Survey of Identified Small World Networks

Sorelle Friedler
Dept. of Computer Science
Swarthmore College

Spring 2004

Abstract

This survey gives an overview of small-world network qualifications as well as a classification of some known examples. These examples are classified into socially created and naturally occuring, where socially created examples are based on underlying social relationships among people, socially influenced have human social relationships and socialization as factors in determining the network but are not defined by them, and naturally occuring networks are independent of social relationships. It is also suggested that these classifications make sense not only when considering the subject of the network, but also in relation to the clustering coefficient ratio of the actual sample as compared to what a random sample would produce for the same size graph.

This survey is intended as a starting point especially for researchers interested in adding to the literature on known examples of small world networks, since there is no currently comprehensive listing of known small world networks and their characteristics.

1 Introduction

1.1 What is a Small-World Network?

A small world network is a graph or network of points that has small highly clustered groups with connections between the groups. Intuitively, this is demonstrated by professional acquaintances - everyone in an office will know each other (this is highly clustered), then when someone changes jobs two small clusters will become connected. This means that paths between any two randomly chosen nodes in the graph should be short. This phenomenon is demonstrated in many real world examples including social networks, neural networks, and the network of people who have acted in movies with Kevin Bacon.

1.2 Motivation and Future Work

Small World networks are becoming increasingly important in many areas of Computer Science, Biology, and Sociology, just to name a few. However, papers about small-world networks still reference relatively few examples and a comprehensive list of known small-world networks still needs to be compiled. Future work could include providing such a list. This survey provides a beginning for that work. Also, it would be useful to have a categorization based not only on subject area, but on inherent characteristics of specific small world networks possibly based on clustering coefficient ratios. It seems, from looking at the table in section 4, that natural breaks in characteristics of small world networks may occur at clustering coefficient ratios of

100 and 1000, further work could explore these characteristic groups. This paper suggests that these breaks may be connected to qualitative differences in type of network.

2 Background

Everyone knows the idea of "six degrees of separation," the idea that any two people in the world are connected by a chain of six aquaintances, but where did this idea start and what exactly does it mean? In 1960 Stanley Milgram conducted an experiment where he gave envelopes containing instructions and names of target people in Boston to random people in Nebraska and Kansas. The source people in Nebraska and Kansas were instructed to pass the envelope on to someone they knew who they believed would be more likely to know the destination person. The path lengths (number of people the envelope was passed to before it reached its destination) averaged 6. This began the idea of "six degrees of separation." [23] Despite popular misconception however, the path length of 6 is the average number, not an upper limit, and the original experiment was limited to the US. The Milgram experiment did not claim the popular version of the result - any two people in the world can be connected by a chain of six or less acquaintances.

More recently, Watts and his colleagues at Columbia University conducted a similar experiement, but on an international scale and using email instead of letters. They confirmed Milgram's results with emails reaching their targets in an average of 5 to 7 steps. Of course there are many problems to be considered in experiments such as these - data sampling is not random and is not actually global since only people with email can participate. Given, however, that approximately 10 million people world-wide have e-mail access, this is a very good sample size. [16]

These experiments demonstrate that the social networks which underly much of network science have special characteristics. What are these special characteristics and how can we determine if a network falls into this new category?

3 Definitions

3.1 Clustering Coefficient

Given a vertex v, let the number of vertices connected to v by a path length of 1 (v's neighbors) be k. Consider the graph consisting only of vertices that are neighbors of v - this graph can have at most $k(k-1)/2$ edges. C_v is the number of these edges that exist divided by $k(k-1)/2$. The clustering coefficient, C, for the original graph is the average of C_v over all v.

3.2 Characteristic Path Length

The characteristic path length, L, of a graph is the number of edges in the shortest path between two vertices averaged over all pairs of vertices.

3.3 Regular Lattice

A regular lattice is a graph which is connected to each of its neighbors and no other vertices. For example, a grid on 16 vertices that looks like a tic-tac-toe board is a regular lattice. Lattices can also be configured in a ring in which case the number of neighbors can be chosen and then each vertex is connected to that many of its nearest neighbors.

3.4 Random Graph

To make a random graph, take a regular ring lattice and with a probability of p=1 for each vertex and edge pair in the regular ring lattice, change that edge to connect to a randomly chosen vertex over the entire ring (no duplicate edges allowed). Call the new graph a random graph with clustering coefficient C_{random} and characteristic path length L_{random}.

3.5 Small World Network

To make a small world network by randomizing a regular ring lattice, pick a value of p between 0 and 1 and do the same procedure as described above. The resulting graph will have a high clustering coefficient, like a regular ring lattice, and have small characteristic path length, like a random graph.

