Lecture Notes: Social Networks: Models, Algorithms, and Applications
Lecture 1: Jan 17, 2012
Scribes: Preethi Ambati and Azar Aliyev

Example 1: An algorithmic view of the small world phenomenon

The story starts with Stanley Milgram, an experimental psychologist, who did a lot of work in 50s-60s and was very well known for innovative experiments. One such popular experiment was the small world experiment. This experiment was conducted in 1967 [1], and was one of the first and best known of several experiments that Milgram conducted.

Experiment: Milgram sent 96 packages to a group of randomly selected people in Omaha. They were randomly selected by looking up the telephone directory. Each package contained the following:

1. An official looking document that looks like a passport. It had Milgram's home institution's(Harvard) logo on it.
2. Name, address and occupation(stock broker) of Milgram's friend in Boston, MA.
3. Instructions to get the package to the destination (Milgram's Friend) by following the rule where each person can only send the package to an acquaintance. An acquaintance was defined as being on "first name basis".

Each individual that participated in this experiment was also asked to write a trace i.e., the details of the individual that they were sending the package to.

By looking at the completed paths(the paths of packages that reached the destination), Milgram was able to figure out the number of hops each package took to reach the destination. Finally the average hop length was computed from this data.

Results: The results of Milgram's experiment were as follows:

1. 18 packages reached destination.
2. The average hop length was 5.9.

The second result is the origin of the popular notion of "six degrees of separation" [1].

[1] Six degrees of separation refers to the idea that everyone is on average approximately six steps away, by way of introduction, from any other person on Earth. Refer to [2] where average path length[2] to Kevin Bacon was calculated by looking up the information from IMDB.

[2] Average path length of the network[3] is calculated by finding the shortest path between all pairs of nodes, adding them up, and then dividing by the total number of pairs. This shows us, on average, the number of steps it takes to get from one member of the network to another.

[3] Network is formed by assigning each actor to a node and by adding an edge between two actors if they have appeared together in a movie.

Problems: There were many problems with this experiment, few of which are mentioned below:

1. Only a small fraction of the original number of packages reached the destination. What happened to the rest of the packages?
2. The incomplete paths(those packages that did not make it) might correspond to longer paths. So, taking into account only the completed paths could result in under estimation of the average path length.
3. All sources came from a single city.
4. Destination was just one individual.

Despite the aforementioned shortcomings, the results of this experiment are widely accepted now. Many versions of this experiment have been performed since then. For example, in a recent study that was done on the Facebook platform application named "Six Degrees" by Karl Bunyan [3], the degrees of separation between different people was calculated. Facebook had over 5.8 million users as seen from the group's page. The average separation for all users of the application is 5.73 degrees whereas the maximum degree of separation is 12. Variants of the experiment such as these confirm that average path length in social networks tend to be a small concept. Furthermore, this result has been observed in other kinds of networks like technological networks, biological networks etc.

In fact, a well-known paper by Watts [4] has a table that contained observations relevant to the "small world experiment" of Milgram. Ignoring the clustering coefficient and some other columns from the original table, a snapshot of it looks like below:

	Number of Nodes/Vertices	**Average Degree**	**Average Path Length**
Film actor network	225,226	61	3.65
Power grid network	4,941	2.67	18.7
C.Elegans	282	14	2.65

The above table contains information about the following networks:

1. Film actor network - In the film actor network, nodes are film actors and two actors are connected by an edge if they acted together in a film(the list of film actors is drawn from The Internet Movie Database).
2. Power grid network - In the power grid network, the nodes are substations and generation stations whereas the edges are high voltage lines.
3. C.Elegans [4] - It is a simple-minded worm that has about 300 neurons in its brain. The nodes in this network are neurons and edges are the synapses.

All of the networks mentioned above are relatively sparse. In other words, their average degree is small compared to the large number of nodes/vertices. Despite this paucity of edges, the average path length is quite small. Therefore the notion of social networks being small world networks extends to other kinds of networks.

[4] Caenorhabditis elegans (C. elegans) is a free-living, non-parasitic soil nematode worm of length of 1.2mm, with almost transparent body and lives in soil. Its nervous system consists of 302 neurons, each of the size approx. 5 μm. C. elegans is one of the simplest organisms with a nervous system.

Based on Milgram's original experiment and other subsequent experiments, the notion of small world experiment has become well established. Kleinberg [5] observed that Milgram's experiment contains algorithmic discovery as well. The observation made by Kleinberg is this: not only is Milgram's experiment saying that an average of 6 hops are needed to get from Omaha to Boston, but these short paths are actually being discovered by the individuals in this network with very limited information.

