The small-world phenomenon

13.1 Six Degrees of Separation

Milgram Experiment
Each person in the experiment are requested to send a mail to a "target". Rather than directly mailing, participants were asked to send mail to one of his acquaintances, and the latter will send to his acquaintance and so on, until it reaches target. The selection of the "passer" is based on the goal to deliver the mail to the target as soon as possible. The result of the experiment shows that one third of the mail finally reached target, with median of six steps. It shows that short paths exist in social network.
Two principles to be developed:

- Short paths exist.
- Short paths can be found.

13.1.1 A Simple Model of Six Separation

Suppose everyone knows 100 people, so after 5 steps away, one can reach 100^5 people, which is larger than the population of the world.
This model doesn't consider overlap of people's acquaintance and will only build a tree structure for the population, which is naive.

13.1.2 Watts-Strogatz Model

It can be modeled as an (r, k) network as Fig. 13.1. In Fig. 13.1, any node is located in a grid, and connected to all the nodes within r radius away from it. The grid can be viewed as geographic proximity. Besides those nodes in r radius, a node can also uniformly randomly link to k other nodes in the grid. We let k to be sufficient small, so within i steps, a node u can reach $\Theta(k^i)$ nodes.

13.2 Decentralized Search

Decentralized search is the procedure that search for a short path from node s to t. Suppose s want to send a message to t, it must first send the message to one of his contacts. **The only knowledge**

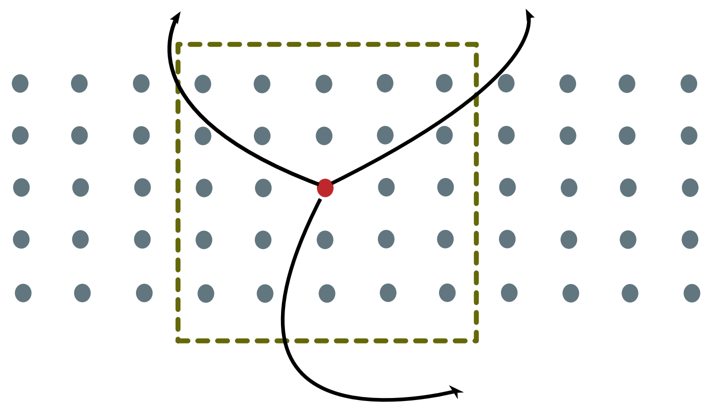

Figure 13.1: Watts-Strogatz Model

s has is its contacts in r radius, its k random links, and the position of t in the grid. After he pass the message to one of the contacts w, it is now w's turn to select his contact, following what s has done.
It is found that decentralized search in Watts-Strogatz network will fail most of the time. The reason is that the k links in the model are generated too randomly. To assure the success, we'd better have some clue of the "gradient" of the links in the grid.
We first generalize the previous (r, k) model into (r, k, q) model.
Here, each node still has link to all other nodes in its r radius range. However, the k random edges is generated in a way that decays with distance, controlled by q. Suppose $d(u, v)$ denotes grid distance between node u and v, the probability of an edge existing between them is proportional to $d(u, v)^{-q}$.
To illustrate how decentralized search is applied, we give a simple (r, k, q) model. Rather than two dimensional grid we mentioned before, here we propose a one dimensional grid for us to place the nodes on. Further, let $r = 1, k = 1, q = 1$. So node i has two local contacts $i-1$ and $(i+1)\%n$, where n is the total number of nodes. i also has one long-range contact. The model is called augmented ring and is shown in Fig. 20.11, 20.12 and 20.13 in the text book.

13.2.1 Myopic Search

We will analyze myopic search on the simple augmented ring we propose previously, which is one version of decentralized search. In myopic search, each nodes will pass the message to one of its three contacts, which is the closest to target t, in the sense of grid distance.
Let the number of steps of passing a message from node s to t is a random variable X. Our goal is to show that $E(X)$ is relatively small, in other words, there is short path from s to t, and can be constructed by myopic search.
We also define **phase** j: In phase j, the search in which the message's distance from the target is between 2^j and 2^{j+1}. So $j < \log_2 n$(will be denoted simply $\log n$ below).
Then we obtain:

$$X = X_1 + X_2 + ... + X_{\log n} \tag{13.1}$$

Here X_i is the number of steps in phase i.
We will show that $E(X_i) = \theta(\log n)$, and $E(X) = \theta((\log n)^2)$.
Let the probability of existing an edge between u and v is $(d(u,v)^{-1})/Z$, and we know that there is at least 1 edge from u to any other nodes, so we get:

$$\begin{aligned}
&\sum_{v \in V, v \neq u} (d(u,v)^{-1}/Z) \geq 1 \\
&\Rightarrow Z \leq 2(1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + ... + \frac{1}{n/2}) \\
&\Rightarrow Z \leq 1 + \ln(n/2) \leq \log n
\end{aligned} \tag{13.2}$$

So we obtain the probability of u links to v is $\frac{(d(u,v)^{-1})}{Z} \geq \frac{(d(u,v)^{-1})}{\log n}$.
Following we will analyze the probability of message passing in phase j for at least i steps.
Suppose node u is the median node for passing message from s to t, and it is in phase j, with distance to t as d. If next step, message will be passed out of phase j, it must go into the $d/2$ range of t (due to the distance from phase j to t being 2^j), which lies on the left and right side of t. Considering the farthest nodes in this range is $3d/2$ from u, the probability of there is at least one of them link to u is:

$$d \cdot \frac{(d(u,v)^{-1})}{\log n} \geq d \cdot \frac{2}{3d \log n} = \frac{2}{3 \log n} \tag{13.3}$$

So the probability of staying in phase j for at least i steps is:

$$(1 - \frac{2}{3 \log n})^{i-1} \tag{13.4}$$

Finally we obtain:

$$\begin{aligned}
&E(X_j) = 1 \cdot Pr[X_j = 1] + 2 \cdot Pr[X_j = 2] + 3 \cdot Pr[X_j = 3] + ... \\
&= Pr[X_j \geq 1] + Pr[X_j \geq 2] + Pr[X_j \geq 3] + ... \\
&\leq 1 + (1 - \frac{2}{3 \log n}) + (1 - \frac{2}{3 \log n})^2 + ... \\
&= \frac{3 \log n}{2}
\end{aligned} \tag{13.5}$$

From Eq. 13.1, we get that $E[X] \leq \frac{3(\log n)^2}{2}$. From above analysis, we construct a short paths from s to t which has a upper boundary much smaller than the number of nodes.