Mapping the Interactions of the World

By Michele Albach
March 9th 2014

A Summary of Network Study and Application

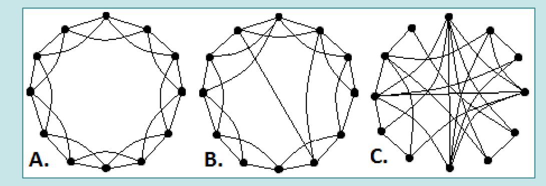

Figure modified from Watts (2003).

Figure 3 Network A is not rewired, or was rewired with a *p* of 0, network B was rewired with a *p* of 0.1 and C with a *p* of 1.

1. What's the Big Deal with Networks?

A network is a system that can be represented by individual nodes, each connected by edges that signify contact. A network is a simple mathematical system that can be used to model all sorts of natural real world interactions, from connected brain neurons to the internet to human interactions. In order to represent different types of interactions, networks can be directed like the hyperlinks that make up the World-Wide-Web, or undirected like connected power stations in the electrical system. One important aspect of networks is their tendency to have a largest component. Components are sections of a network that are totally unconnected from each other, and most of the time networks will have one component that connects over 90% of the nodes. This is because if there are multiple large components, each with high amounts of nodes, it is very likely for a connection to form somewhere that will connect the two (Newton, 2010). Examples of undirected and directed networks, as well as a largest component, are shown in Figure 1.

Figure modified from Watts (2003).

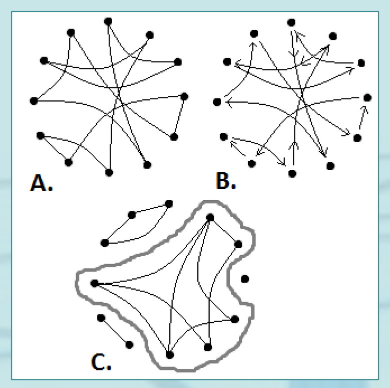

Figure 1. Network A shows an undirected network, network B a directed network, and network C has its largest component circled

3. Snail Mail, Email, and Shortest Path Length

Inspired by the small-world problem, Stanley Milgram and Jeffrey Travers (1969) popularized network study in 1967 by sending out and tracking the routes of letters. In order to determine if anyone in the world could be reached in a few short steps, Travers and Milgram distributed 96 letters that were all addressed to the same person. With these letters were instructions. They told the name, occupation, and geographic location of the target person and asked the letter holder to mail the letter to someone they knew on a first name basis, that they thought was closer than them to the target. Each person was also asked to keep track of who had had the letter. Of the 96 letters, only 18 were delivered to the target. However those 18 letters had been passed through an average of only 5.9 people. Travers and Milgram called this number the world's shortest path length.

In a network, the shortest path length is the average of the smallest amount of steps that can be taken to get from any node to any other node. Travers and Milgram found this to be approximately 6, supporting the already circulating theory of six degrees of separation as well as the small-world theory. This was huge news, and was not questioned for 30 years until a psychology professor looked into the paper. She found that all of the letters were originally delivered to a similar geographic location, and that the distance travelled by the letters was not far enough to count the results as a shortest path length for the whole world (Kleinfeld, 2002). However, other research has been done to since confirm the work of Travers and Milgram. Dodds *et al.* (2003) sent out 60 000 emails to find one of 18 targets with similar instructions to the original experiment. Despite a very low completion rate, the average path length of delivered emails was found to be between 5 and 7 people.

