The Small-World Phenomenon: An Algorithmic Perspective

Jon Kleinberg *

Abstract

Long a matter of folklore, the "small-world phenomenon" — the principle that we are all linked by short chains of acquaintances — was inaugurated as an area of experimental study in the social sciences through the pioneering work of Stanley Milgram in the 1960's. This work was among the first to make the phenomenon quantitative, allowing people to speak of the "six degrees of separation" between any two people in the United States. Since then, a number of network models have been proposed as frameworks in which to study the problem analytically. One of the most refined of these models was formulated in recent work of Watts and Strogatz; their framework provided compelling evidence that the small-world phenomenon is pervasive in a range of networks arising in nature and technology, and a fundamental ingredient in the evolution of the World Wide Web.

But existing models are insufficient to explain the striking algorithmic component of Milgram's original findings: that individuals using local information are collectively very effective at actually *constructing* short paths between two points in a social network. Although recently proposed network models are rich in short paths, we prove that no *decentralized* algorithm, operating with local information only, can construct short paths in these networks with non-negligible probability. We then define an infinite family of network models that naturally generalizes the Watts-Strogatz model, and show that for one of these models, there is a decentralized algorithm capable of finding short paths with high probability. More generally, we provide a strong characterization of this family of network models, showing that there is in fact a *unique* model within the family for which decentralized algorithms are effective.

*Department of Computer Science, Cornell University, Ithaca NY 14853. Email: kleinber@cs.cornell.edu.
Supported in part by an Alfred P. Sloan Research Fellowship, an ONR Young Investigator Award, and NSF Faculty Early Career Development Award CCR-9701399.

1 Introduction

The Small-World Phenomenon. A social network exhibits the *small-world phenomenon* if, roughly speaking, any two individuals in the network are likely to be connected through a short sequence of intermediate acquaintances. This has long been the subject of anecdotal observation and folklore; often we meet a stranger and discover that we have an acquaintance in common. It has since grown into a significant area of study in the social sciences, in large part through a series of striking experiments conducted by Stanley Milgram and his co-workers in the 1960's [10, 14, 9]. Recent work has suggested that the phenomenon is pervasive in networks arising in nature and technology, and a fundamental ingredient in the structural evolution of the World Wide Web [13, 15, 2].

Milgram's basic small-world experiment remains one of the most compelling ways to think about the problem. The goal of the experiment was to find short chains of acquaintances linking pairs of people in the United States who did not know one another. In a typical run of the experiment, a *source* person in Nebraska would be given a letter to deliver to a *target* person in Massachusetts. The source would initially be told basic information about the target, including his address and occupation; the source would then be instructed to send the letter to someone she knew *on a first-name basis* in an effort to transmit the letter to the target as efficaciously as possible. Anyone subsequently receiving the letter would be given the same instructions, and the chain of communication would continue until the target was reached. Over many trials, the average number of intermediate steps in a successful chain was found to lie between five and six, a quantity that has since entered popular culture as the "six degrees of separation" principle [4].

Modeling the Phenomenon. Naturally, the empirical validation of the phenomenon has led to a rash of analytical work aimed at answering the following general question:

> ($*$) Why should there *exist* short chains of acquaintances linking together arbitrary pairs of strangers?

Most of the early work on this issue, beginning with analysis of Pool and Kochen that pre-dated Milgram's experiments [12], was based on versions of the following explanation: random networks have low diameter. (See for example the book of surveys edited by Kochen [8].) That is, if every individual in the United States were to have a small number of acquaintances selected uniformly at random from the population — and if acquaintanceship were symmetric — then two random individuals would be linked by a short chain with high probability. Even this early work recognized the limitations of a uniform random model; if A and B are two individuals with a common friend, it is much more likely that they themselves are friends. But at the same time, a network of acquaintanceships that is too "clustered" will not have the low diameter that Milgram's experiments indicated.

