CS599: Structure and Dynamics of Networked Information (Spring 2005)
03/30/2005: Small World
Scribes: Yuriy Brun and Shyam Kapadia

We have observed that random graphs do not model social networks well, as social networks exhibit a lot of clustering, whereas triangles or other small cycles are rare in sparse random graphs. Instead, we considered a model proposed by Watts and Strogatz [3] that superimposed random edges on top of a clustered graph, e.g., a grid graph, thus obtaining both high clustering and small diameter. However, returning to Milgram's experiment, this does not yet answer the question of how people were actually able to *find* these short paths. Indeed, we saw that when the random links were uniformly random, then no local rule can find these short paths. So we want to adjust the model, and in particular the random long-range link distribution, to explain how people were actually able to find paths.

We assume that the underlying "clustered" graph is the 2-dimensional grid, and study the effect of different distributions of long-range links. Let's say Prob$[v \to w]$, the probability that v connects to w via the long-range link, is a function of $d_{v,w}$, the distance between v and w. Intuitively, it seems that connections between distant individuals are less likely, so the probability should be monotone decreasing.

If Prob$[v \to w]$ is an inverse exponential function of $d_{v,w}$, then it decreases very rapidly, and long-range links are too unlikely. Hence, we use a polynomially decreasing function in $d_{v,w}$ [1], i.e., Prob$[v \to w] \sim (d_{v,w})^{-\alpha}$ for a constant $\alpha \geq 0$.

To understand this distribution well, we need to calculate its normalizing constant, i.e., the γ such that Prob$[v \to w] = \frac{1}{\gamma} d^{-\alpha}$. By noticing that for each d, there are $\Theta(d)$ nodes at distance d from any given node v, we can calculate

$$\sum_w d_{v,w}^{-\alpha} \approx \sum_{d=1}^{n} d d^{-\alpha} = \sum_{d=1}^{n} d^{1-\alpha} = \begin{cases} \Theta(n^{2-\alpha}) & \text{for } \alpha < 2 \\ \Theta(\log n) & \text{for } \alpha = 2 \\ \Theta(1) & \text{for } \alpha > 2 \end{cases}$$

In our previous class, we proved that for $\alpha = 0$, the case where the destination of each long-range link is a uniformly random node (independent of the distance between v and w), nodes cannot route efficiently based solely on local information. For very large α, we also expect local routing (or any routing, for that matter) to not work well, as long distance links will be exceedingly rare. Of the three cases (1) $\alpha = 2$ (2) $\alpha < 2$ and (3) $\alpha > 2$, it seems thus most promising that $\alpha = 2$ may work well for local routing. Indeed, we will show that it is the only exponent for which local routing can work in poly-logarithmic time.

The case $\alpha = 2$

Claim 1 *For $\alpha = 2$, we can route locally in $\Theta(\log^2(n))$ steps in expectation.*

Proof. The algorithm is the simple greedy rule of always forwarding the message to the neighbor closest to the destination.

Let s be the source node and t the destination node, at distance d from s. We will show that within an expected $O(\log n)$ steps, the distance of the message to t is halved. (Notice that the time here is independent of d.) In order to prove this, we let $B_{d/2}(t) = \{v | d_{v,t} \leq d/2\}$ denote all nodes in the circle of radius $d/2$ around t.

Because there are $\Theta(d^2)$ nodes in the set $B_{d/2}(t)$, and each is at distance at most $\Theta(d)$ from s, we can lower-bound the probability for s's long-range link to end in $B_{d/2}(t)$ as follows:

$$\frac{1}{\Theta(\log n)} \sum_{v \in B_{d/2}(t)} d_{s,v}^{-\alpha} \geq \Theta\left(\frac{1}{\log n} \cdot d^2 \cdot d^{-\alpha}\right) = \Theta\left(\frac{1}{\log n}\right)$$

Notice that this gives a notion of "uniformity of scales": the probability of halving the distance is the same independent of what the distance itself is. Similarly, if we think of circles around v of size 2^k for $k = 0, \ldots$, then v has equal probability of having its long-range link in any of the annuli between circles of radius 2^k and 2^{k+1}.

By our calculations above, any single long-distance edge reaches $B_{d/2}(t)$ with probability at least $\Theta\left(\frac{1}{\log n}\right)$. If not, the message is moved to another node, no further away, which again has at least the same probability of reaching $B_{d/2}(t)$, etc. Hence, the number of steps until a long-range edge will reach $B_{d/2}(t)$ is lower-bounded by a negative binomial random variable with parameter $\Theta\left(\frac{1}{\log n}\right)$. In particular, the expected number of steps to hit $B_{d/2}(t)$ is $\Theta(\log n)$, and the actual number is sharply concentrated around the expectation. As the distance of the message from t is halved every $\Theta(\log n)$ steps (independently of the current distance), t will be reached after $\Theta\left(\log^2 n\right)$ steps. ■

The case $0 \leq \alpha < 2$

We already saw that for $\alpha = 0$, local routing takes a polynomial number of steps. The problem was that links were too unstructured, and the chain was unlikely to encounter any node with long-range link into a (sufficiently small) polynomial-sized ball around the destination. Here, we will generalize the construction to show that the same problem occurs for all $\alpha < 2$.

