**Lecture Notes**: Social Networks: Models, Algorithms, and Applications
**Lecture 9**: February 14, 2012
**Scribes**: Mauricio Monsalve and **Yamini Mule**

# 1 Descentralized search and $\alpha$

Reading: Chapter 20 from Kleinberg and Easley[1].
Paper: "Geographic Routing on Social Networks", PNAS 2005, Liben-Nowell et al.

**Kleinberg model** Let us consider the Kleinberg model $K(n, r, q, \alpha)$ where $n$ = no. of nodes, $r$ = radius for short range edges per node, $q$ = no. of long range edges added to each node, $\alpha$ = parameter for the probability distribution of long range edges.

**Results** The two results mentioned are as follows.

- If $\alpha = 0$ (i.e. other end points are chosen with uniform probability), then expected path length we obtain from any is $O(n^{2/3})$.

  This is a negative result of Watts-Strogatz's model with in the network there is no grade of how to proceed.

- If $\alpha$=2 then expected path length we get from GEOGREEDY is $O(\log^2 n)$.

As we tune the parameter $\alpha$, we get models that are quite different to do decentralized models.

**Some Intuitions** Why does $\alpha$=2 make decentralized search efficient?

Imagine a circle with vertex v, ball of radius d centered at 'v' and ball of radius $2d$ centered at 'v'. The node is annulus simple because area is $\theta(d^2)$. What is the probability that there is a long range path between some vertex u at distance $d$? Observe that at distance $d$ lie approximately $d^2$ vertices and that the probability of a long range edge falls with $1/d^2$. Then, when computing the probability, summing $O\left(\delta(d^2)\right)$ elements of probability $O(d^2)$ cancels out the distance d terms. Therefore, that probability is approximately constant or, better said, upper bounded by constants.

**Key point** Grid points are separated by one point.

Imagine we are sitting on $v$. What is the long range edge from $v$ to some node in annulus?

Constant $\leq$ probability [there is a "long range" edge from v to some node in the annulus] $\leq$Constant. The constants are independent of $d$.

**Observation** We will see what happens if $\alpha$ is altered. When $\alpha = 2$, long range edges of different "scales" are equally likely. When $\alpha < 2$, "short" long range links become unlikely. We favor longer jumps as $d$ increases as number of points increase. If $\alpha = 0$, longer long range edges (edges added at random base graph) are much more likely. Then we no longer have shorter long range links. When $\alpha > 2$, long range links become unlikely as the probability falls quite rapidly as we increase the distance $d$, the probability of getting to annulus falls as the annulus size grows. So, the early long range links observed in Milgram's experiment become hard to happen.

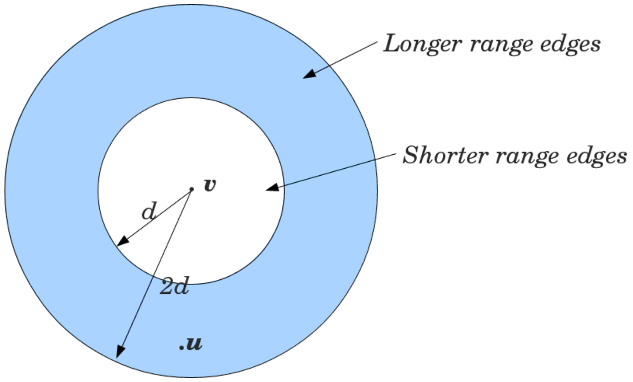

Figure 1: Sketch for the *heuristic proof* for the case of $\alpha = d^{-2}$.

## 2 The Live Journal Case Study

Do the questions we have been asking make sense for an online social network? This section is based in the PNAS paper[1].

There is a large social network associated with geographic network. This data was gathered during Feb 2004. At that time the numbers of bloggers are 1,312,454 in the live journal community. A blog typically contains reports on the user's personal life, reactions to world events and commentary on other blogs. In Live journal system, each blogger provides a profile, including his or her location, interests, and a list of bloggers whom he or she considers to be a friend. Of these 1.3 million bloggers, there are 495,836 in the United States who list a hometown and state that we find in the United States Geological survey and thus able to map to a longitude and latitude. The resolution of our geographic data is limited to the level of towns and cities. These towns has average population of about 1,305 where 80% of nodes of friendship relations are reciprocal and approximately 80% of nodes are in a connected giant component. The clustering coefficient which is high (0.2) is a characteristic of social networks. The degree distribution is heavy tailed, possibly log-normal.