Quantified, this means that small world graphs have $C \gg C_{random}$ and $L \simeq L_{random}$. [23]

4 Social Networks

Since Watts and Strogatz first codified these ideas in 1998, there has been an explosion of research into small world networks in areas ranging from Biology to Linguistics to Hollywood. A list of many known examples is given in section 11. As you might expect, these examples start global networks of personal relationships or some important subset of such a network. Diseases are one example which operate on this type of network.

Diseases spread based on social and trasmission networks which are known to be small world networks. This means that given any transmission network on which a disease is operating, it is important to understand the characteristic path length and the clustering coefficient for that network to understand how the disease will spread. A network with a high degree of clustering and a short path length is likely to be prone to epidemic, while a network with smaller clustering and longer path length may escape an outbreak relatively unharmed. The underlying small world topology for these networks also affects how diseases evolve since the clustering coefficient influences competition patterns among strains - a more clustered network will create more competition among child strains of a disease and so more evolved and potentially deadly diseases. [19]

Sexually Transmitted Diseases are an important subset of diseases which rely heavily on their underlying small world relationship network. Gathering data about precisely how this graph is connected is unfortunately difficult due to its sensitivity as well as the usual reasons of scale and reliability, so precise information about global characteristic path lengths and clustering coefficients for various STDs are not currently available. [10] [12] [13] [20] I imagine that when these statistics become available and it becomes clear just how many steps away any given person is likely to be to someone with an STD, due to the small world nature of the network of sexual relationships, the information will be staggering. This information would be important for sexual health workers as well as epidemiologists and could potentially be used to decide how to best manage the current HIV/AIDS crisis in many parts of the world based on the particular characteristics of the local relationship network.

The SARS outbreak, which was first noticed in China in February of 2003, is another example of a disease operating on a small world social and interaction network. SARS, however, demonstrates that the infection rate of a disease can not necessarily be predicted soley based on the underlying topology of the potential infection network. Instead, the saving grace with the SARS outbreak has been that while SARS sits on a small world network, its infection rates are linear as would normally be expected in a regular lattice network. The linear infection rates have allowed health officials to use isolation as an effective way of curbing the spread of the disease. [18]

5 Professional Networks

Professional networks and social networks are often indistinguishable when simply considering these networks as networks of interpersonal relationships. Instead, we consider professional networks to be networks of professional collaboration.

If we consider a graph of Hollywood actors with connections between actors who have been in movies together, we find that Kevin Bacon is one of the people at the center of a large connected component. From this observation (and pop culture speculation), Kevin Bacon numbers are created - any actor who has been in a movie with Bacon has a number of 1, any actor who has been in a movie with someone with a Bacon number of 1 has a number of 2, and so on. The Bacon numbers indicate the distance of the path on the graph between the given actor and Kevin Bacon. The practical use of Kevin Bacon numbers? The same as the profession - entertainment. [3] [23]

Within the mathematical professional community, there is long standing folklore about the mathematician Paul Erdős and connections to him through common research papers. Erdős is given Erdős number 0, anyone who's written a paper with him is given number 1, anyone who's written a paper with someone with Erdős number 1 is given an Erdős number of 2, and so on in the same manner as with Kevin Bacon numbers (and again, the use is pure mathematical entertainment).

In his paper, Jerrold Grossman explains how he analyzed the data of the Mathematical Reviews, which has catalogued most published mathematics papers since 1940, and created a collaboration graph with authors as vertices and edges between authors who had written a paper together. In this graph with around 337,000 authors there is on large connected component with about 208,000 authors one of which is Paul Erdős. It is this connected component that is considered (authors not in this section have an Erdős number of infinity). Grossman finds the characteristic path of this connected component to be between 7 and 8 (around 7.33) and the clustering coefficient to be 0.15 (10,000 times higher than would be expected for a random graph of the same size, making the mathematical community the most clustered of the examples given in this paper). [2] [11]

6 Political Networks

When we think of political networks we must first think of the social networks and underlying interpersonal relationships which form a foundation for modern politics. Then professional networks - politicians serving from the same state or cosponsoring a bill - should come to mind. Informally analyzing these connections and making use of the, most likely, short paths between any two politicians is simply good politics. It is, however, the networks which could affect policy decisions, and not simply reelection, to which I am referring in this section. Given the current political climate, the obvious immediate example of a political small world network is terrorism.

Terrorist networks are currently small scattered groups who are in contact with each other but have no overarching leadership. This description intuitively fits the description of a small world network - there are highly clustered sections with connections between sections to provide for short path length (possibly representing communication time). There is a group called the Counter Terrorist Lab which appears to be studying the application of small world networks to terrorist networks, however all their results which are specific to terrorism are classified, so it is hard to know much about exactly what is known. [9] [1]

7 Linguistic Networks

Linguistic, technological, and transportation networks are neither completely based on interpersonal interactions, nor completely removed from them. As might then be expected, it can be seen in the chart in section 12 that these networks are less clustered than the social, political, and professional networks, and more clustered than the biological networks we will discuss later.

