Mapping the Interactions of the World:

A Summary of Network Study and Application

By Michele Albach

April 7th 2014

An Introduction to Networks

Our world is filled with complex systems, and an easy way to model these systems is through networks. Networks are systems that can be represented by individual nodes connected by edges signifying some sort of connection. From human interactions, to electrical systems, to brain neurons, we are effected by the relations of networks every day. Networks can be undirected, like the connections represented by acquaintances when modeling human interaction, or can be directed, like the directed hyperlinks that make up the World-Wide-Web. The nodes of a network can be classified by their degree, which is the amount of connections protruding from it. Networks can also be defined as having components, meaning sections that are not at all connected to each other. Normally, a network will have a large component, which contains the majority (typically over 90%) of all nodes. It is highly unlikely for a network to have more than one large component, as each component will contain a large percentage of all nodes, and any one connection from a node in one component to the other will combine the components making one large component (Newman, 2010). This is important because it would suggest that natural networks, like human interactions, are mostly connected. In examples like the internet, if nodes were not all connected then it would sometimes be impossible to send an email from one component to another. Figure 1 shows examples of directed and undirected networks, as well as highlights the largest component in an example network.

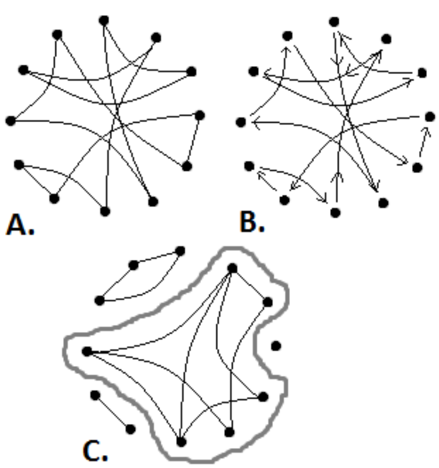

Figure 1. Network A shows an undirected network, network B shows a directed network, and network C has its largest component circled.

In order to study, model, and predict the interactions of networks, years of research have been contributed towards determining the qualities of natural occurring networks and creating a model that encompasses both the structural and dynamic aspects of these qualities. This research became popularized with a manual study of the social distance between any 2 people in the world using hand written letters (Travers and Milgram, 1969), and has progressed in many different directions. Examples of this include the development of a model encompassing a small world with clusters (Watts and Strogatz, 1998), the study of preferential attachment and the resulting scale-free networks (Barabási and Albert, 1999), and a more in-depth study of the searchability and dynamics of an evolving network (Kleinberg, 2000; Kossinets and Watts, 2006). This paper will outline the history of network research, describing each theory and experiment to eventually lead to the current best model of the natural network. Research will also be verified using rewritten code to obtain similar results. The applications of the developed model will then be described, as well as the theorized next step in network research.

Determining Shortest Path Length

The study of networks became prevalent when Travers and Milgram (1969) conducted an experiment in which they distributed 96 letters throughout the United States that contained the name, occupation, and place of residence of one stockbroker in Boston. The letter also contained instructions to mail the letter to someone that was known on a first-name basis by the letter holder whom was thought to be closer to the recipient, and to record the amount of people that the letter had been passed to. Of the 96 letters, 18 arrived at their final destination and had been passed through an average of 5.9 people. What Travers and Milgram had found was an approximation for the shortest path length of the world. The shortest path length is defined as the average distance in nodes traversed to get from one node to any other node. This led to the conclusion that a person was only 6 steps away from any other person in the world, and supported the already popular theory '6 degrees of separation' visualized by Figure 2. Although influential and thoroughly cited, Travers' and Milgram's experiment was proved 30 years later to be flawed. A psychology professor noted while attempting to recreate their experiment for a class project that their choices of letter recipients were not unbiased and were all in similar geographic locations (Kleinfeld, 2002). However, further research that has been conducted has found similar results despite faults in the original experiment to keep Travers' and Milgram's results valid (Newman, 2010). One of these repeat experiments was performed by Dodds *et al.* (2003), in which 60 000 email messages were sent out in an attempt to reach one of 18 targets. Despite a very low completion rate of 1.5% (thought to be due to the apathy of participants), the shortest path length was found to be between 5 and 7 people, confirming Travers' and in Milgram's research.