The Live Journal social network displays a surprising and variable relationship between geographic distance and probability of friendship. The network shows short paths discoverable by using geography alone, even though existing models predict the opposite. We present rank based-friendship as it is unique in providing two desirable properties simultaneously: (i) it matches the experimental observations regarding the relationship between geography and friendship and (ii) it has a mathematical proof that networks exhibiting rank-based friendship will contain discoverable short paths. The Live Journal network exhibits rank-based friendship and does not contain discoverable short paths. Thus we can say that rank-based as a mechanism that has been empirically observed in real networks and theoretically guarantees small-world properties. Rank-based friendship explains geographic routing to a destination city; our data do not allow conclusions about routing within a city. On average about one-third of Livejournal friendships are independent of geography and may derive from other dimensions, like occupation and interests. These edges may play a role in local routing within the destination city and may also supplement the geographic links in global routing to the city.

**Geo-routing Experiment** Question: On this network is decentralized search efficient?

Geo-routing Experiment: In our simulation, messages are forwarded by using the geographically greedy routing algorithm GEOGREEDY [3]. If a person 'u' currently holds the message and wants to eventually reach a target 't' then he/she considers her set of friends and chooses as the next step in the chain the friend in this set which is geographically closest to 't'. If 'u' is closer to the target than all of his/her friends, then he/she gives up, and the chain terminates. When sources and targets are chosen randomly, we find that the chain successfully reaches the city of the target in approximately 13% of the trails, with a mean completed chain length of slightly more. We focus on global routing in these simulated experiments. For a target t' chosen uniformly at random from the network, the average population of t's city is 1,306 and always under 8,000. Therefore, the success of geography in narrowing the search from 5,00,000 users across the United States to the on average 1,300 residents in a particular city.

GEOGREEDY with "random reset": A success rate of 13% in this simulated experiment shows a surface similarity to Milgram's original experiment [2]. This experiment, however routes messages only to the destination city and does not suffer problems of voluntary participation. Our simulated participants have a much narrower choice of actions, as they are restricted to friends geographically closest to the target. We conclude that, even under restrictive forwarding conditions, geographic information is sufficient to perform global routing in a significant fraction of cases. This simulated experiment may be taken as a lower bound on the presence of short discoverable paths. We modify the routing algorithm as an individual 'u' that has no friend geographically closer to the target instead forwards the message to a person selected at random from u's city. Under this modification chain completes 80% of the time with average path length 12. The completion rate is not 100% because a chain may still fail by landing at a location in which no inhabitant has a friend closer to the target.

**Claim** Live Journal social networks permits efficient decentralized search.

Note that the heterogeneity of town populations makes Kleinberg's model not so usable. It is not easy to see why decentralized search still works.

**The Geographic Basis of Friendship** Friendship is more likely within geographic proximity and even with the internet. It is still an important factor deciding a contact.

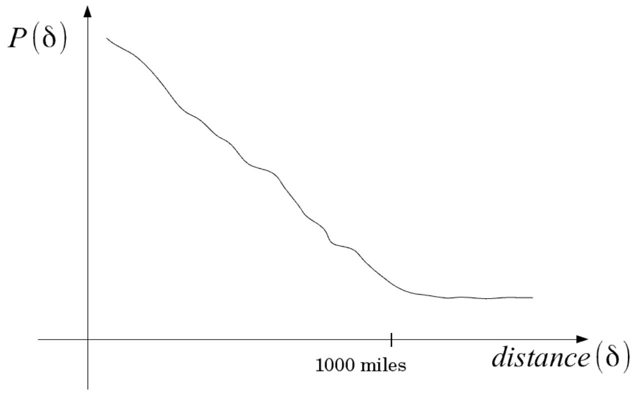

Figure 2: The geographic basis of friendship

Geographically, greedy routing in the Livejournal social network under a restrictive model allows 13% of paths to reach their destination city. Restrictive global-routing scheme enjoys a high success rate. So a question arises i.e., is there a model that explains efficiency of decentralized search in population of highly varying density?

The answer lies in moving two dimensions to the right dimensional space. A simple model explains the phenomenon. The relation between friendship and geographic distance in the Livejournal network is shown in the figure. For distance $\delta$, let $p(\delta)$ denotes the proportion of pairs u, v separated by distance $d(u, v) = \delta$ who are friends. As $\delta$ increases, $p(\delta)$ decreases, indicating that geographic proximity indeed increases the probability of friendship. However, for larger distances the curve flattens to a constant probability of friendship between people regardless of the geographic distance between them.

They basically fit one over $\delta$ fix well. Notice that it is not $1/(\delta^2)$. So this is inverse linear rate. This is not really a contradiction because Kleinberg model doesn't apply here due to high variable of density. Watts-Strogatz give a compelling model of social networks, simultaneously accounting for the presence of short connecting chains and the high clustering coefficients of real social networks, but the goal of this model is to explain the existence of short paths rather than to give an explanation of navigability. A major step towards explaining this is taken by Kleinberg. He modeled social networks by a $k$-dimensional grid of people, where each person knows his immediate geographic neighbors in every cardinal direction, and the probability of a long-distance link from $u$ to $v$ is proportional to $1/d\ (u,\ v)\ \alpha$, for some constant $\alpha \geq 0$. Kleinberg showed that short paths can be discovered in these networks if $\alpha = k$; more surprisingly, he proved that this is the only value of $\alpha$ for which these networks are navigable. He has subsequently generalized the model and results to an abstract characterization of group structures in social networks.