Human Language can be viewed as a graph with words as nodes and connections between nodes if the two words cooccur significantly. Cancho and Sole use what they call basic and improved techniques to create the graph and determine which nodes should be connected. Both methods result in a small world graph. The specific results are shown in the table in part 12, the basic technique results are labelled unrestricted and the improved results are labelled restricted. [8]

Language can also be viewed as a graph with connections between words if they have similar meanings. When done with English, this graph was found to be less clustered and with a longer characteristic path length than the cooccurance graph. This result is as might be expected since, intuitively, we can see that given any two random words - "this" and "result" from the beginning of this sentance for example - they are more likely to cooccur in a sentance than to have the same meaning. [7] [6]

8 Technological Networks

The World Wide Web is a small world network that can be viewed as a graph with pages as nodes and edges representing links between pages. These links can represent relationships, interests, or can be general links to common pages like search engines. If the edges are undirected, representing a link in either direction even if only one page links to the other (your home page links to google, and google does not link back), the clustering coefficient is about six times higher than if the edges in the graph are directed, however either graph effectively shows through its high clustering coefficient and low characteristic path length that the world wide web is a small world network. This result could affect search engines and networking algorithms. [4]

If we consider the text and links to and from personal web pages and available mailing lists, it is possible to discern friendship groups, which is a way to use the internet to examine social networks on a larger scale, but with some error. Adamic and Adar considered this information for MIT and Stanford groups and found both groups to be about 70 times more clustered than the same sized random network, though MIT's average path length was shorter than Stanford's, probably due to MIT's smaller size. [5]

The power grid is also based somewhat on human connections - power lines will be more frequent and clustered within a city than between cities, in the same way that you are more likely to know your neighbor than someone from out of state. The power grid is, however, more heavily based on power supply and demand along with other factors, so it must be considered as a separate small world network. As a network based mostly on factors independent of human interaction, it is much less clustered than the social and professional networks mentioned earlier. The power grid of the western United States, chosen since it was a known and completely mapped network, was examined by Watts and Strogatz and found to be a small world network. [3] [23]

9 Transportation Networks

The Boston subway system can be viewed as a graph with stations as nodes and edges between nodes which are connected by direct tunnels. Latora and Marchiori reject the classic analysis of small world networks for this system, arguing that examining clustering coefficients in such a sparsely connected graph which has a high likelyhood of having a corresponding disconnected random graph is illogical and also doesn't take into account distance between stations (cost). Latora and Marchiori develop an idea of global and local efficiency, in the case of subway systems efficiency would represent how close to the ideal (where each station is directly connected to every other station) the system is. Latora and Marchiori argue that small world networks are those which have high both global and local efficiency. The Boston Subway system has high global efficiency, but low local efficiency, so it is not a small world network, but when the subway system is combined with the bus system it has both high global efficiency and high local efficiency, so the Boston transportation system as a whole is a small world network. [15]

10 Biological Networks

Not only are small world networks found in situations relating to human relationships, they are also found in naturally occuring situations our bodies and in ecological systems. These biological networks are the least clustered, the closest to the random graphs of the same size, of the examples we are considering.

The neural network of the nematode worm C. Elegans and the metabolic network of E. Coli are both known to be small world networks. [3] [14] [22] [21] [23] Ecological food webs can also be made into graphs with species as the vertices and undirected lines between species who have any interactions as predators or prey. Montoya and Sole analyze the graphs for three known food webs, the Ythan estuary (two graphs represent this food web, the second has 42 added metazoan parasite species), Silwood park, and Little Rock lake. The results demonstrate that these food webs are small world networks. The specifics of all of these examples are shown in the table in section 12. [17]

11 Characterization of Examples

1. Socially Created
   (a) Social Networks
      i. Six Degrees of Separation - Milgram [23]
      ii. Six Degrees of Separation - Email [16]
      iii. Diseases [19]
         A. Sexually Transmitted Diseases [10] [12] [13] [20]
         B. SARS [18]
   (b) Professional Networks
      i. Kevin Bacon and the Actors Network [3] [23]
      ii. Erdos [2] [11]
   (c) Political Networks
      i. Terrorist Networks [9] [1]
2. Socially Influenced
   (a) Linguistic Networks
      i. Human Language Usage [8]

ii. English Word Meaning [7] [6]

(b) Technological Networks

i. WWW [4]

ii. Friendship Groups on the Internet [5]

iii. The Power Grid [3] [23]

(c) Transportation Networks

i. The Boston Transportation System (subway and bus) [15]

3. Naturally Occuring

(a) Biological Networks

i. Neural Networks [3] [14] [23]

ii. Metabolic Networks [22] [21]

iii. Food Webs [17]

12 Table of Examples

This table contains the actual and random characteristic path lengths and clustering coefficients as well as the ratios between actual and random data for all of the examples given in this survey. The examples are ordered based on their ratio of actual clustering coefficient to the clustering coefficient for a random network of the same size, least clustered to most clustered. Since the Boston subway system, STDs, SARS, and the six degrees of separation examples did not provide information about clustering coefficients their data is not included.