Recently, Watts and Strogatz proposed a model for random networks that interpolates between these two extremes, by dividing the edges of the network into "local" and "long-range" contacts [15]. The paradigmatic example they studied was a "re-wired ring lattice," constructed roughly as follows. One starts with a set V of n points spaced uniformly on a circle, and joins each point by an edge to each of its k nearest neighbors, for a small constant

k. These are the "local contacts" in the network. One then introduces a small number of edges in which the endpoints are chosen uniformly at random from V — the "long-range contacts". Watts and Strogatz argued that such a model captures two crucial parameters of social networks: there is a simple underlying structure that explains the presence of most edges, but a few edges are produced by a random process that does not respect this structure. Their networks thus have low diameter (like uniform random networks), but also have the property that many of the neighbors of a node u are themselves neighbors (unlike uniform random networks). They showed that a number of naturally arising networks exhibit this pair of properties (including the connections among neurons in the nematode species *C. elegans*, and the power grid of the Western U.S.); and their approach has been applied to the analysis of the hyperlink graph of the World Wide Web as well [1].

The Present Work. Let us return to Milgram's experiment. We claim that it really contains two fundamentally surprising discoveries: first, that such short chains should *exist* in the network of acquaintanceships; and second, that people should be able to *find* these chains knowing so little about the target individual. From an analytical point of view, the first of these discoveries is existential in nature, the second *algorithmic* — it reveals that individuals who know only the locations of their direct acquaintances can still, collectively, construct a short path between two points in the network. We therefore propose to study the following natural companion to Question $(*)$ above:

> $(**)$ Why should arbitrary pairs of strangers be able to *find* short chains of acquaintances that link them together?

It is important to note that Question $(**)$ raises issues that lie truly beyond the scope of Question $(*)$: one could imagine a world in which short chains exist but individuals, faced with a disorienting array of social contacts, are unable to find them. It is therefore natural to ask what properties a social network must possess in order for its members to be able to find short chains through it.

In this work, we study "decentralized" algorithms by which individuals, knowing only the locations of their direct acquaintances, attempt to transmit a message from a source to a target along a short path. Our central findings are the following.

- First, we show that existing models are insufficient to explain the success of such decentralized algorithms in finding short paths through a social network. In a class of networks generated according to the model of Watts and Strogatz, we prove that there is no decentralized algorithm capable of constructing paths of small expected length (relative to the diameter of the underlying network).

- We then define an infinite family of random network models that naturally generalizes the Watts-Strogatz model. We show that for one of these models, there is a decentralized algorithm capable of finding short paths with high probability.

- Finally, we prove the stronger statement that there is in fact a *unique* model within the family for which decentralized algorithms are effective.

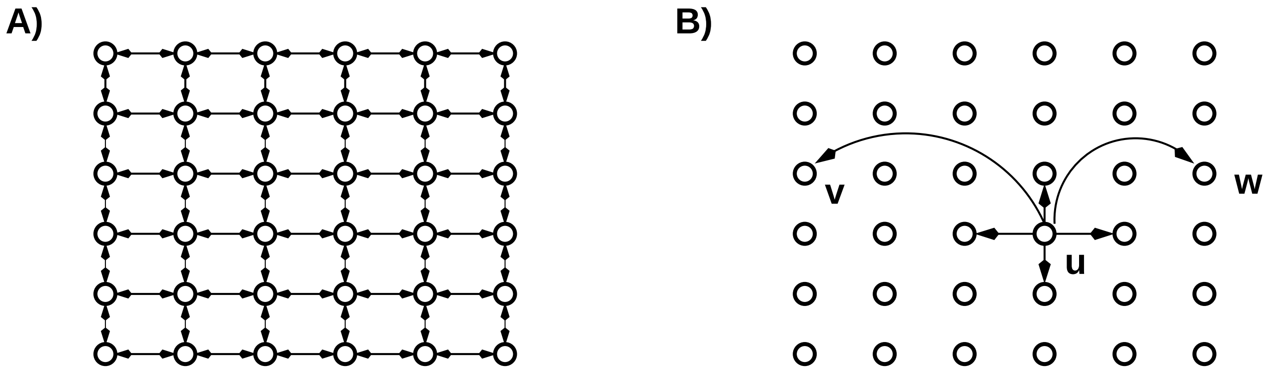

Figure 1: (A) A two-dimensional grid network with $n = 6$, $p = 1$, and $q = 0$. (B) The contacts of a node u with $p = 1$ and $q = 2$. v and w are the two long-range contacts.

The Model: Networks and Decentralized Algorithms. We now give precise definitions for our network model and our notion of a decentralized algorithm; we then provide formal statements of the main results.