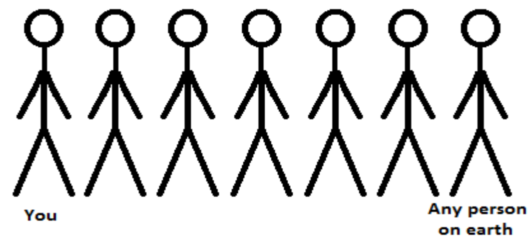

Figure 2. The expression '6 degrees of separation' supported by Travers' and Milgram's experiment is represented visually. People standing beside each other represent acquaintances.

Linking Friends of Friends

After determining that short path lengths do exist in the world, the next step was to design a network that models this. However, one other quality needed to be upheld; connections in the world are not random but tend to be clustered. In applicable terms, friends of friends are likely to also be friends. Watts and Strogatz (1998) defined clustering in a

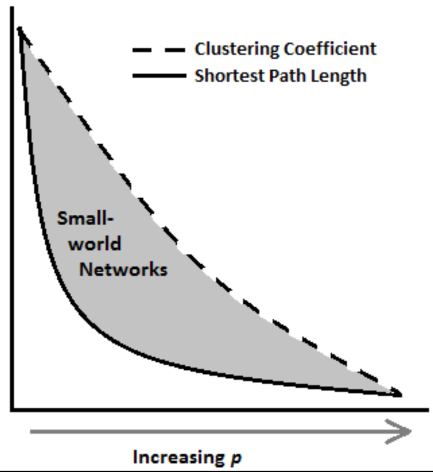

Figure 3. The shaded area represents small-world networks due to a clustering coefficient that is still relatively high but a low shortest path length. Figure is modified from Watts (2003).

measurable term called the clustering coefficient, which is calculated by taking the ratio of all linked pairs of connections to a node over all possible pairs of connections to a node. Upon studying this quantity in networks, they noticed that as networks go from ordered to disordered, both clustering and the shortest path length tend to go from high to low. But when an intermediate level of randomness is selected, high clustering and low shortest path lengths are exhibited. This is because shortest path length decreases at a much faster rate than clustering does, leaving a space of networks that have a low shortest path length without compromising high clustering, as shown by Figure 3. This is exactly what is exhibited in natural networks. This type of network was called a 'small-world network'.

In order to demonstrate this, Watts and Strogatz implemented an algorithm that starts with an ordered network similar to network A in Figure 4 and randomly rewires each connection with a probability of *p* (the increasing term in Figure 3). Figure 4 shows 2 other networks, network C was rewired with a high *p* value (1) and exhibits both low clustering and shortest path length, and network B was rewired with an intermediate *p* value (0.1) and exhibits the small world attributes (high clustering and low shortest path length). In order to test the hypothesis of

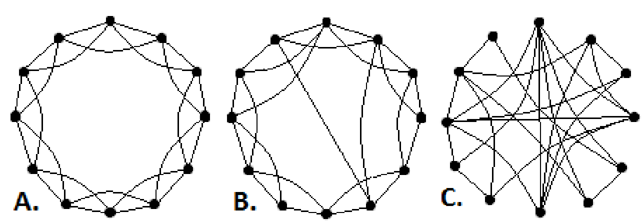

Figure 4. Network A is not rewired, or is created with a rewiring probability of 0, network B was rewired with a probability of 0.1 and C with a probability of 1. Modified from Watts (2003).

Watts and Strogatz, their code was rewritten with slight differences in the rewiring process but with the same basic method. The code was run for randomly rewired networks of 500 nodes and an average of 50 connections per node and the results are presented in Table 1. These results correspond with the results achieved by Watts and Strogatz, as a low rewiring probability yielded high clustering and path lengths, a high rewiring probability yielded low clustering and path lengths, and an intermediate probability yielded high clustering and low path lengths, demonstrating the small world phenomenon.

Table 1. The clustering coefficient and average shortest path length of networks with 500 nodes and 50 connections per node rewired at different probabilities.

Rewiring probability	Clustering Coefficient	Average Shortest Path Length
0	0.735	4.773
0.1	0.556	1.695
1	0.298	1.59

Spreading and Percolation

The primary application of the study of networks is network dynamics. Understanding how a rumour, a disease, or a fad percolates through a network allows for predictions to be made. This could contribute to society in many ways from helping to stop epidemics, or predicting what songs will become number one hits. The study of epidemics was initiated by Kermack and McKendrick (1927), in which they described how individuals go through 3 stages during an epidemic; susceptible, infected, and removed. Susceptible means a node does not have the condition, infected means a node has the condition and can infect other nodes, and removed means that a node can no longer infect other nodes (either due to death or immunity in terms of disease). A network or population as a whole will also go through three stages due to the three individual stages. First, most of the population will still be susceptible resulting in a slow rate of infection, called the slow growth phase. Then the numbers will even out and rate of infection will increase during the explosive phase, until the level of infected individuals is much larger than that of susceptible individuals, decreasing rate of infection during the burn out phase. Together, these result in a logistic growth rate as shown by Figure 5. The three individual stages were abbreviated to call this the SIR model of infection.

