *The Elements of Integration and Lebesgue Measure*, John Wiley and Sons, Inc., New York, 1995.

[17] A. Okabe, B. Boots, K. Sugihara and S. N. Chiu, *Spatial Tessellations - Concepts and Applications of Voronoi Diagrams,* John Wiley and Sons, Inc., New York, second edition, 2000.

[18] P. Billingsley, *Probability and Measure*, John Wiley and Sons, Inc., New York, third edition, 1995.

[19] J. C. Oxtoby, "Homeomorphic measures in metric spaces," *Proceedings of the American Mathematical Society*, vol. 24, no. 3, pp. 419-423, March 1970.

[20] J. Stilwell, *Geometry of Surfaces*, Springer-Verlag, New York, 1992.