In designing our network model, we seek a simple framework that encapsulates the paradigm of Watts and Strogatz — rich in local connections, with a few long-range connections. Rather than using a ring as the basic structure, however, we begin from a two-dimensional grid and allow for edges to be *directed*. Thus, we begin with a set of nodes (representing individuals in the social network) that are identified with the set of lattice points in an $n \times n$ square, $\{(i, j) : i \in \{1, 2, \ldots, n\}, j \in \{1, 2, \ldots, n\}\}$, and we define the *lattice distance* between two nodes (i, j) and (k, ℓ) to be the number of "lattice steps" separating them: $D((i, j), (k, \ell)) = |k - i| + |\ell - j|$. For a universal constant $p \geq 1$, the node u has a directed edge to every other node within lattice distance p — these are its *local contacts*. For universal constants $q \geq 0$ and $r \geq 0$, we also construct directed edges from u to q other nodes (the *long-range contacts*) using independent random trials; the i^{th} directed edge from u has endpoint v with probability proportional to $[D(u, v)]^{-r}$. (To obtain a probability distribution, we divide this quantity by the appropriate normalizing constant $\sum_v [D(u, v)]^{-r}$; we will call this the *inverse* r^{th}*-power distribution*.)

This model has a simple "geographic" interpretation: individuals live on a grid and know their neighbors for some number of steps in all directions; they also have some number of acquaintances distributed more broadly across the grid. Viewing p and q as fixed constants, we obtain a one-parameter family of network models by tuning the value of the exponent r. When $r = 0$, we have the uniform distribution over long-range contacts, the distribution used in the basic network model of Watts and Strogatz — one's long-range contacts are chosen independently of their position on the grid. As r increases, the long-range contacts of a node become more and more clustered in its vicinity on the grid. Thus, r serves as a basic structural parameter measuring how widely "networked" the underlying society of nodes is.

The algorithmic component of the model is based on Milgram's experiment. We start with two arbitrary nodes in the network, denoted s and t; the goal is to transmit a message from s to t in as few steps as possible. We study *decentralized algorithms*, mechanisms whereby the message is passed sequentially from a current message holder to one of its (local or long-range) contacts, using only local information. In particular, the message holder u in a given step has knowledge of

(i) the set of local contacts among all nodes (i.e. the underlying grid structure);
(ii) the location, on the lattice, of the target t; and
(iii) the locations and long-range contacts of all nodes that have come in contact with the message.

Crucially, u does not have knowledge of the long-range contacts of nodes that have not touched the message. Given this, u must choose one of its contacts v, and forward the message to this contact. The *expected delivery time* of a decentralized algorithm — a primary figure of merit in our analysis — is the expected number of steps taken by the algorithm to deliver the message over a network generated according to an inverse r^{th}-power distribution, from a source to a target chosen uniformly at random from the set of nodes. Of course, constraining the algorithm to use only local information is crucial to our model; if one had full global knowledge of the local and long-range contacts of all nodes in the network, the shortest chain between two nodes could be computed simply by breadth-first search.

The reader may worry that assumption (iii) above gives a decentralized algorithm too much power. However, it only strengthens our results: our lower bounds will hold even for algorithms that are given this knowledge, while our upper bounds make use of decentralized algorithms that only require assumptions (i) and (ii).

Statement of Results. Our results explore the way in which the structure of the network affects the ability of a decentralized algorithm to construct a short path.

When $r = 0$ — the uniform distribution over long-range contacts — standard results from random graph theory can be used to show that with high probability there *exist* paths between every pair of nodes whose lengths are bounded by a polynomial in $\log n$, exponentially smaller than the total number of nodes. However, there is no no way for a decentralized algorithm to find these chains:

Theorem 1 *There is a constant α_0, depending on p and q but independent of n, so that when $r = 0$, the expected delivery time of any decentralized algorithm is at least $\alpha_0 n^{2/3}$. (Hence exponential in the expected minimum path length.)*

As the parameter r increases, a decentralized algorithm can take more advantage of the "geographic structure" implicit in the long-range contacts; at the same time, long-range contacts become less useful in moving the message a large distance. There is a value of r where this trade-off can be best exploited algorithmically; this is $r = 2$, the inverse-square distribution.

