

How Do Networks Become Navigable?

Aaron Clauset* and Christopher Moore†,*

*Computer Science Department and †Department of Physics and Astronomy,
University of New Mexico, Albuquerque NM 87131

{aaron,moore}@cs.unm.edu

(Dated: October 13, 2003)

Networks created and maintained by social processes, such as the human friendship network and the World Wide Web, appear to exhibit the property of *navigability*: namely, not only do short paths exist between any pair of nodes, but such paths can easily be found using only local information. It has been shown that for networks with an underlying metric, algorithms using only local information perform extremely well if there is a power-law distribution of link lengths. However, it is not clear why or how real networks might develop this distribution. In this paper we define a decentralized “rewiring” process, inspired by surfers on the Web, in which each surfer attempts to travel from their home page to a random destination, and updates the outgoing link from their home page if this journey takes too long. We show that this process does indeed cause the link length distribution to converge to a power law, achieving a routing time of $O(\log^2 n)$ on networks of size n . We also study finite-size effects on the optimal exponent, and show that it converges polylogarithmically slowly as the lattice size goes to infinity.

I. INTRODUCTION

It is well-known that a wide variety of social networks have the so-called *small world* property; networks of size n have diameter $O(\log n)$, meaning that between any two nodes there exists a path of size $O(\log n)$. This property is also shared by sparse random graphs; however, random graphs lack several other important properties of real-world networks, such as clustering and heavy-tailed degree distributions (see [1] for a review).

Another property of social networks which distinguishes them from random graphs, and which has received somewhat less attention, is their *navigability*. That is, not only do short paths exist, but it is easy to find them using only local information. The pioneering work of Milgram in the 1960s [2] showed that people can, at least some of the time, find short paths to other distant people by pursuing the greedy strategy where they pass the message to whichever of their own acquaintances they feel is “closest” to the target recipient. This experiment has recently been repeated using e-mail networks [3]. In a random graph, although a short path exists, a local algorithm must be lucky to find it as it can do little better than a random walk on the network.

Kleinberg [4, 5] addressed the issue of navigability by considering a small-world model consisting of a d -dimensional lattice with long-range links added to it. Unlike the Watts-Strogatz model [6] in which the long-range links are uniformly distributed, in the Kleinberg model pairs of nodes a distance ℓ apart are connected with some probability $P(\ell)$, where $P(\ell) \sim \ell^{-\alpha}$ and the number of outgoing links per node is fixed. While a finite-dimensional lattice is obviously a gross oversimplification of the social spaces in which we live, or the conceptual spaces in which one Web page seems “closer” to another, this model captures the essential features of the navigation problem: how can we minimize the routing time given local information about some metric?

Kleinberg studied the performance of the greedy algorithm, in which each node passes the message to whichever of its neighbors, either local or long-range, that is closest to the destination on the underlying lattice. He showed that if $\alpha = d$ this algorithm achieves a routing time of $O(\log^2 n)$ on lattices of size n . However, if $\alpha \neq d$, the routing time grows as $O(n^\beta)$ for some $\beta > 0$. Note that if when $\alpha = d$, we integrate over spheres of radius ℓ , we find that the distribution of link lengths is $f(\ell) \sim \ell^{-1}$, with a cutoff when ℓ reaches the system size; this distribution provides the right mix of long-, medium-, and short-range links for the greedy algorithm to quickly “zero in” on its destination. Similar results have been obtained for the case where the underlying graph is a tree, representing a hierarchical structure, or an overlapping set of trees representing multiple affiliations in society [7, 8].

While Kleinberg’s work largely characterizes the *static* properties a network needs to have to be navigable, it leaves open the question of how or why networks might evolve these properties. Indeed, the topology of a social network is constantly being modified by its members. If, whenever these members find it difficult to search the network, they modify their own connections in an effort to make future searches easier, this dynamical process should make the network more navigable over time.

In this paper, we consider a dynamic network model inspired by surfers on the Web, each of whom controls the outgoing links from their home page. We start at a source node x , and choose a random destination node y . Based on the (metric) distance between x and y , we set a threshold T_{thresh} on the number of (topological) steps we feel the journey should take. If after T_{thresh} steps we have not reached our destination, we stop searching, and rewire x ’s long-range link to the place where we gave up. In the human-friendship network, this might correspond to gaining new acquaintances in the course of a search; on the Web, it corresponds to creating a bookmark/linking to the relevant pages, which other surfers can then use [9].

