1 Six Degrees of Separation

The small-world phenomenon – the principle that we are all linked by short chains of acquaintances, or "six degrees of separation" [2] – has long been the subject of anecdotal fascination, and more recently has become the subject of both experimental and theoretical research. At its most basic level, it is a statement about social networks — in particular, about the network with one node corresponding to each person in the world, and an edge joining two people if they know each other on a first-name basis. When we say that this graph is a "small world," we mean, informally, that almost every pair of nodes is connected by a path with an extremely small number of steps.

Given the complexity of the network, however, and the difficulty in determining its structure, it seems like a daunting task to test this hypothesis: how would one try to actually verify, empirically, the claim that most pairs of nodes in this graph are connected by short paths?

The social psychologist Stanley Milgram [7, 8] took up this challenge in the 1960s, conducting an experiment to test the small-world property by having people explicitly construct paths through the social network defined by acquaintanceship. To this end, he chose a *target person* in the network, a stockbroker living in a suburb of Boston, and asked a collection of randomly chosen "starter" individuals each to forward a letter to the target. He provided the target's name, address, occupation, and some personal information, but stipulated that the participants could not mail the letter directly to the target; rather, each participant could only advance the letter by forwarding it to a single acquaintance that he or she knew on a first-name basis, with the goal of reaching the target as rapidly as possible. Each letter thus passed successively from one acquaintance to another, closing in on the stockbroker outside Boston.

The letters thus acted as virtual "tracers," mapping out paths through the social network. Milgram found that the median length among the completed paths was six, providing the first concrete evidence for the abundance of short paths connecting far-flung pairs of individuals in society, as well as supplying the basis for the number "six" in the resulting pop-cultural mantra. One needs to be careful in interpreting this finding, of course: many of the chains never reached the target, and the target himself was a relatively "high-status" individual who may have been easier to reach than an arbitrary person (see e.g. the recent critique by Judith Kleinfeld [5]). But since Milgram's work, the overall conclusion has been accepted at least at a qualitative level: social networks tend to exhibit very short paths between essentially arbitrary pairs of nodes.

2 Structure and Randomness

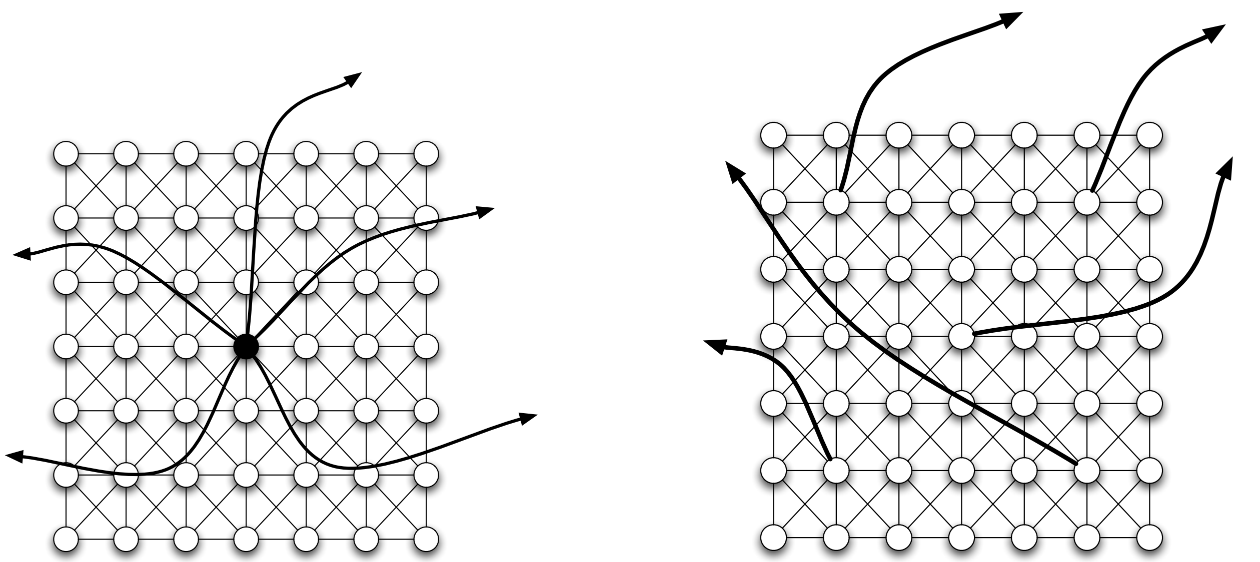
Should we be surprised by the fact that the paths between seemingly arbitrary pairs of people are so short? Here's a basic argument suggesting that short paths are at least compatible with intuition. Suppose each of knows more than 100 other people on a first-name basis (in fact, for most people, the number is significantly larger). Then, taking into account the fact that each of your friends has at least 100 friends other than you, you could in principle be two steps away from over $100 \cdot 100 = 10,000$ people. Taking into account the 100 friends of these people brings us to more than $100 \cdot 100 \cdot 100 = 1,000,000$ people who in principle could be three steps away. In other words, the numbers are growing by powers of 100 with each step, bringing us to 10 billion after five steps.