Theorem 2 *There is a decentralized algorithm $\mathcal{A}$ and a constant α_2, independent of n, so that when $r = 2$ and $p = q = 1$, the expected delivery time of $\mathcal{A}$ is at most $\alpha_2(\log n)^2$.*

This pair of theorems reflects a fundamental consequence of our model. When long-range contacts are formed independently of the geometry of the grid, short chains will exist but the nodes, operating at a local level, will not be able to find them. When long-range contacts are formed by a process that is related to the geometry of the grid in a specific way, however, then short chains will still form *and* nodes operating with local knowledge will be able to construct them.

We now comment on the ideas underlying the proofs of these results; the full details are given in the subsequent sections. The decentralized algorithm $\mathcal{A}$ that achieves the bound of Theorem 2 is the following simple rule: in each step, the current message-holder u chooses a contact that is as close to the target t as possible, in the sense of lattice distance. Note that algorithm $\mathcal{A}$ makes use of even less information than is allowed by our general model: the current message holder does not need to know anything about the set of previous message holders. To analyze an execution of algorithm $\mathcal{A}$, we say that it is in *phase* j if the lattice distance from the current message holder to the target is between 2^j and 2^{j+1}. We show that in phase j, the expected time before the current message holder has a long-range contact within lattice distance 2^j of t is bounded proportionally to $\log n$; at this point, phase j will come to an end. As there are at most $1 + \log n$ phases, a bound proportional to $(\log n)^2$ follows. Interestingly, the analysis matches our intuition, and Milgram's description, of how a short chain is found in real life: "The geographic movement of the [message] from Nebraska to Massachusetts is striking. There is a progressive closing in on the target area as each new person is added to the chain" [10].

The impossibility result of Theorem 1 is based, fundamentally, on the fact that the uniform distribution prevents a decentralized algorithm from using any "clues" provided by the geometry of the grid. Roughly, we consider the set U of all nodes within lattice distance $n^{2/3}$ of t. With high probability, the source s will lie outside of U, and if the message is never passed from a node to a long-range contact in U, the number of steps needed to reach t will be at least proportional to $n^{2/3}$. But the probability that *any* message holder has a long-range contact in U is roughly $n^{-2/3}$, so the expected number of steps before a long-range contact in U is found is at least proportional to $n^{2/3}$ as well.

More generally, we can show a strong characterization theorem for this family of models: $r = 2$ is the *only* value for which there is a decentralized algorithm capable of producing chains whose length is a polynomial in $\log n$:

Theorem 3 *(a) Let $0 \leq r < 2$. There is a constant α_r, depending on p, q, r, but independent of n, so that the expected delivery time of any decentralized algorithm is at least $\alpha_r n^{(2-r)/3}$.*

(b) Let $r > 2$. There is a constant α_r, depending on p, q, r, but independent of n, so that the expected delivery time of any decentralized algorithm is at least $\alpha_r n^{(r-2)/(r-1)}$.

The complete proof of this theorem is given in Section 3. The proof of (a) is analogous to that of Theorem 1. The proof of (b), on the other hand, exposes a "dual" obstacle for decentralized algorithms: with a large value of r, it takes a significant amount of time before the message reaches a node with a long-range contact that is far away in lattice distance. This effectively limits the "speed" at which the message can travel from s to t.

Although we have focused on the two-dimensional grid, our analysis can be applied more broadly. We can generalize our results to k-dimensional lattice networks, as well as slightly "irregular" lattices; in the k-dimensional case, a decentralized algorithm can construct paths of length polynomial in $\log n$ if and only if $r = k$. Our results suggest a general model for