4. Maintaining Clustering

The next big step in the world of network study came with Watts and Strogatz (1998). They noticed that natural networks tend to not only have shortest path lengths, but also be clustered. In order to quantify clustering, they defined the clustering coefficient as the average ratio of all linked pairs of connections to a node over all possible pairs of connections. What this measures in terms of human interaction is what fraction of the pairs of your friends are also friends. Watts and Strogatz wanted to find a network that has a small shortest path length with a large clustering coefficient, as this would model real networks that seem to be small but are still clustered. They implemented an algorithm that starts with a regular lattice and rewires it with a probability of *p*, as shown by Figure 3. Upon analyzing networks of increasing disorder (rewired with an increasing *p*), they found that both shortest path length and clustering coefficient decrease. However, they noted that when an intermediate *p* is used, shortest path length is already low with a still relatively high clustering coefficient. This was because shortest path length decreases with an increasing *p* at a much faster rate than clustering coefficient, as shown by Figure 4. As a result, an intermediate *p* (about 0.1) causes small-world networks, which exhibit small path lengths and high clustering.

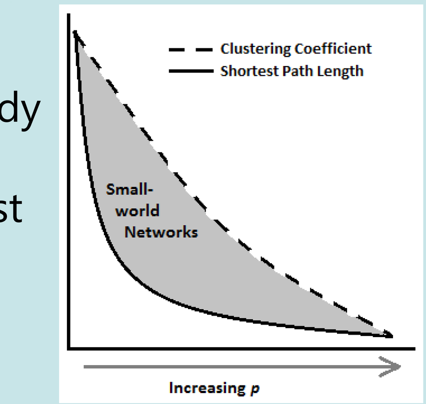

Figure modified from Watts (2003).

Figure 4. The shaded area represents small-world networks due to a clustering coefficient that is still relatively high but a low shortest path length.

2. The Small-World Problem

Imagine you're at a cocktail party. You meet a stranger and begin to make conversation, and soon enough you realize that your new friend went to high school with your current boss. You both casually remark that it's a small world! But is it really? The whole world has a population of over 7 billion people, how is it that we say that it is a small world without giving it a second thought? Say this party was in New York. Out of the 8 million people that make up New York, what's the chance that you and the person that you happen to start talking to at a party share an acquaintance?

This is what was commonly known as the small-world problem, and was a circulating social theory in the late 1960s. The theory stated that all social connections could be modelled as a network, and that any person in the world could be reached through a small number of steps (Watts, 2003). This theory is also known as the 'Six degrees of separation' theory explained by Figure 2 which was popularized by a play written by John Guare.

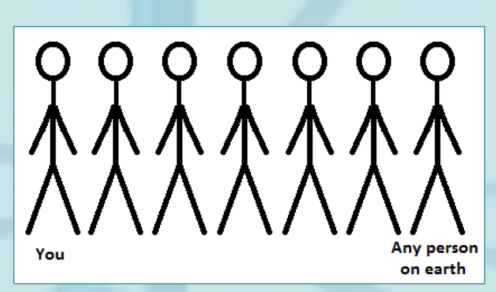

Figure 2. The expression '6 degrees of separation' which is an example of the small-world problem is represented visually. People standing beside each other represent acquaintances.

The picture on the right is a network representation of the connections that make up the internet. Each node is a class C subnets group of computers, which is just a group of computers with similar internet addresses typically managed by a single organization. When you use the internet, information doesn't just jump from one computer to another. Information has to go from node to node, similarly to any other network, to get to its destination.

Picture taken from: http://openmatt.org/tag/web/

5. The Rich Get Richer

Watts and Strogatz missed an important feature of natural networks; the property that nodes with high degrees (having lots of connections) tend to be more likely to earn new connections than nodes with few connections. This is referred to as preferential attachment or homophily, and is commonly described by sociologists as the 'rich get richer' effect. The result of preferential attachment is a high possibility of extremes. For example, the distribution of wealth across the United-States shows preferential attachment, resulting in most families having similar amounts of money with the odd family having very low amounts or very high amounts.

Mathematically, this changes the distribution curve used to describe the degrees of nodes. It had been assumed that networks would follow the Poisson distribution, but Barabási and Albert (1999) showed that it actual follows a power law distribution resulting in the networks to be scale free. A scale free network allows for extremes, which is a better description of natural networks than what was proposed by Watts and Strogatz. A description of the Power Law, Poisson, and Normal distributions is shown in Figure 5.