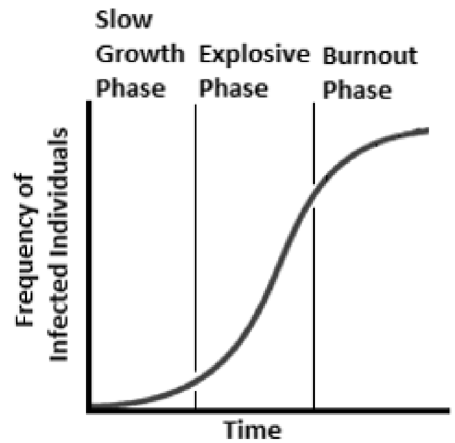

Figure 5. A model of the logistic rate of infection in the SIR model. Modified from Watts (2003).

Watts and Strogatz (1998) applied this model to their study of networks to find that small-world networks become infected at a much faster rate than non-random or totally random networks. They found that slight rewiring resulted in a couple connections across the network, like what is shown in network B of Figure 4. These connections were called bridges, because they became very frequently used in shortest path lengths, as they acted as a shortcut across the network. The result of adding a couple bridges was a great decrease in the shortest path length. The implication of this is that a very small amount of rewiring was needed to add shortcuts, decrease path lengths, and therefore increases the overall rate of infection.

The Rich Get Richer

Watts' and Strogatz's model failed to incorporate another important quality of naturally occurring network, called preferential attachment (Newman and Watts, 1999). Often referred to as homophily, this means that nodes tend to make new connections to other nodes that are similar to it. Kossinets and Watts (2009) later found that homophily originates from long generations of biased preferences in network dynamics. The typical result of preferential attachment is that highly connected nodes earn more connections, creating extremes concerning the various degrees of nodes. In other words, most nodes have very few connections, but some nodes have very large amount. This is commonly referred to as 'The Rich get Richer' effect. Barabási and Albert (1999) investigated this concept to find that its implication caused networks to become scale-free. A scale free network is one that is distributed as a power law as opposed to a Poisson distribution which had been assumed by Watts and Strogatz. The difference between both distributions as well as a normal distribution are highlighted in Figure 6. Power law distributions decrease at a much slower rate than Poisson distributions, which allow for the extremes to be more common. In the case of networks, that means that it allows for huge and frequent degree extremes, which had been caused by preferential attachment.

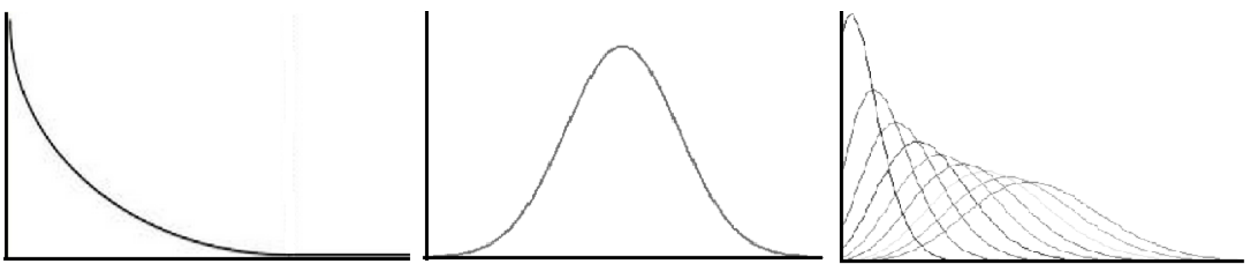

Figure 6. Examples of Power Law, Normal, and Poisson distributions. Each is displayed on a frequency graph. A normal distribution (center) is a distribution for a range of options when the average is the most likely to occur. This type of distribution will typically describe the frequencies of human heights, or test scores in a large class. 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. This distribution was originally used in 1898 in a study of how many soldiers had been accidentally killed by kicking horses. 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. The power law was first used to describe the distribution of wealth in the United States, demonstrating 'The Rich get Richer' effect. In other words, most families have similar incomes but some are incredibly rich and some are incredibly poor.