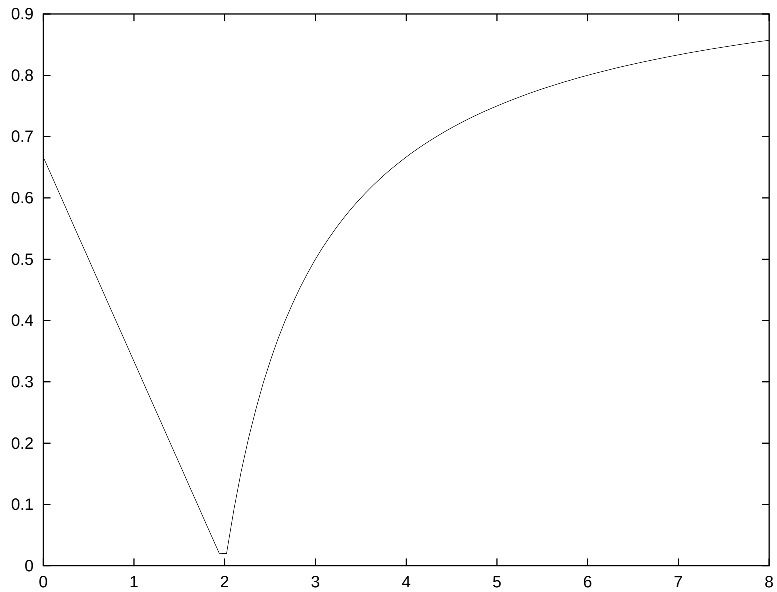

Figure 2: The lower bound implied by Theorem 3. The x-axis is the value of r; the y-axis is the resulting exponent on n.

the crucial role that geographic cues play in helping individuals make connections through a complex social network, capturing phenomena observed in empirical studies [10, 7]. We believe it would be interesting to extend these models to incorporate other types of social cues that arise in these settings, and more generally to consider other ways in which an algorithmic perspective can help provide insight into processes that shape and take place within large social networks.

Other Related Work. There has been work aimed at modeling the way in which individuals in Milgram's experiments chose recipients for their letters. Some of this work is related in spirit to what we do here, though using very different perspectives and models. Killworth and Bernard [7], in their "reverse small-world experiments," asked a set of respondents to explain how they chose to send letters in a run of the small-world experiment, and used this information to look for common principles at an empirical level. At an analytical level, White [16] investigated the probability that a chain would "die out" through an individual's failure to participate, and Hunter and Shotland [5] studied the passage of a chain through different social "categories." In the context of a "referral system" for the World Wide Web, Kautz, Selman, and Shah [13] ran simulations of communication in an abstract social network in which each individual was given pre-defined *accuracy* and *responsiveness* parameters. The distinction between the mere existence of short paths linking points on the World Wide Web, and the ability of agents to find them, has also been raised recently in work of Albert, Jeong, and Barabasi [2, 6].

2 Upper Bound for the Inverse-Square Distribution

We now present proofs of the theorems discussed in the previous section. When we analyze a decentralized algorithm, we can adopt the following equivalent formulation of the model, which will make the exposition easier. Although our model considers all long-range contacts as being generated initially, at random, we invoke the "Principle of Deferred Decisions" — a common mechanism for analyzing randomized algorithms [11] — and assume that the long-range contacts of a node v are generated only when the message first reaches v. Since a decentralized algorithm does not learn the long-range contacts of v until the message reaches v, this formulation is equivalent for the purposes of analysis.

A comment on the notation: $\log n$ denotes the logarithm base 2, while $\ln n$ denotes the natural logarithm, base e.

Proof of Theorem 2. Since $p = q = 1$, we have a network in which each node u is connected to its four nearest neighbors in the lattice (two or three neighbors in the case of nodes on the boundary), and has a single long-range contact v. The probability that u chooses v as its long-range contact is $d(u,v)^{-2}/\sum_{v \neq u} d(u,v)^{-2}$, and we have

$$\begin{aligned}
\sum_{v \neq u} d(u,v)^{-2} &\leq \sum_{j=1}^{2n-2} (4j)(j^{-2}) \\
&= 4 \sum_{j=1}^{2n-2} j^{-1} \\
&\leq 4 + 4\ln(2n-2) \leq 4\ln(6n).
\end{aligned}$$

Thus, the probability that v is chosen is at least $[4\ln(6n)d(u,v)^2]^{-1}$.

The decentralized algorithm $\mathcal{A}$ is defined as follows: in each step, the current message-holder u chooses a contact that is as close to the target t as possible, in the sense of lattice distance. For $j > 0$, we say that the execution of $\mathcal{A}$ is in *phase* j when the lattice distance from the current node to t is greater than 2^j and at most 2^{j+1}. We say $\mathcal{A}$ is in phase 0 when the lattice distance to t is at most 2. Thus, the initial value of j is at most $\log n$. Now, because the distance from the message to the target decreases strictly in each step, each node that becomes the message holder has not touched the message before; thus, we may assume that the long-range contact from the message holder is generated at this moment.