There's nothing mathematically wrong with this reasoning, but it's not clear how much it tells us about real social networks. The difficulty already manifests itself with the second step, where we conclude that there may be more than $10,000$ people within two steps of you. As we've seen, social networks abound in closed triads — sets of three people who mutually know each other — and in particular, many of your 100 friends will know each other. As a result, when we think about the nodes you can reach by following edges from your friends, many of these edges go from one friend to another, not to the rest of world. The number $10,000$ came from assuming that each of your 100 friends was linked to 100 *new* people; and without this, the number of friends you could reach in two steps could be much smaller.

So the effect of triadic closure in social networks works to limit the number of people you can reach by following short paths. And, indeed, at an implicit level, this is a large part of what makes the small-world phenomenon surprising to many people when they first hear it: the social network appears from the local perspective of any one node to be highly clustered, not the kind of massively branching structure that would more obviously reach many nodes along very short paths.

The Watts-Strogatz model. Can we make up a simple model that exhibits both features: many closed triads, but also very short paths? Duncan Watts and Steve Strogatz proposed such a model in 1998 [9], essentially by directly combining these two features in a single network. Paraphrasing their original suggestion slightly (but keeping the main idea intact), let's imagine that everyone lives on an $n \times n$ grid — we can imagine the grid as a model of geographic proximity, or potentially some more abstract kind of social proximity. We'll assume that every node knows all other nodes within a radius of r grid steps, for some constant value of r; these are the people you know simply because you live close to them. Then, for some other constant value k, everyone also knows k other nodes selected uniformly at random from the grid; these are the people you happen to know far away, for random reasons having nothing to do with the proximity imposed by the grid. (See Figure 1(a).)

So this gives us a way to construct a random directed graph, in which each node has roughly r^2+k out-links (roughly r^2 out-links to the nearby nodes, and k more random ones). Let's now consider the properties of the network — for all these properties, the question is not whether they hold for certain, but whether they hold with high probability (since clearly there may be certain rare random outcomes where the k edges out of each node don't buy



(a) *Local structure plus random edges*

(b) *Very few random edges still cause a small world*

Figure 1: The Watts-Strogatz model arises from a highly clustered network (such as the grid), with a small number of random links added in.

you anything).

Why are the paths short? To begin with, a network constructed this way clearly has many closed triads, since many of a node's friends are nearby on the grid, and hence know each other according to the model. Now, what about the number of hops in the shortest path between two far-apart nodes on the grid? This is where the k random edges out of each node prove very useful. Just following the random edges, without using the grid, a given node v can reach k other nodes in one step. Since these edges are chosen uniformly at random, these k friends of v are in fact likely to have very little overlap in who *their* random friends are. So in two steps using random edges only, v can reach close to $k \cdot k = k^2$ other nodes. This reasoning continues to hold for larger numbers of steps, with some loss due to the eventual overlap in sets of friends; but after j steps using random edges only, v can reach close to k^j other nodes. The $n \times n$ grid has n^2 nodes total, and setting $k^j = n^2$, we see that we have a chance of reaching everyone once j reaches $\log_k n^2 = 2 \log_k n$. It takes some fairly intricate analysis to do this precisely, accounting for the increasing amount of overlap among random friends as the number of steps increases, but this is in fact approximately the maximum number of hops needed to get between any pair of nodes in this model — in other words, logarithmic in the total size of the population.