References:

Barabási AL, Albert R. 1999. Emergence of scaling in random networks. Science. 286(5439):509-512.

Dodds PS, Muhamad R, Watts DJ. 2003. An experimental study of search in global social networks. Science. 301(5634):827-829.

Kleinfeld, JS. 2002. The small world problem. Society. 39(2):61-66.

Newman MEJ. 2010. Networks. New York: Oxford University Press; 772 p.

Travers J, Milgram S. An experimental study of the small world problem. Sociometry. 1969;32(4):425-443.

Watts DJ. 2003. Six degrees: the science of a connected age. New York: W.W. Norton and Company; 374 p.

Watts DJ, Strogatz SH. Collective dynamics of 'small world' networks. Nature. 1998;393:440-442.

6. Connecting the Dots

Our world is filled with networks, from biological networks like the neuron system to electrical networks like the power grid. Once we understand how to model the networks that surround us, we might be able to predict or even control them. For example, a better understanding of the network that makes up the internet could make it possible to prevent viruses, or an understanding of the power grid could prevent power outages. All the people mentioned above have made major contributions towards having this understanding. Travers and Milgram helped us to understand that networks are not quite as vast as they seem, but can actually be interpreted as small. Watts and Strogatz suggested ways in which this small world attribute can be maintained without losing clustering, which allows for a more accurate model of natural networks. Barabási and Albert determined the distribution that is followed by the degrees of nodes, allowing for a better understanding of power distribution. The next big step in network study is to examine the dynamics of a network. Natural networks are rarely static, but relationships are constantly forming and breaking. Being able to model a changing network is a step that will hopefully result in positive contributions towards understanding our world.

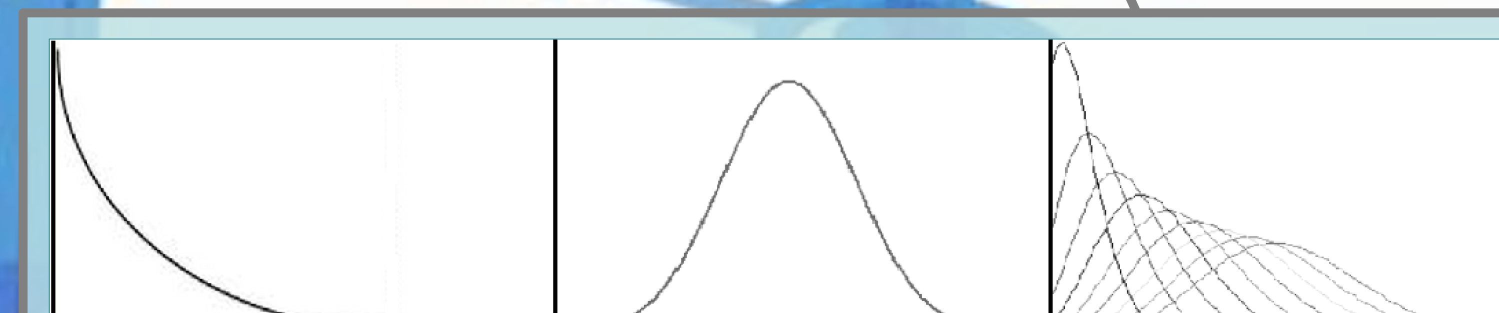

Figure 5. Examples of Power Law, Normal, and Poisson distributions. Each is displayed on a frequency graph. A normal distribution (center) is a distribution for a range of options where the average is the most likely to occur. The Poisson distribution (right) is an adaptation of the normal distribution used for when the mean is already known, and each possible outcome should be equally rare. Each separate line is representative of a different mean. The power law (left), which is what is followed by the degree spread in a network, is used to describe populations where most outcomes are the same, but there are extremes that would skew the data if represented in a normal distribution.