Suppose we are in phase j, $\log(\log n) \leq j < \log n$, and the current message holder is u. What is the probability that phase j will end in this step? This requires the message to enter the set B_j of nodes within lattice distance 2^j of t. There are at least

$$1 + \sum_{i=1}^{2^j} i = \frac{1}{2}2^{2j} + \frac{1}{2}2^j + 1 > 2^{2j-1}$$

nodes in B_j, each is within lattice distance $2^{j+1} + 2^j < 2^{j+2}$ of u, and hence each has a probability of at least $(4\ln(6n)2^{2j+4})^{-1}$ of being the long-range contact of u. If any of these nodes is the long-range contact of u, it will be u's closest neighbor to t; thus the message enters B_j with probability at least

$$\frac{2^{2j-1}}{4\ln(6n)2^{2j+4}} = \frac{1}{128\ln(6n)}.$$

Let X_j denote total number of steps spent in phase j, $\log(\log n) \le j \le \log n$. We have

$$\begin{aligned} EX_j &= \sum_{i=1}^{\infty} \Pr\left[X_j \ge i\right] \\ &\le \left(1 - \frac{1}{128\ln(6n)}\right)^{i-1} \\ &= 128\ln(6n). \end{aligned}$$

An analogous set of bounds shows that $EX_j \le 128\ln(6n)$ for $j = \log n$ as well. Finally, if $0 \le j \le \log(\log n)$, then $EX_j \le 128\ln(6n)$ holds for the simple reason that the algorithm can spend at most $\log n$ steps in phase j even if all nodes pass the message to a local contact.

Now, if X denotes the total number of steps spent by the algorithm, we have

$$X = \sum_{j=0}^{\log n} X_j,$$

and so by linearity of expectation we have $EX \le (1 + \log n)(128\ln(6n)) \le \alpha_2(\log n)^2$ for a suitable choice of α_2.

3 Lower Bounds for Other Distributions

We first expand our model of a decentralized algorithm slightly; it will correspondingly strengthen the result to show a lower bound for this new model. An algorithm initially has knowledge of the grid structure, all the local contacts, and the locations of s and t. In step i, some set S_i of nodes has touched the message. At this point, the algorithm has knowledge of all long-range contacts of all nodes in S_i. (Following our style of analysis, the long-range contacts of other nodes will be constructed only as the message reaches them.) Based on this information, it chooses any contact v of any node in S_i that has not yet received the message — v need not be a contact of the current message holder — and it sends the message to v. The set S_{i+1} thus contains one element more than S_i, and the algorithm iterates. This is the same as our initial model of a decentralized algorithm, except that we do not count steps in which the algorithm "backtracks" by sending the message through a node that has already received it.

For technical reasons, we will add one additional feature to the algorithms we consider. An algorithm will run for an infinite sequence of steps; initially it behaves as above, and once the message reaches t, the message remains at t in all subsequent steps. Thus, when we consider the i^{th} step of a given algorithm, we need not worry that it has already terminated by this step.

We now prove the two parts of Theorem 3; note that part (a) implies Theorem 1 by setting $r = 0$. As in Section 2, we will invoke the Principle of Deferred Decisions [11] in the analysis.

Proof of Theorem 3a. We consider an arbitrary decentralized algorithm of the type described above, and consider the expected number of steps required for the message to travel from s to t, for nodes s and t generated uniformly at random from the grid.

Note that because we have the freedom to choose the constant α_r, we may also assume that n is at least as large as some fixed absolute constant n_0. The probability that a node u chooses v as its i^{th} out of q long-range contacts is $d(u,v)^{-r}/\sum_{v\neq u} d(u,v)^{-r}$, and we have

$$\begin{aligned}
\sum_{v\neq u} d(u,v)^{-r} &\geq \sum_{j=1}^{n/2} (j)(j^{-r}) \\
&= \sum_{j=1}^{n/2} j^{1-r} \\
&\geq \int_1^{n/2} x^{1-r}\, dx \\
&\geq (2-r)^{-1}((n/2)^{2-r}-1) \\
&\geq \frac{1}{(2-r)2^{3-r}} \cdot n^{2-r},
\end{aligned}$$

where the last line follows if we assume $n \geq 2^{3-r}$. Let $\delta = (2-r)/3$.