The exact details of the model are less crucial for our discussion than some of the high-level points. First, the premise behind the Watts-Strogatz model is to think of the small-world phenomenon as arising from a superposition of structure and randomness – the structured, orderly grid by itself does not have particularly short paths linking far-apart pairs of nodes, but if we sprinkle a small number of random links into such a network, then the distance between nodes drops dramatically. In fact, this mixture of structure and randomness clearly parallels that distinction between strong and weak ties [1] that we saw earlier in the course — the proximity-based edges on the grid correspond to strong ties, creating many closed triads, while the random edges correspond to weak ties; the distance between the two endpoints of a random edge may increase significantly if the edge were to be deleted. It is the strong ties that make the world clustered, while it is these weak ties, as in Granovetter's original conception, that make the world "small," by creating an abundance of short paths.

Once we understand how this type of "hybrid" network leads to short paths, we in fact find that a surprisingly small amount of randomness is needed to achieve the same effect. Suppose, for example, that instead of allowing each node to have k random friends, we only allow one out of every k nodes to have a *single* random friend — keeping the proximity-based edges as before, as illustrated schematically in Figure 1(b). (We can think of this as corresponding to a technologically earlier time, when most people only knew their near neighbors, and a few people knew someone far away.) Even this network will have short paths between all pairs of nodes. To see why, suppose that we conceptually group $k \times k$ subsquares of the grid into "towns." Now, consider the small-world phenomenon at the level of towns, each town contains approximately k people who each have a random friend, and so the town collectively has k links to other towns selected uniformly at random. So this is just like the previous model, except that towns are now playing the role of individual nodes — and so we can find short paths between any pair of towns. But now to find a short path between any two people, we first find a short path between the two towns they inhabit, and then use the proximity-based edges to turn this into an actual path in the network on individual people.

This, then, is the crux of the Watts-Strogatz model: introducing a tiny amount of randomness into the world is enough to make it "small," with short paths between every pair of nodes.

3 Decentralized Search

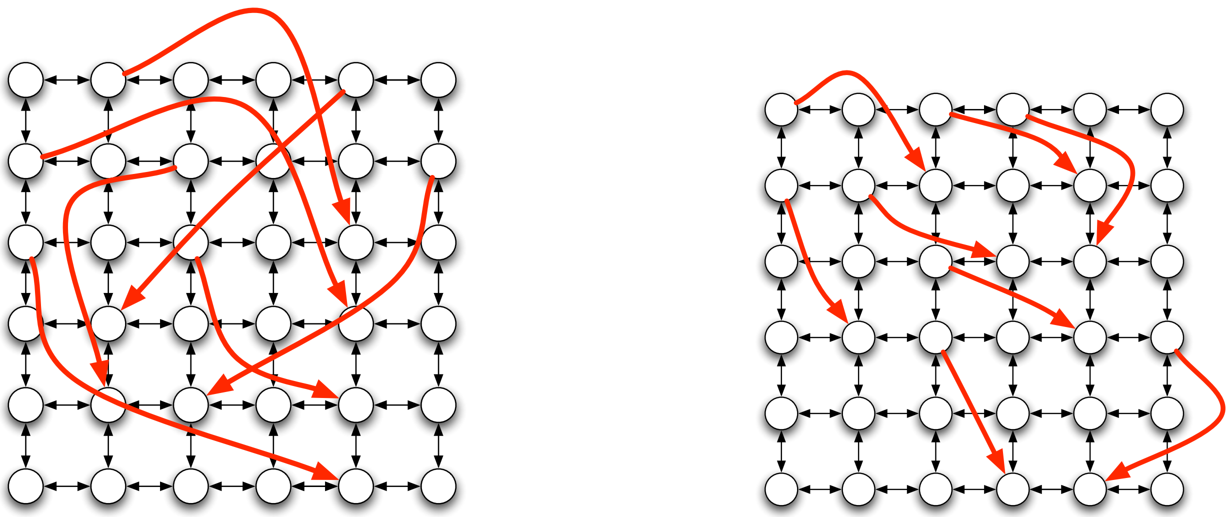
Let's go back to Stanley Milgram's original experiment, though, because it contains a further surprising point beyond just the presence of short paths. The point is this: not only did short paths turn out to exist in the social network, but people, using knowledge only of their own acquaintances, were able to collectively construct paths to the target. This was a necessary consequence of the way Milgram formulated the task for his participants; if one really wanted the *shortest* path from a starting person to the target, one should have instructed the starter to forward a letter to *all* of his or her friends, who in turn should have forwarded the letter to all of their friends, and so forth. This "flooding" of the network would have reached the target as rapidly as possible; but for obvious reasons, such an experiment was not a feasible option. As a result, Milgram was forced to embark on the much more interesting experiment of constructing paths by "tunneling" through the network, with the letter advancing just one person at a time — a process that could well have failed to reach the target, even if a short path existed.