Our main result is that this process does indeed cause the distribution of link lengths to converge to a power law $f(\ell) \sim \ell^{-\alpha_{\text{rewired}}}$. The precise value of α_{rewired} varies with lattice size, and differs somewhat from Kleinberg’s prediction $\alpha = d$. We believe this is due to finite-size effects; to support this belief, we directly construct networks with power-law link length distributions (as opposed to rewiring them), measure the exponent α_{opt} that minimizes the routing time, and find that it converges rather slowly to d as $n \rightarrow \infty$, roughly as $d - O(1/\log^2 n)$. However, even though the exponents differ somewhat, our rewiring process produces networks whose routing times match or improve those of Kleinberg’s optimum.

Other types of dynamical models have been studied (e.g. [11, 12, 13, 14, 15]). However, the model discussed here appears to be the first to optimize for navigability.

II. THE REWIRING PROCESS AND RESULTS

One restriction of Kleinberg’s model of a social space consists of a d -dimensional lattice of size n , where each node u is connected to its $2d$ nearest neighbors and has a single long-range link to a node v , chosen with probability $P(\ell) \sim \ell^{-\alpha}$ where $\ell = |u - v|$ is the Manhattan distance (the L_1 norm). Such networks are denoted $G(n, d, 1, 1, \alpha)$ in [4]; here the two 1s indicate the radius of the local connections and the number of long-range links per node. Our goal is to show that our rewiring process causes networks with a range of initial link length distributions to converge to this form, where $\alpha = \alpha_{\text{rewired}}$ and is close to Kleinberg’s prediction $\alpha = d$. We report here on experiments where $d = 1$, for which we can feasibly study networks of size up to $n = 10^7$; our results for $d = 2$ are qualitatively similar.

In each round of the rewiring process, we choose the source node x uniformly, and choose the destination y according to a *demand distribution* $Q(x, y)$. One method of selection is to choose a distance ℓ according to a distribution $Q(\ell)$, and then choose y randomly from among the nodes that satisfy $|x - y| = \ell$. Here, we take $Q(\ell)$ to be uniform, which for $d = 1$ means that y is simply a uniformly random node; we comment in the Conclusions on the effect of other demand distributions.

Starting at x , the greedy algorithm produces a path $x = x_0, x_1, x_2, \dots, x_T = y$ where T is the routing time, and where for each $0 \leq i < T$, x_{i+1} is the outgoing neighbor of x_i that minimizes $|x_{i+1} - y|$ (with ties broken randomly). The rewiring process works as follows: if $T \geq T_{\text{thresh}}$, i.e., if the greedy algorithm reaches some x_i with $i = T_{\text{thresh}}$, then we discontinue the search and change x ’s outgoing link to point to x_i .

In our experiments, we choose the threshold T_{thresh} uniformly at random from the interval $[1, \ell]$. Such a naive selection avoids making assumptions about the network’s size, performance or topology. In particular, it avoids assuming the $O(\log^2 n)$ routing time which both the rewired networks and Kleinberg’s networks achieve.

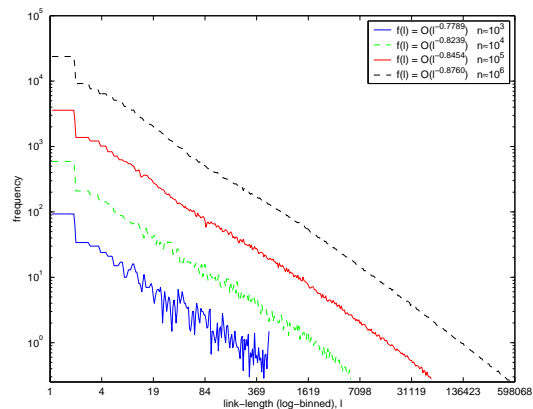


FIG. 1: Link length distributions with power-law fits for networks of size $n = \{10^3, 10^4, 10^5, 10^6\}$ after running the rewiring process until $T_{\text{rewired}} \leq 1.01 \cdot T_{\text{opt}}$. The initial graphs had $\alpha_0 = \infty$, i.e., only self-loops.