A question that naturally follows this observation is this: can we come up with models of social networks that not only display the "small world" property, but also possess the property that decentralized(local) search can lead to the discovery of short paths? If we take Milgram's experiment seriously and think about it, maybe we need models that are not just capturing the topological features of "small world" networks but also algorithmic properties like the one mentioned above.

In Kleinberg's paper, there is a negative result and a positive result.
Negative result: In the Watts-Strogatz random graph model [6], decentralized search can produce paths that are arbitrarily long relative to the distance between pairs of nodes.
Positive result: Consider the following network which is a grid of 4X4 size (arbitrary size). Each

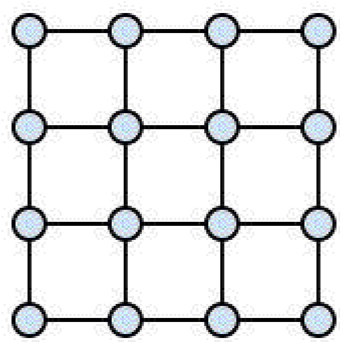

Figure 1: 4X4 Grid

node gets k edges, connecting each node at random to other nodes. K here is a parameter of the model. Each node u is connected by a long range link to v with probability proportional to $d(u, v)^{-q}$, where q is another parameter that is ≥ 0 and $d(u, v)$ is a distance between nodes u and v. In our case, the distance is the shortest path between nodes. Now consider the following two cases:
Case1: When q = 0, v is chosen uniformly at random. This is basically the Watts - Strogatz model. In this case, long range links are "too random". They are structurally helpful (i.e., they make a "small world" graph) but they are not algorithmically helpful.
Case2: When q is very large, randomness is no longer present because all the "long range" links from a vertex u connect to vertices that are close by.

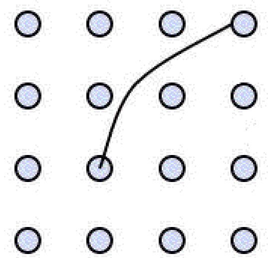

Figure 2: Long range link

In either case (when q = 0 or q is very large), decentralized search does not work. Now the question is - "is there a sweet spot?" i.e., a value of q for which the network is neither too random nor too deterministic. For q = 2, decentralized search is successful. This is the positive result.

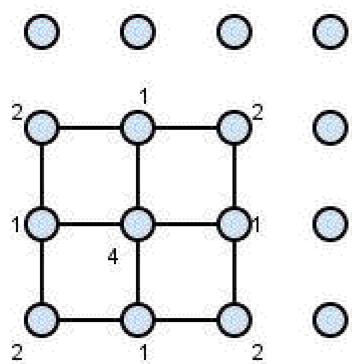

Figure 3: q = 2

Example 2: Equilibrium traffic on networks

Consider a road network. The diagram figure above shows the morning rush hour traffic. There are roads in other directions but we ignore them. Roadways have labels and each label shows commute time on that roadway as a function of the number of cars on that roadway. In the diagram above, x denotes the number of cars on a roadway.
Now suppose that 4000 cars need to enter at A and get to B.
One possible solution would be as below:
Send 2000 cars in the direction $A \to C \to B$ and the remaining 2000 cars in the direction of $A \to D \to B$. With this approach every car gets to B from A in time $(\frac{2000}{100}) + 45 = 65$ units. As every car will spend the same amount of time on a roadway we will achieve a social optimum solution. The social optimum solution here is obtained by modelling the problem as an optimization

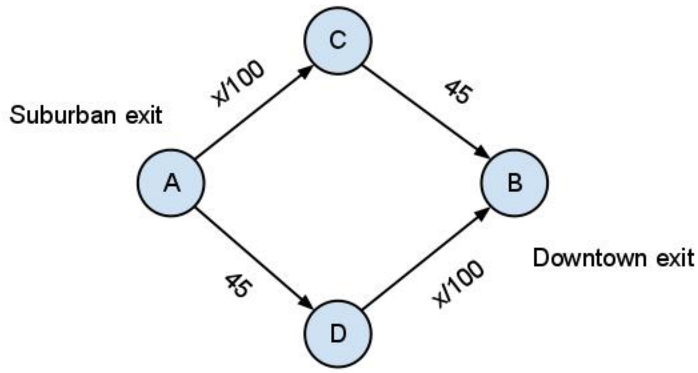

Figure 4: Road Network

problem that a ”central planner” would solve.
The other solution would be as below:
This solution is obtained by modelling each driver as a rational, self-interested individual, such that he knows the entire network. However the driver does not know what decision other drivers will make. If we have a solution different from the first solution (the social optimum solution), at least one driver will have an incentive to deviate. i.e., send some number of cars (a number other than 2000) in upper path and the rest along lower path. Let’s say 1500 cars are sent in the upper