So the success of the experiment raises fundamental questions about the power of collective search: even if we posit that the social network contains short paths, why should it have been structured so as to make this type of decentralized routing so effective? Clearly the network contained some type of "gradient" that helped participants guide messages toward the target. As with the Watts-Strogatz model, which sought to provide a simple framework for thinking about short paths in highly clustered networks, this type of search is also something we can try to model: can we construct a random network in which decentralized routing succeeds, and if so, what are the qualitative properties that are crucial for success?

A model for decentralized search. To begin with, it is not difficult to model the kind of decentralized search that was taking place in the Milgram experiment. Starting with the grid-based model discussed above, we suppose that a starting node s is given a message that it must forward to a target node t, passing it along edges of the network. Initially s only knows the location of t on the grid, but, crucially, it does not know the random edges out of any node other than itself. Each intermediate node along the path has this partial information as well, and it must choose which of its neighbors to send the message to next. These choices amount to a collective procedure for finding a path from s to t — just as the participants in the Milgram experiment constructed paths to the target person. We will evaluate different search procedures according to their *delivery time* — the expected number of steps required to reach the target, over a randomly generated set of long-range contacts, and randomly chosen starting and target nodes.

What should we view as a good delivery time? Qualitatively, the main feature of the model from the previous section is that the length of paths between arbitrary nodes is proportional to $\log n$, which is a very slowly growing function of n. Motivated by this, we'll take as our dividing line the difference between delivery times that grow like $\log n$ to some fixed power (e.g. $\log n$ or $(\log n)^2$), and those that grow like n to some fixed power (e.g. $n^{1/3}$ or $n^{1/2}$). There is an exponential gap in how quickly these two kinds of functions grow; the former will be viewed as "good" delivery times, while the latter will be viewed as "bad." Of course, it's important to keep in mind that these distinctions are really about how the functions scale as n increases; for medium values of n, they may not actually be so far apart from each other.

Given all this, we can start with the grid-based model of the previous section, and ask how well decentralized search can perform in this network. Perhaps surprisingly, the answer is negative: one can show that the delivery time of any decentralized algorithm in the grid-based model must be at least proportional to $n^{2/3}$ [4]. The analysis is not something we'll be able to go into here, but the underlying informal idea is that any method for decentralized search can potentially make some initial progress toward the target, but it rapidly becomes "lost" — the random links lead in arbitrary directions, and we are left with essentially a hopeless maze-solving problem in order to determine which of the random links are reasonable first steps on short paths to the target.



(a) *A small clustering exponent* (b) *A large clustering exponent*

Figure 2: With a small clustering exponent, the random edges tend to span long distances on the grid; as the clustering exponent increases, the random edges become shorter.

Thus, the grid-based model turns out to be a case where there is a gap between the existence of short paths and the ability of decentralized action to find them: although we know that there are paths of length proportional to $\log n$, no decentralized method is capable of finding them. This is clearly not the end of the story; the obvious question is whether we can find a natural variation on the network model in which efficient decentralized search becomes possible.

Generalizing the network model. At some level, the problem with the grid-based model is that the random links are too "unstructured" to be useful in getting all the way to the target. To deal with this, we can vary the model so that even the random links are related a bit more closely to the geometry of the grid.

Specifically, let's introduce one extra quantity, a *clustering exponent* q. We have nodes on a grid as before, and each node still has edges to each other node within r grid steps. But now, each of its k random edges is generated as follows. For two nodes v and w, let $d(v, w)$ denote the number of grid steps between them. (This is their distance if one had to walk along adjacent nodes on the grid.) In generating a random edge out of v, we have this edge link to w with probability proportional to $d(v, w)^{-q}$.