Claim 2 *For $\alpha < 2$, local routing requires a polynomial number of steps in expectation.*

Proof. Let $\delta = \frac{2-\alpha}{3}$, and $B(t)$ be the ball of radius n^δ around t.

This time, we want to upper-bound the probability that a given node has a long-range link into $B(t)$. Even when two nodes are very close together, say $d = 1$, the probability of a long-range link between them is at most $\frac{1}{\gamma} = n^{\alpha-2} = n^{-3\delta}$. As there are only $\Theta\left(n^{2\delta}\right)$ nodes in $B(t)$, the probability for a given node to have a long-range link into $B(t)$ is at most $\Theta\left(n^{-\delta}\right)$. Thus, it takes n^δ steps in expectation to encounter a long-range link into $B(t)$. On the other hand, if no such long-range link is encountered, then the chain takes at least n^δ small steps. In either case, the total number of steps is $\Omega\left(n^\delta\right)$. ■

The case $\alpha > 2$

For small α, the problem is that, while short paths exist, they cannot be found with local information alone, as the random links are "too random". For large α, the problem is that there are not many long-range links — the argument below needs to be strengthened only slightly to show that the diameter of the graph is actually polynomial in n.

Claim 3 *For $\alpha > 2$, the expected number of steps required to route from s to t is polynomial in n.*

Proof. Following our intuition that long-distance links are unlikely, we first calculate the probability that a link is longer than some given number m. Using again that there are $\Theta(d)$ nodes at distance d from a given node, this is at most

$$\Theta\left(\sum_{d=m}^{\infty} d d^{-\alpha}\right) = \Theta\left(\sum_{d=m}^{\infty} d^{1-\alpha}\right) = \Theta\left(m^{2-\alpha}\right)$$

Hence, the probability that a link's length is at least, say, $n^{1/3}$, is at most $\Theta(n^{\frac{2-\alpha}{3}})$. So if we only take $n^{\frac{\alpha-2}{3\alpha}}$ steps, the probability of having any of them encounter a long-range link longer than $n^{1/3}$ is polynomially small ($O(n^{\frac{2-\alpha}{3}(1-1/\alpha)})$). But if none of them encounter any longer links, then the total distance is at most $n^{1/3} \cdot n^{\frac{\alpha-2}{3\alpha}} = n^{\frac{\alpha-2}{\alpha}} = o(n)$, so the destination cannot be reached. (Notice that instead of $n^{1/3}$ and $n^{\frac{\alpha-2}{3\alpha}}$, we could have chosen other values $\beta > 0$ and $\gamma > 0$ such that the first n^β steps don't see a link of length greater than n^γ, so long as $\beta + \gamma < 1$.) ■

Based on simulations, it seems that the behavior is actually worse for $\alpha > 2$ than for $\alpha < 2$.

Notice that the entire asymptotic analysis holds for large n, i.e., n going to infinity. It does not apply necessarily for small finite n. Indeed, it has been verified experimentally that the correct exponent α for finite n is $\alpha = 2 - f(n)$, where $f(n) \to 0$ as $n \to \infty$. Determining the exact nature of $f(n)$ is still open.

In the analysis, it is important to notice that while, for $n \to \infty$, any poly-logarithmic function is exponentially smaller than any polynomial one, this may not be so for finite, even fairly large, n. This affects the observation above about the "right" exponent, and also may help in explaining the apparent gap between the short paths observed by Milgram, and the apparently much longer paths guaranteed by our proof. In addition, our analysis only used one long-range link per node. If nodes have many more long-range links, the predicted path lengths will be smaller.

While our analysis was done for a 2-dimensional grid, we did not use many properties of it. In fact, the only property we used was that there were $\Theta(d)$ nodes at distance d from a given one. The analysis extends easily to r-dimensional grids; the unique exponent for which local routing can be accomplished is then $\alpha = r$.

It also extends to hierarchical tree structures. For professions or other interests, a grid is perhaps not the right type of model. Instead, we may consider all professions to form a hierarchy ("scientific" vs. "arts" vs. "finance" etc., further subdivided into different sciences, subcategories, etc.). Each person is located at a leaf of this hierarchy, and generates $\Omega(\log^2 n)$ links to other people. The probability of linking to a given node decreases in the distance between the two nodes in the tree. It can be shown that for an appropriate choice of distribution, we still obtain local routing, and the distribution is closely related to the one studied above, in that it satisfies "uniformity over scales".

Kleinberg [2] proves a more general result for a certain class of set systems which includes metric spaces that "grow uniformly". Whether a similar result holds for arbitrary metric spaces is currently open.

References

[1] J. Kleinberg. The small-world phenomenon: An algorithmic perspective. In *Proc. 32nd ACM Symp. on Theory of Computing*, pages 163–170, 2000.

[2] J. Kleinberg. Small-world phenomena and the dynamics of information. In *Proc. 13th Advances in Neural Information Processing Systems*, 2001.

[3] D. Watts and S. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393:440–442, 1998.