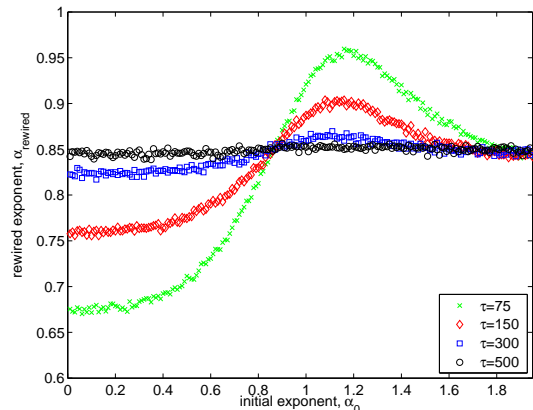


FIG. 2: Convergence of α_{rewired} for a range of initial link length exponents α_0 , where $n = 10^5$. At τ increases, α_{rewired} converges to $\alpha_{\text{rewired}} \approx 0.85$ regardless of the value of α_0 .

For our initial conditions, the link length distributions have exponent α_0 ranging from 0, a uniform distribution as in the Watts-Strogatz model, to ∞ , where all the “long-range links” are simply self-loops. We run τ rounds per node of the rewiring process; that is, τn times we choose nodes x, y , run the greedy algorithm from x to y and rewire if $T \geq T_{\text{thresh}}$. We then compare the final link length distribution to Kleinberg’s optimum $f(\ell) \sim \ell^{-1}$.

Figure 1 illustrates the resulting distributions for networks of size $n = \{10^3, 10^4, 10^5, 10^6\}$ and $\alpha_0 = \infty$. We see that after τn rounds (we describe our choice of τ in the next section) the rewiring process has built a power-law distribution of link lengths, $f(\ell) \sim \ell^{-\alpha_{\text{rewired}}}$ with $\alpha_{\text{rewired}} \approx \{0.78, 0.82, 0.85, 0.88\}$ respectively. Figure 2 shows that a range of initial distributions converge to the same final distribution. Starting with initial exponents α_0 ranging from 0 to 2, the exponent of the rewired link length distributions all converge to $\alpha_{\text{rewired}} \approx 0.85$ (here $n = 10^5$) as the rewiring continues.

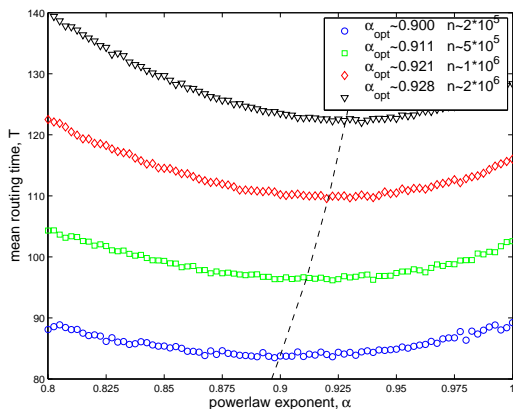


FIG. 3: Mean routing time $T(\alpha)$ for several network sizes. The dashed line shows the path α_{opt} , defined as the minimum of the best quadratic fit, follows as n increases.

III. FINITE-SIZE EFFECTS

While our rewiring process produces a power-law link length distribution, the exponent is noticeably different from the value $\alpha = 1$ (more generally, $\alpha = d$) which Kleinberg proved in the limit $n \rightarrow \infty$. Notably, Figure 2 shows that networks with an initial exponent $\alpha_0 = 1$ actually move away from this value as they are rewired.

This deflection appears to be due to finite-size effects, and in fact the exponent α_{opt} that minimizes the routing time on finite lattices turns out to be rather different from Kleinberg's value even at reasonably large n . We examined system sizes over six orders of magnitude, $10^1 \leq n \leq 10^7$. For each size, we constructed networks with α ranging from 0 to 1, and measured the mean routing time $T(\alpha)$ over $50n$ trials for each α . We estimated α_{opt} as the minimum of the best quadratic fit of $T(\alpha)$.

Figure 3 illustrates $T(\alpha)$ for several values of n and clearly shows that $\alpha_{\text{opt}} < 1$. The dashed line illustrates α_{opt} 's slow approach to 1 as n increases. Figure 4 shows the dependence of α_{opt} on n ; it is fit rather closely by the form $\alpha_{\text{opt}} = 1 - A/\log^2 n$ for $A \approx 2.85$. The fact that α_{opt} converges polylogarithmically, as opposed to polynomially, indicates that finite-size effects are quite severe.

However, while α_{rewired} differs from d (and from α_{opt}) our rewired networks achieve a mean routing time equivalent to networks with exponent α_{opt} , and better than those with exponent d . Fig. 5 compares the mean routing times for the three types of networks. All three are closely fit with a curve of the form $T(n) \sim \log^2 n$, just as Kleinberg's analysis predicts [4, 5].