Searchability

Going back to Travers and Milgram, Kleinberg (2000) and Watts *et al* (2002) had noticed that their original done-by-hand experiment had proved something else as well as the existence of short paths, it had proved that the typical person could *find* these short paths. This does not seem too complicated at first, but think about trying to determine how far away you are from any person, say Johnny Depp. The first thing you might do is think of your acquaintances, and pick someone who either lives in Los Angeles or is trying to make it in the acting world. If you're lucky, you might find that they know a director who has met Johnny Depp making you 3 steps away from the actor, but if you are anything more than that it could be impossible to find that connection by yourself. However, the participants of Travers' and Milgram's experiment were able to find short paths. This implies that natural networks are not only *small*, but they are *searchable*. How can an individual know how to reach a specific target that they know very little about, and seems so far away? The answer was that they didn't have to. All that each person had to do was send the message to the person that they thought was closest, no individuals have to find the target by themselves. As a result, distances that seem incredibly long between nodes can be traversed in 2 short steps, seemingly breaking the triangle inequality, which states that taking 2 steps can never result in a shorter path than taking one step directly to the target. The key to understanding this is to break the network up into different levels based on groupings that cause connections, for example occupation or geographic location. Because networks work as a whole to find targets, steps used to access a target can come from any grouping, thus allowing the triangle inequality to be broken. This is further explained in Figure 7. In order to determine the groups that individuals use to find targets, McCarty *et al* (2001) did

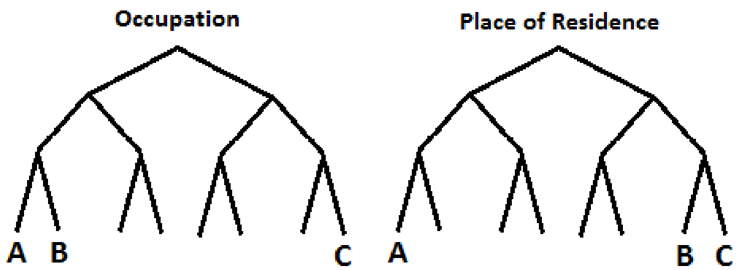

Figure 7. In both levels, C is far from A. However if A uses its occupation to connect to B, then B uses its place of residence to connect to C, A can reach C easier in 2 steps than in one, breaking the triangle inequality. Modified from Watts (2003).

an experiment that they called the 'reverse small world' experiment. What they did was set up a similar situation to Travers' and Milgram's experiment, but they did not go through with the process. Instead all that they did was interview possible participants, telling the participants to ask whatever questions they would like to know about the target in order to determine where they would send the letter to get closer. The most frequently asked questions were the name of the target, their place of residence, and their occupation. The exact same information that was supplied by Travers and Milgram when they first found that short paths exist. So it was concluded that networks are searchable because they can be broken up into different networks, based on the groupings where the connections originated. When broken up this way, individuals are able to use whatever groupings seem most appropriate, and the use by the whole network of a different level at each step results in seemingly long jumps to be taken in shorter steps.

Networks that are broken up into levels of the groups causing affiliation, are called affiliation networks. In order to study and understand these networks they can be drawn as bipartite graphs (Newman, Strogatz and Watts, 2001). Bipartite graphs have 2 opposing sides, and nodes on one side can only connect to nodes on the other side, an example is shown in Figure 8. A bipartite graph can be converted to a network by excluding one side (either the groups or the individuals in Figure 8) and drawing a network of the other side, connected nodes if they were connected to the same opposing node in the bipartite graph. For example in Figure 8, individual B would be connected to individual C as they are both connected to group A, however individual B would not be connected to individual D because they do not share any groups. The advantage of drawing two levels as a bipartite graph is that the origins of connections can be seen, as opposed to just seeing the connections between nodes. That way, observations can be made on a deeper level in terms of the dynamics of a network.

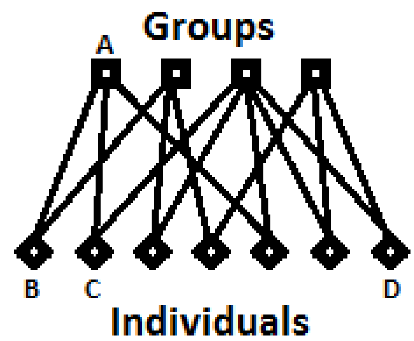

Figure 8. An example of a bipartite graph. This particular graph shows not only the connections amongst individuals, but also the origins of those connections (the groups they are a part of). Modified from Watts (2003).