Let U denote the set of nodes within lattice distance pn^{δ} of t. Note that

$$|U| \leq 1 + \sum_{j=1}^{pn^{\delta}} 4j \leq 4p^2 n^{2\delta},$$

where we assume n is large enough that $pn^{\delta} \geq 2$. Define $\lambda = ((2-r)2^{7-r}qp^2)^{-1}$. Let $\mathcal{E}'$ be the event that within λn^{δ} steps, the message reaches a node other than t with a long-range contact in U. Let $\mathcal{E}'_i$ be the event that in step i, the message reaches a node other than t with a long-range contact in U; thus $\mathcal{E}' = \cup_{i\leq\lambda n^{\delta}} \mathcal{E}'_i$. Now, the node reached at step i has q long-range contacts that are generated at random when it is encountered; so we have

$$\Pr\left[\mathcal{E}'_i\right] \leq \frac{q|U|}{\frac{1}{(2-r)2^{3-r}} \cdot n^{2-r}} \leq \frac{(2-r)2^{3-r}q \cdot 4p^2 n^{2\delta}}{n^{2-r}} = \frac{(2-r)2^{5-r}qp^2 n^{2\delta}}{n^{2-r}}.$$

Since the probability of a union of events is bounded by the sum of their probabilities,

$$\begin{aligned}
\Pr\left[\mathcal{E}'\right] &\leq \sum_{i\leq\lambda n^{\delta}} \Pr\left[\mathcal{E}'_i\right] \\
&\leq \frac{(2-r)2^{5-r}\lambda qp^2 n^{3\delta}}{n^{2-r}} \\
&= (2-r)2^{5-r}\lambda qp^2 \leq \frac{1}{4}.
\end{aligned}$$

We now define two further events. Let $\mathcal{F}$ denote the event that the chosen source s and target t are separated by a lattice distance of at least $n/4$. One can verify that $\Pr\left[\mathcal{F}\right] \geq \frac{1}{2}$. Since $\Pr\left[\overline{\mathcal{F}} \vee \mathcal{E}'\right] \leq \frac{1}{2} + \frac{1}{4}$, $\Pr\left[\mathcal{F} \wedge \overline{\mathcal{E}'}\right] \geq \frac{1}{4}$.

Finally, let X denote the random variable equal to the number of steps taken for the message to reach t, and let $\mathcal{E}$ denote the event that the message reaches t within λn^δ steps. We claim that if $\mathcal{F}$ occurs and $\mathcal{E}'$ does not occur, then $\mathcal{E}$ cannot occur. For suppose it does. Since $D(s,t) \geq n/4 > p\lambda n^\delta$, in any s-t path of at most λn^δ steps, the message must be passed at least once from a node to a long-range contact. Moreover, the final time this happens, the long-range contact must lie in U. This contradicts our assumption that $\mathcal{E}'$ does not occur.

Thus, $\Pr\left[\mathcal{E} \mid \mathcal{F} \wedge \overline{\mathcal{E}'}\right] = 0$, hence $E\left[X \mid \mathcal{F} \wedge \overline{\mathcal{E}'}\right] \geq \lambda n^\delta$. Since

$$EX \geq E\left[X \mid \mathcal{F} \wedge \overline{\mathcal{E}'}\right] \cdot \Pr\left[\mathcal{F} \wedge \overline{\mathcal{E}'}\right] \geq \frac{1}{4}\lambda n^\delta,$$

part (a) of the theorem follows.

Proof of Theorem 3b. We now turn to part (b) of the theorem, when $r > 2$. Again we consider an arbitrary decentralized algorithm; and again, as necessary, we may assume that n is larger than some fixed absolute constant n_0. We write $\varepsilon = r - 2$. Consider a node u, and let v be a randomly generated long-range contact of v. The normalizing constant for the inverse r^{th}-power distribution is at least 1, and so for any m, we have

$$\begin{aligned}
\Pr\left[D(u,v) > m\right] &\leq \sum_{j=m+1}^{2n-2} (4j)(j^{-r}) \\
&= 4 \sum_{j=m+1}^{2n-2} j^{1-r} \\
&\leq \int_m^\infty x^{1-r}\, dx \\
&\leq (r-2)^{-1} m^{2-r} = \varepsilon^{-1} m^{-\varepsilon}.
\end{aligned}$$