So we in fact have a different model for each value of q. The original grid-based model corresponds to $q = 0$, since then the links are choxsen uniformly at random; and varying q is like turning a knob that controls how uniform the random links are. In particular, when q is very small, the long-range links are "too random," and can't be used effectively by a decentralized algorithm (as we saw specifically for the case $q = 0$ above); when q is large, the long-range links appear to be "not random enough," since they simply don't provide enough

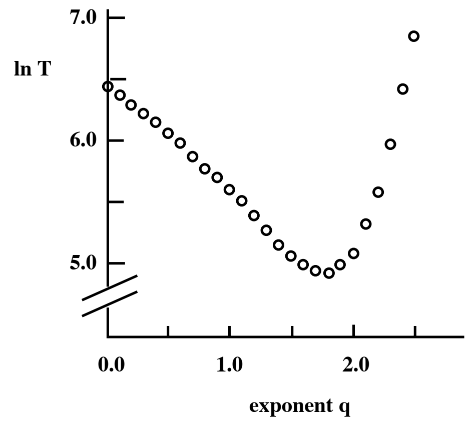


Figure 3: Simulation of greedy routing in the grid-based model with clustering exponent q. Each point is the average of 1000 runs on (a slight variant of) a grid with 400 million nodes. The delivery time is best in the vicinity of exponent $q = 2$, as expected; but even with this number of nodes, the delivery time is comparable over the range between 1.5 and 2.

of the long-distance jumps that are needed to create a small world. Pictorially, this variation in q can be seen in the difference between the two networks in Figure 2. Is there an optimal operating point for the network, where the distribution of long-range links is sufficiently balanced between these extremes to be of use to a decentralized routing algorithm?

In fact there is. The main result for this model is that, when $q = 2$ (so random links follow an inverse-square distribution), then there is a decentralized rule that nodes can follow with delivery time proportional to $(\log n)^2$ [4]. Moreover, when q is any value other than 2, the delivery time of any decentralized algorithm is at least proportional to n^c, for some value c that depends on q. Thus, $q = 2$ is the only exponent where it is possible to achieve a delivery time that is bounded by $\log n$ to some fixed power. Below, we'll give the basic idea why $q = 2$ has this property.

The Inverse-Square Network. The decentralized method that achieves rapid delivery time when $q = 2$ is very simple: each node simply forwards the message to the neighbor that lies closest to the target on the grid (i.e. with distance measured in grid steps). We call this *greedy routing*, since it "greedily" tries to reduce the distance to the target as much as possible in each step. This is a reasonable approximation to one of the main strategies used by participants in Milgram's experiment — many people forwarded the letter to the friend who lived as close to the target as possible [3].

The results about the dependence on q are phrased as functions of n; to get a sense for how the delivery time varies with q for a fixed network size, we can simulate greedy routing on randomly generated networks for different values of q. Figure 3 shows the results for networks on 400 million nodes; the delivery time is indeed best in the vicinity of exponent $q = 2$, as expected; but even with this number of nodes, the delivery time is comparable over the range between 1.5 and 2. This relates to our earlier point, that powers of $\log n$ are

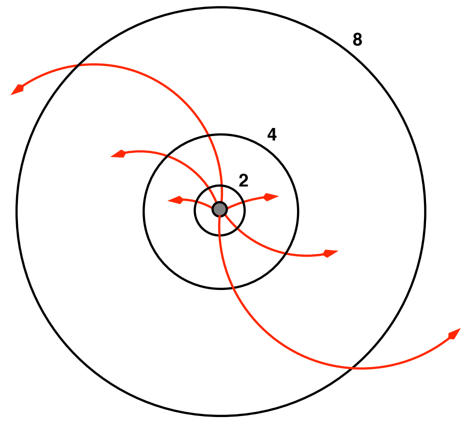


Figure 4: The concentric scales of resolution around a particular node.

eventually smaller than powers of n as n grows, but n may need to grow quite large for this effect to become visible. (For example, you can check that $(\log_2 n)^2$ and $n^{1/3}$ are roughly the same value when n is a billion.)

So now we can return to the question: what's special about $q = 2$? Here's a rough outline of the analysis that shows that greedy routing has delivery time proportional to $(\log n)^2$ in this case. First, let's think about the inverse-square distribution as follows. Consider a node v, and a distance d, and look at the "ring" of all grid points that are at least d grid steps from v, but less than $2d$. (So this is essentially the set of points between two concentric discs of radius d and $2d$.) For a random edge from v, the probability that it links to any node in this ring varies a little depending on how far into the ring the particular node is, but all these probabilities are approximately proportional to d^{-2}; and there are approximately d^2 nodes in this ring. So these two terms approximately cancel out, and we conclude: the probability that a random edge links into *some node* in this ring is approximately independent of the value of d.