Of course, the routing time of the network depends on the number of rewiring rounds. In Figure 6, we show the number of rounds per node τ required to achieve a mean routing time T_{rewired} which is only 1% greater than T_{opt} , i.e., the routing time of a network with exponent α_{opt} . This *rewiring time* grows as $\tau \sim n^{0.77}$ rounds per node, or $n^{1.77}$ rounds total.

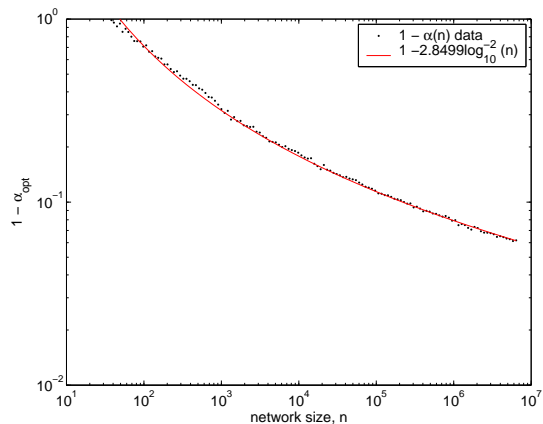


FIG. 4: Finite-size dependence of α_{opt} for n up to 10^7 . The solid line shows a fit to $1 - \alpha_{\text{opt}} \sim 1/\log^2 n$, showing that α_{opt} converges polylogarithmically slowly to d as $n \rightarrow \infty$.

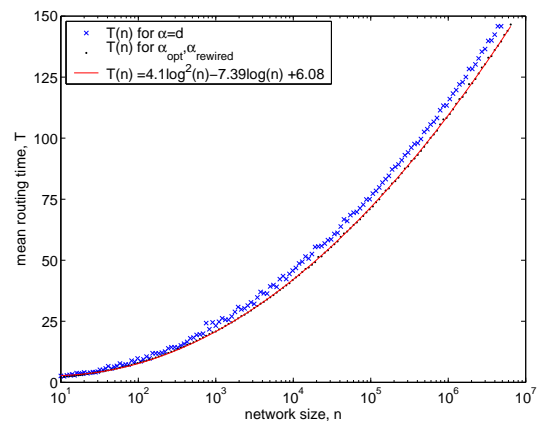


FIG. 5: Mean routing time T for networks with $\alpha = \{d, \alpha_{\text{opt}}, \alpha_{\text{rewired}}\}$. Shown is a fit to $\alpha = \alpha_{\text{rewired}}$ of the form $T \sim \log^2 n$.

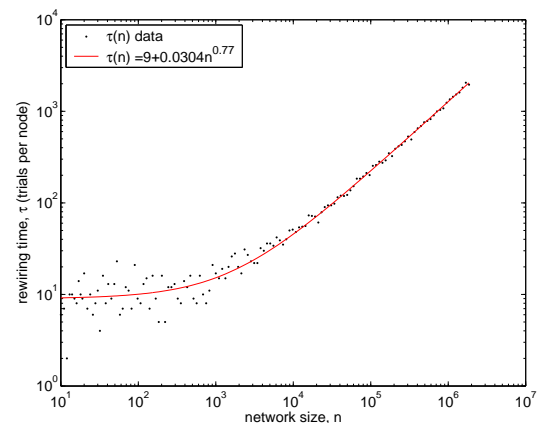


FIG. 6: The rewiring time $\tau(n)$, i.e., the number of rounds per node the rewiring process needs to achieve a routing time $T_{\text{rewired}} \leq 1.01 \cdot T_{\text{opt}}$, within 1% of that of a network with exponent α_{opt} . The solid line is a fit of the form $\tau \sim n^{0.77} + C$.

IV. DISCUSSION AND CONCLUSIONS

Kleinberg [4, 5] explored the navigability of small-world networks built on an underlying d -dimensional space. In the limit of infinite size, he found the mean routing time is minimized by a power-law distribution of link lengths with exponent $\alpha = d$ [10]. Here, we have explained how social networks might develop such a distribution over time, by introducing a simple, decentralized rewiring process that relies only on local information (and is even ignorant of the size of the network).

This process has a natural interpretation: the topology of a social network is constantly being modified by its members, who update their personal connections as they explore and navigate the network. If a member becomes frustrated because the journey to a destination takes too long, they can be expected to change their connections to make similar journeys more quickly in the future.