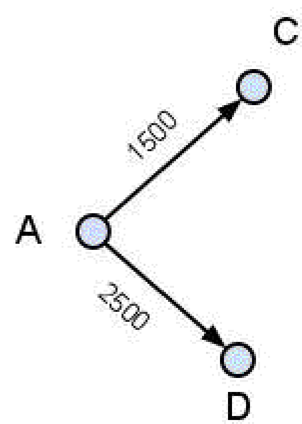

Figure 5: Number of cars along upper and lower paths

path, 2500 cars are sent in the lower path. Then the drivers travelling on the lower path are not playing their ”best response”. The solution in which all drivers are playing their best response is

a "Nash Equilibrium." [5] So in this example, the unique social optimum is identical to the unique Nash Equilibrium.
Now to the same road network add a new route from C to D. This adds capacity to the network. For simplicity, lets assume the commute time from $C \rightarrow D$ is zero. The network now looks like below. By adding an edge, things only get better for social optimum solution. But what happens

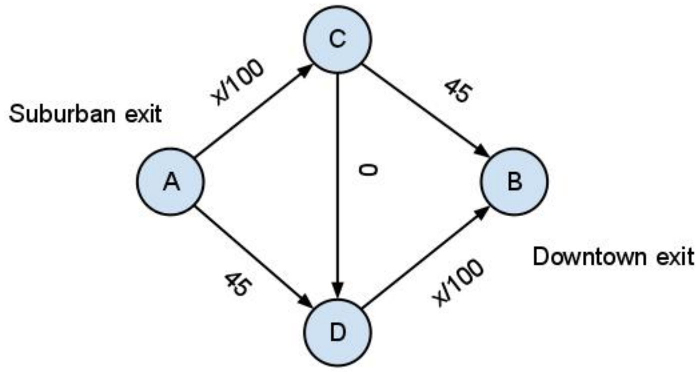

Figure 6: New road network

to the Nash Equilibrium solution?
Now, as the best response for every driver will be the route $A \rightarrow C \rightarrow D \rightarrow B$, all 4000 cars will take it and the commute time per car will be $(\frac{4000}{100}) + 0 + (\frac{4000}{100}) = 40 + 0 + 40 = 80$.
As we see, adding capacity to the network has made the Nash Equilibrium solution worse. This solution is Breasss Paradox [7]. Now, there is a gap between social optimum (≤ 65) and Nash Equilibrium (80). The ratio of Nash Equilibrium cost to the social optimum is called the *price of anarchy*[6]. This was the result of work by Tim Roughgarden and Eva Tardos [8].

Price of anarchy $= \frac{NashEquilibriumcost}{SocialOptimum} \geq \frac{80}{65}$

Tim Roughgarden and Eva Tardos showed in their research that the price of anarchy is $\leq \frac{4}{3}$

References

[1] Jeffrey Travers and Stanley Milgram. An Experimental Study of the Small World Problem. Sociometry (1969).

[2] http://oracleofbacon.org/

[5] Nash Equilibrium is named after John Forbes Nash, who proposed it. The Nash Equilibrium is probably the most widely used "solution concept" in game theory. If there is a set of strategies with the property that no player can benefit by changing her strategy while the other players keep their strategies unchanged, then that set of strategies and the corresponding payoffs constitute the Nash Equilibrium.

[6] The Price of Anarchy is a concept in game theory that measures how the efficiency of a system degrades due to selfish behavior of its agents.

[3] http://en.wikipedia.org/wiki/Six_degrees_of_separation

[4] Watts and Strogatz. Collective dynamics of ”small world” networks. Journal Nature (1988).

[5] Kleinberg. Complex networks and decentralized search algorithms. ICM (2006).

[6] Newman, Watts, Strogatz. Random graph models of social networks. In proceedings of the National Academy of Sciences of the United States of America (2002).

[7] Braess. Uber ein Paradoxon aus der Verkehrsplanung Unternehmensforschung (1968).

[8] Tim Roughgarden and Eva Tardos. How bad is selfish routing? ACM (2002).