Networks Changing Over Time

Now that the structure of the natural network was mostly understood, Kossinets and Watts (2006) began to apply this structure to study the way a network evolves over time. They conducted a yearlong experiment, collecting data through email messages exchanged by over 40 000 students and staff at a university campus. An email message sent resulted in a directed connection in the recorded network, and the more emails that were exchanged between 2 people, the stronger the connection would be on the recorded network. The year was split into periods of 60 days, and for each period a static network of data was collected. Then, by comparing each successive period of data, the evolution of a network over time could be observed. Minimal data about the participants was also collected; age, faculty affiliation, year of university and class list. What Kossinets and Watts found was that in a lack of global perturbations (like a new school year) the spread of a network would eventually come to an equilibrium. This equilibrium was one where students and staff with things in common were highly connected. This being said, it was also found that it would be very difficult for any particular individual to manipulate their own position in the network. The applicable conclusion of Kossinets and Watts was that there is no particular individual or set of individuals who are in the center of the network, the control of passing information is well spread.

Applications of Network Study

Life on earth is not only dominated by the interactions of networks from the interactions of neurons to computers, life on earth *is* a network. Everything we do causes the fall of dominoes, affecting other people in many ways. One major and current network example is the online social networking site Facebook. The site recently celebrated its tenth birthday, and had a lot to celebrate. As of 2013, Facebook had accumulated 1 billion 230 million users. New fads, videos, and ideas can easily be spread all around the world in a matter of days. This can be both harmless and devastating. On January 6th 2014, a Facebook page called 'NekNominations' was added in Australia, starting a dangerous worldwide drinking game that in just under 2 months had been passed from country to country and contributed to the alcohol related death of several people. But the readiness of human interactions to spread a trend is not always a bad thing, when massive flooding hit Calgary, Alberta in the summer of 2013, a page on Facebook was used to collect donations, reunite relatives (since telephone lines were down), and organize groups of volunteers. A third example of majorly spreading trends was the short lived and addicting game of Flappy Bird. For 6 months after its release, Flappy Bird was just as unknown as the hundreds of other free games available for download, but sometime in October 2013, the game went viral, and was the number 1 free game by January 2014. The reason for its incredible popularity over the other games is hard to predict, but its ability to go viral and be spread famous across the world is easily explainable by social networking. It's for reasons like this that the study of networks and their dynamics is so important. Furthermore, the understanding social interaction is directly applicable to the understanding of all other natural networks that dominate our lives, which once understood would allow us to better predict network interactions, and overall improve our way of life.

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.

Kermack WO, McKendrick AG. 1927. A contribution to the mathematical theory of epidemics. Proceedings of the Royal Society of London, Series A. 115:700-721.

Kleinberg J. 2000. The small world phenomenon: an algorithmic perspective. Conference Proceedings of the Annual ACM Symposium on Theory of Computing. 163-170.

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

Kossinets G, Watts DJ. 2006. Empirical analysis of an evolving social network. Science. 311(5757):88-90.

Kossinets G, Watts DJ. 2009. Origins of homophily in an evolving social network. American Journal of Sociology. 115(2):405-450.

Matham A. 2014. The clipped wings of flappy bird [Internet]. The New Indian Express. [updated 2014 February 23, cited 2014 February 24]. Available from: http://www.newindianexpress.com/lifestyle/tech/The-Clipped-Wings-of-the-Flappy-Bird/2014/02/23/article2069828.ece#.UwwLMPldXY4

McCarty C, Killworth PD, Bernard HR, Johnsen EC, Shelley GA. 2001. Comparing two methods for estimating network size. Human Organization. 60(1):28-39.

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

Newman MEJ, Strogatz SH, Watts DJ. 2001. Random graphs with arbitrary degree distributions and their applications. The American Physical Society. 64(26118):1-17.

Newman MEJ, Watts DJ. 1999 Scaling and percolation in the small world model. The American Physical Society. 60(6):7332-7342.

The Guardian [Internet]. 2014. Facebook: ten years of social networking, in numbers. [cited 2014 February 24]. Available from: http://www.theguardian.com/news/datablog/2014/feb/04/facebook-in-numbers-statistics

TheStar.com GTA [Internet]. 2014. NekNomination: the rise and fall of an online drinking game. [cited 2014 February 24]. Available from: http://www.thestar.com/news/gta/2014/02/09/neknomination_the_rise_and_fall_of_an_online_drinking_game.html

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, Dodds PS, Newman MEJ. Identity and search in social networks. Science. 2002;296(5571):1302-1305.

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