This suggests that the inverse-square distribution can be approximately pictured as follows. From the perspective of any node v, we divide all other nodes into concentric rings at distance between 1 and 2, between 2 and 4, between 4 and 8; and more generally between any consecutive powers of 2. (See Figure 4.) This gives us $\log_2 n$ different "scales of resolution" at which other nodes can live, when viewed from v. (In geographic terms, these different scales of resolution can be viewed intuitively as the set of all people who live, for example, on your block, in your neighborhood, in your city, your state, your country, and the whole world.) The point of our reasoning in the previous paragraph is that the inverse-square distribution is approximately equally likely to produce a random edge that links into any of these scales of resolution. This is the key structural property that makes it conducive to effective decentralized routing.

In particular, as nodes attempt to reach the target by following links, their "progress" can be measured by keeping track of which of the $\log n$ scales of resolution they're in, as viewed from the target. One can show, using the kind of reasoning we've just been going through, that the message will spend only about $\log n$ steps at each scale of resolution before finding a link that leads into the next closer scale (or even closer). As there are $\log n$ scales of resolution in total, and about $\log n$ steps are spent in each, this gives a delivery time proportional to $(\log n)^2$.

So really, the decentralized search is working because it's rapidly "funneling" its way toward the target, metaphorically working its way from the scale of countries down to states, then to cities, then neighborhoods, then blocks. In this way, it's not unlike how the U.S. Postal Service uses the address on an envelope for routing (since, ignoring the ZIP code, an address exactly specifies scales of resolution, including information about the country, state, city, street, and finally the exact street number). But the point is that U.S. postal routing works well because they've invested enormous resources in designing it to do so; on the other hand, the corresponding funnel-down patterns are arising from completely random structures in the inverse-square network.

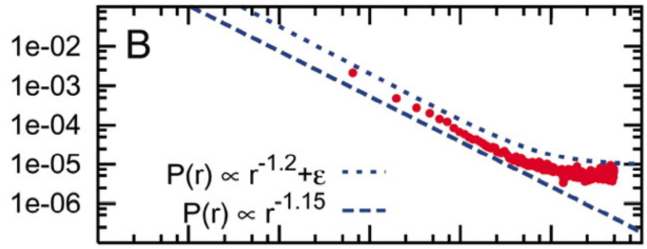
Earlier we asked whether a concrete model for analyzing decentralized search might suggest, at a more general level, some qualitative structures that lead to efficient routing. This idea of scales of resolution is one such qualitative pattern, and it can be used to generalize the model considerably, away from the specifics of a regular grid structure. A number of such generalizations have been pursued; we discuss one in the next section, motivated by the goal of empirically testing the grid model.

4 Empirical Analysis and Rank-Based Friendship

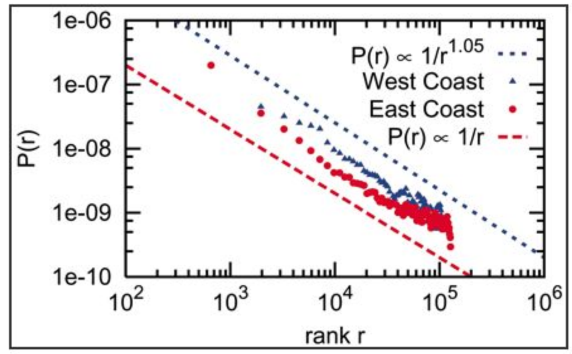
It's natural to ask whether we can find, in real social networks, some analogue of this special exponent $q = 2$. Clearly this will require some amount of translation between the abstract models and reality, since we don't live on a grid, our friendships are not created at random, and so forth. However, using the grid as a metaphor for geography, it's clear how to begin approaching this task: we should find a large dataset that specifies where people live and who their friends are; we can then try to determine how the probability of an edge in the social network decreases with the distance between its endpoints and see if it looks at all like an inverse-square law.

In the past few years, the rich data available on social networking sites has made it much easier to get information of this sort, at a large scale. Liben-Nowell et al. [6] used the blogging site LiveJournal for precisely this purpose, analyzing roughly 500,000 users who provided a U.S. ZIP code for their home address, as well as links to their friends on the system. Note that LiveJournal is serving here primarily as a very useful "model system," containing data on the geographic basis of friendship links on a scale that would be enormously difficult to obtain by more traditional survey methods. From a methodological point of view, it is an interesting and fairly unresolved issue to understand how closely the structure of friendships defined in on-line communities correspond to the structure of friendships as we understand them in off-line settings.