We set $\beta = \frac{\varepsilon}{1+\varepsilon}$, $\gamma = \frac{1}{1+\varepsilon}$, and $\lambda' = \frac{\min(\varepsilon,1)}{8q}$. We will assume n has been chosen large enough that $n^\gamma \geq p$. Let $\mathcal{E}'_i$ be the event that in step i, the message reaches a node $u \neq t$ that has a long-range contact v satisfying $D(u,v) > n^\gamma$. Let $\mathcal{E}' = \cup_{i \leq \lambda' n^\beta} \mathcal{E}'_i$ be the event that this happens in the first $\lambda' n^\beta$ steps. We have

$$\begin{aligned}
\Pr\left[\mathcal{E}'\right] &\leq \sum_{i \leq \lambda' n^\beta} \Pr\left[\mathcal{E}_i\right] \\
&\leq \lambda' n^\beta \cdot q\varepsilon^{-1} n^{-\varepsilon\gamma} \\
&= \lambda' q \varepsilon^{-1} \leq \frac{1}{4}.
\end{aligned}$$

As in part (a), we define $\mathcal{F}$ to be the event that s and t are separated by a lattice distance of at least $n/4$. Observe that $\Pr\left[\mathcal{F} \wedge \overline{\mathcal{E}'}\right] \geq \frac{1}{4}$. Let X denote the random variable equal to the number of steps taken for the message to reach t, and let $\mathcal{E}$ denote the event that the message reaches t within $\lambda' n^\beta$ steps. We claim that if $\mathcal{F}$ occurs and $\mathcal{E}'$ does not occur, then $\mathcal{E}$ cannot occur. For if $\mathcal{E}'$ does not occur, then the message can move a lattice distance of at most n^γ in each of its first $\lambda' n^\beta$ steps. This is a total lattice distance of at most

$$\lambda' n^{\beta+\gamma} = \lambda' n < n/4,$$

and so the message will not reach t given that $\mathcal{F}$ occurs.

Thus $E\left[X \mid \mathcal{F} \wedge \overline{\mathcal{E}'}\right] \geq \lambda' n^{\beta}$. Since

$$EX \geq E\left[X \mid \mathcal{F} \wedge \overline{\mathcal{E}'}\right] \cdot \Pr\left[\mathcal{F} \wedge \overline{\mathcal{E}'}\right] \geq \frac{1}{4}\lambda' n^{\beta},$$

part (b) of the theorem follows.

References

[1] L. Adamic, "The small world Web," manuscript, available at http://www.parc.xerox.com/istl/groups/iea/www/smallworld.html.

[2] R. Albert, H. Jeong, A.-L. Barabasi, "The diameter of the World Wide Web," *Nature* 401, 130 (1999).

[3] B. Bollobas, *Random Graphs* (Academic Press, London, 1985).

[4] J. Guare, *Six Degrees of Separation: A Play* (Vintage Books, New York, 1990).

[5] J. Hunter and R. Shotland, "Treating data collected by the small world method as a Markov process," *Social Forces* 52, 321 (1974).

[6] J. Kaiser, Ed., "It's a small Web after all," *Science* 285, 1815 (1999).

[7] P. Killworth and H. Bernard, "Reverse small world experiment," *Social Networks* 1, 159 (1978).

[8] M. Kochen, Ed., *The Small World* (Ablex, Norwood, 1989).

[9] C. Korte and S. Milgram, "Acquaintance networks between racial groups: Application of the small world method," *J. Personality and Social Psych.*, 15, 101 (1978).

[10] S. Milgram, "The small world problem," *Psychology Today* 1, 61 (1967).

[11] R. Motwani and P. Raghavan, *Randomized Algorithms* (Cambridge University Press, Cambridge, 1995).

[12] I. de Sola Pool and M. Kochen, "Contacts and influence," *Social Networks* 1, 5 (1978).

[13] H. Kautz, B. Selman, M. Shah, "ReferralWeb: Combining Social Networks and Collaborative Filtering," *Communications of the ACM*, 30, 3 (March 1997).

[14] J. Travers and S. Milgram, "An experimental study of the small world problem," *Sociometry* 32, 425 (1969).

[15] D. Watts and S. Strogatz, "Collective dynamics of small-world networks," *Nature* 393, 440 (1998).

[16] H. White, "Search parameters for the small world problem," *Social Forces* 49, 259 (1970).