Our results show that rewiring causes a wide range of initial topologies to converge to a power-law distribution of link lengths, very similar to Kleinberg's. The exponent α_{rewired} we obtain differs significantly from d , its optimal value on infinite lattices. We attribute this deflection to finite size effects which cause the optimal exponent α_{opt} to converge polylogarithmically as $n \rightarrow \infty$. However, the rewired network achieves the same mean routing time as power-law networks with exponent α_{opt} , and better routing times than those with $\alpha = d$. Specifically, the mean routing time as a function of system size is $O(\log^2 n)$ as predicted by Kleinberg's analysis. The number of rounds of the rewiring process needed to achieve this routing time grows as a low-degree polynomial of n .

In addition to creating and maintaining a power-law distribution of link lengths, we believe this rewiring process to be *adaptive*. For instance, we conjecture that if new nodes are added to the network, or if certain nodes or links are removed, it will dynamically optimize for these new situations. We also believe, based on preliminary results, that if the demand distribution $Q(x, y)$ is not uniform, e.g. if certain destinations are more popular than others, or if the source and the destination are correlated (both of which are true in any real network), that it will optimize routing times for the source-destination pairs that appear more frequently. These adaptive properties would be particularly useful in networks where nodes are constantly coming on- and off-line and where the demand of each destination rises and falls over time, such as live peer-to-peer networks, distributed sensor networks or massively parallel computers.

Our discussion contemplates a "social space" consisting of a finite-dimensional lattice, an obviously poor model for the complex social spaces we routinely navigate. An interesting study would be an analogous rewiring process for networks whose underlying structure is hierarchical [7], involves multiple group affiliations [8] or is otherwise structured, as in the peer-to-peer network Freenet [16] with the modifications described in [17]. We leave these as directions for future work.

Acknowledgments

The authors are grateful to Mark Newman and Jon Kleinberg for helpful discussions. This work was funded by NSF grant PHY-0200909, and the Los Alamos National Laboratory.

-
- [1] M.E.J. Newman, "The structure and function of complex networks." *SIAM Review* **45** (2003) 167-256.
- [2] S. Milgram, "The small world problem." *Psychology Today* **2** (1967) 60-67.
- [3] P.S. Dodds, R. Muhamad, and D. J. Watts "An experimental study of search in global social networks." *Science* **301** (2003) 827-829.
- [4] J. Kleinberg. "The small-world phenomenon: an algorithmic perspective." *Proc. 32nd ACM Symposium on Theory of Computing* (2000) 163-170.
- [5] J. Kleinberg, "Navigation in a small world." *Nature* **406** (2000) 845.
- [6] D.J. Watts and S.H. Strogatz, "Collective dynamics of small-world networks." *Nature* **393** (1998) 440-442.
- [7] J. Kleinberg, "Small-world phenomena and the dynamics of information." *Advances in Neural Information Processing Systems* **14** (2001) 431-438.
- [8] D.J. Watts, P.S. Dodds and M.E.J. Newman, "Identity and search in social networks." *Science* **296** (2002) 1302-1305.
- [9] Note that our model conflates links and bookmarks; a surfer's bookmarks are located at her home node and may be utilized in others' navigation.
- [10] In [5], Kleinberg noted that for a $n = 20,000$, $d = 2$ graph, the optimal value of α was slightly below d .
- [11] A. Vazquez. "Growing networks with local rules: preferential attachment, clustering hierarchy and degree correlations", *Phys. Rev. E* **67** (2003) 056104.
- [12] P.L. Krapivsky and S. Redner. "Organization of Growing Random Networks". *Phys. Rev. E* **63** (2001) 066123.
- [13] S.S. Manna and A. Kabakcioglu, "Scale-free network on Euclidean space optimized by rewiring of links." Preprint, `cond-mat/0302224`.
- [14] M. Baiesi and S.S. Manna, "Scale free networks from a self-organizing Hamiltonian dynamics." Preprint, `cond-mat/0305054`.
- [15] Ph. Blanchard, A. Ruschhaupt, and T. Krueger, "Small world graphs by the iterated 'my friends are your friends' principle." Preprint, `cond-mat/0304563`.
- [16] I. Clarke, et al, "Freenet: A Distributed Anonymous Information Storage and Retrieval System," in *Designing Privacy Enhancing Technologies*, LNCS 2009, H. Federath, ed., Springer-Verlag, Berlin (2001).
- [17] A. Goel, H. Zhang, and R. Govindan. "Using the Small-World Model to Improve Freenet Performance." *Proc. IEEE Infocom* (2002).