**Long range links** They are looking for another way to explain long range links by ranking the rest of the world as distance. We have the nearest individual, the second nearest, and so on, and label their distance as 1, 2, etc.

Pick the other end point w of a long range link $(v, w)$ with a probability proportional to $1/rank_v(w)$.

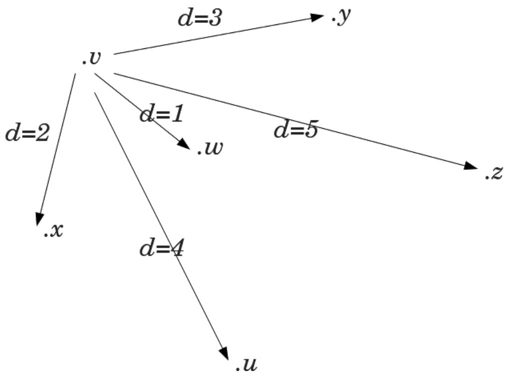

Figure 3: Ranks and distance

Note that Kleinberg model is a specific case of this rank-based model. Thinking in terms of ranks, their distances respect the ranks.

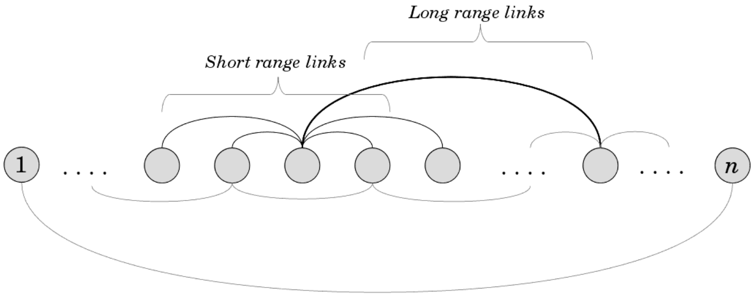

Figure 4: Kleinberg model is a special case of this rank-based model.

**Theorem 1** *Let $N$ be an arbitrary $k$-dimensional grid of locations, and let $P$ be an arbitrary $n$-person population on $N$ with rank-based friendship. For an arbitrary source person $s$ and a target person $t$ chosen uniformly at random from $P$, let $\rho$ denote the path from $s$ to the city of $t$ found by GEOGREEDY. Then the expected length of $\rho$ is at most $clog^3n$, for a constant $c$ independent of $n$, $N$, $s$, and $P$ but dependent on $k$.*

**Proof:** Consider a message traveling from source person $s$ to target person $t$ by using GEOGREEDY. We claim that if $t$ is chosen uniformly at random from $P$, then the expected number of steps before the message reaches a person within distance $d(s,\ t)/2$ of $t$ is at most$c$·log2 $n$. After the distance to $t$ is halved log $n$ times, the message will have arrived at its destination. Thus the expected number of steps to reach $t$ is at most $c$·log3 $n$. To prove the claim, we show that the probability that a person forwards the message to someone within the small $d(s,\ t)/2$-radius neighborhood around $t$is at least 1/log $n$ times the relative densities of the small neighborhood around $t$ compared with a larger neighborhood containing both $s$ and $t$. By taking expectations over $t$ and appropriately approximating the densities of these neighborhoods, we show that the expected number of steps before the message reaches the small neighborhood around $t$ is at most $c \log 2n$. □

There is significant evidence from real-world message-passing experiments that an effective routing strategy typically begins by making long geography-based hops as the message leaves the source and ends by making hops based on attributes other than geography. Thus there is a transition from geography-based to nongeographic-based routing at some point in the process. They plotted that ratio and obtained something cool, proving that decentralized search is still efficient, yet people don't select their friends thinking about the efficiency of decentralized search.

### To think about

- Asymmetry of decentralized search. People have thoughts about this. It is not the same way to go from Washington to Omaha than to vice versa. Depending on the structure of social networks, the status of target has significance.
- Distance may not be geographic at all. There can be more general definitions of distance like social status, community membership, etc.
- In the paper, west coast friendship and east coast friendship are different. East coast is much more geographically bounded.

## References

[1] D. Liben-Nowell, J. Novak, R. Kumar, P. Raghavan and A. Tomkins. Geographic routing in social networks. PNAS, 102(33):11623–11628, August 2005.
[2] Milgram, S. (1967) Psychol. Today 1, 61–67
[3] Kleinberg, J. M. (2000) Nature 406 , 845.