A number of things have to be done in order to align the LiveJournal data with the basic grid model, and perhaps the most subtle involves the fact that the population density



(a) *Rank-based friendship on LiveJournal*



(b) *Rank-based friendship: East and West coasts*

Figure 5: The probability of a friendship as a function of geographic rank on the blogging site LiveJournal.

of the users is extremely non-uniform (as it is for the U.S. as a whole). In particular, the inverse-square distribution is useful for routing when nodes are uniformly spaced in two dimensions; what's a reasonable generalization to the case in which they can be spread very non-uniformly?

Rank-Based Friendship. One approach that works well is to determine link probabilities not by physical distance, but by *rank*. Let's suppose that as a node v looks out at all other nodes, it *ranks* them by proximity: the rank of a node w is equal to the number of other nodes that are closer to v than w is. Now, suppose that for some exponent p, node v creates a random link as follows: it chooses a node w as the other end with probability proportional to $\text{rank}(w)^{-p}$. We will call this *rank-based friendship* with exponent p.

Which choice of exponent p would generalize the inverse-square distribution for uniformly-spaced nodes? If a node w in a uniformly-spaced grid is at distance d from v, then it lies on the circumference of a disc of radius d, which contains about d^2 closer nodes — so its rank is approximately d^2. Thus, linking to w with probability proportional to d^{-2} is approximately the same as linking with probability $\text{rank}(w)^{-1}$, so this suggests that exponent $p = 1$ is the right generalization of the inverse-square distribution.

In fact, Liben-Nowell et al. were able to prove that for essentially any population density, if random links are constructed using rank-based friendship with exponent 1, the resulting network allows for efficient decentralized search with high probability. In addition to generalizing the inverse-square result for the grid, this result has a nice qualitative summary: to construct a network that is efficiently searchable, create a link to each node with probability that is inversely proportional to the number of closer nodes.

Now one can go back to LiveJournal and see how well rank-based friendship fits the distribution of actual social network links: we consider pairs of nodes where one assigns the other a rank of r, and we ask what fraction of these pairs are actually friends, as a function of r. Does this fraction decrease approximately like r^{-1}? Since we're looking for a power-law relationship between the rank r and the fraction f of rank-r pairs connected by an edge, we can use an idea that we saw earlier in our discussion of power laws: rather than plotting f as a function of r, we can plot $\log f$ as a function of $\log r$, see if we find an approximately straight line, and then estimate the exponent p as the slope of this line.

Figure 5(a) shows this result for the LiveJournal data; we see that much of the body of the curve is approximately a straight line sandwiched between slopes of -1.15 and -1.2, and hence close to the optimal exponent of -1. It is also interesting to work separately with the more structurally homogeneous subsets of the data consisting of West-Coast users and East-Coast users, and when one does this the exponent becomes very close to the optimum of -1. Figure 5(b) shows this result: The lower dotted line is what you should see if the points followed the distribution rank^{-1}, and the upper dotted line is what you should see if the points followed the distribution $\text{rank}^{-1.05}$.

The fit to the optimal distribution for search is thus much closer than one might have expected — and this is all the more surprising given that users on LiveJournal are certainly not explicitly engineering their social structure to support rapid decentralized search. Rather, this distribution is clearly arising from certain processes or selective pressures that are more subtle, and manifested in aggregate. Understanding what the forces are that drive a network toward this operating point is an interesting topic of current research.

References

[1] Mark Granovetter. The strength of weak ties. *American Journal of Sociology*, 78:1360–1380, 1973.

[2] John Guare. *Six Degrees of Separation: A Play.* Vintage Books, 1990.

[3] Peter D. Killworth and H. Russell Bernard. Reverse small world experiment. *Social Networks*, 1:159–192, 1978.

[4] Jon Kleinberg. Navigation in a small world. *Nature*, 406:845, 2000.

[5] Judith Kleinfeld. Could it be a big world after all? The 'six degrees of separation' myth. *Society*, April 2002.

[6] David Liben-Nowell, Jasmine Novak, Ravi Kumar, Prabhakar Raghavan, and Andrew Tomkins. Geographic routing in social networks. *Proc. Natl. Acad. Sci. USA*, 102(33):11623–11628, August 2005.

[7] Stanley Milgram. The small-world problem. *Psychology Today*, 2:60–67, 1967.

[8] Jeffrey Travers and Stanley Milgram. An experimental study of the small world problem. *Sociometry*, 32(4):425–443, 1969.

[9] Duncan J. Watts and Steven H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393:440–442, 1998.