	L_{actual}	L_{random}	$\frac{L_{actual}}{L_{random}}$	C_{actual}	C_{random}	$\frac{C_{actual}}{C_{random}}$
Food Web: Ythan Estuary 1	2.28	2.09	1.09	0.21	0.09	2.33
Food Web: Little Rock	2.22	1.60	1.39	0.35	0.14	2.5
Food Web: Ythan Estuary 2	2.43	2.26	1.08	0.22	0.06	3.67
Food Web: Silwood	3.40	3.23	1.05	0.15	0.03	5
C. Elegans Neural Network	2.65	2.25	1.18	0.28	0.05	5.6
Metabolic Network: Reaction Graph	2.62	1.98	1.32	0.59	0.09	6.6
Metabolic Network: Substrate Graph	2.9	3.04	0.95	0.32	0.026	12.3
Power Grid	18.7	12.4	1.51	0.080	0.005	16
Friends on WWW: MIT	6.4			0.21		70
Friends on WWW: Stanford	9.2			0.23		70
WWW: directed	4.23			0.081	1.05×10^{-3}	77.1
English	3.16	2.5	1.26	0.53	0.002	265
WWW: undirected	3.1			0.108	2.3×10^{-4}	469.6
Human Language: Restricted	2.67	3.06	0.87	0.437	1.55×10^{-4}	2819.4
Kevin Bacon	3.65	2.99	1.22	0.79	0.00027	2925.9
Human Language: Unrestricted	2.63	3.03	0.87	0.687	1.55×10^{-4}	4432.3
Erdős	7.33			0.15		10,000

References

[1] Counter-terrorism lab (ct-lab). Website: http://www.dsi.unive.it/ massimo/ct-lab/.

[2] Facts about erdos numbers and the collaboration graph. Website: http://personalwebs.oakland.edu/ grossman/trivia.html.

[3] Kevin bacon, the small-world, and why it all matters. Website: http://www.santafe.edu/sfi/publications/Bulletins/bulletinFall99/workInProgress/smallWorld.html.

[4] Lada A. Adamic. The small world web. *Proceedings of the Third European Conference on Research and Advanced Technology for Digital Libraries*, pages 443 – 452, 1999.

[5] Lada A. Adamic and Eytan Adar. Friends and neighbors on the web. *Social Networks*, 25(3), 2001.

[6] Ying-Cheng Lai Adilson E. Motter, Alessandro P. S. de Moura and Partha Dasgupta. Topology of the conceptual network of language. *Physical Review E*, 65, 2002.

[7] Philip Ball. Small world network: English words are connected by just three degrees of separation. Website: http://www.nature.com/nsu/020701/020701-2.html, Jul 2002.

[8] Ramon Ferrer Cancho and Richard V. Sole. The small-world of human language. *Proc. Roy. Soc. of London Series B*, 268:2261 – 2266, 2001.

[9] Steven M. Cherry. It's a small world after all: A tour of the theory of networks reveals some surprising links. *IEEE Spectrum*, Jun 2002. Book review of Linked by Albert-Laszlo Barabasi.

[10] Ken T. D. Eames and Matt J. Keeling. Modeling dynamic and network heterogeneities in the spread of sexually transmitted diseases. *Proc. Natl. Acad. Sci. USA*, 99(20), 2002.

[11] Jerrold W. Grossman. Patterns of collaboration in mathematical research. *SIAM News*, 35(9), 2002.

[12] James Holland Jones and Mark S. Handcock. An assessment of preferential attachemtn as a mechanism for human sexual network formation. 2002.

[13] James Holland Jones and Mark S. Handcock. Likelihood-based inference for stochastic models of sexual network formation. 2003.

[14] Helmut Kroger. Small world networks - what the fuss is all about and are neurons organized this way? Website: http://www.cap.ca/edu/lec-talks/H-Kroger2.html.

[15] Vito Latora and Massimo Marchiori. Is the boston subway a small-world network? *Physica A*, 2002.

[16] Stefan Lougren. Six degrees of e-mail separate wired world? Website: http://news.nationalgeographic.com/news/2003/08/0807_030807_email.html, Aug 2003.

[17] Jose M. Montoya and Ricard V. Sole. Small world patterns in food webs. *Journal of Theoretical Biology*, 214:405 – 412, 2002.

[18] Norio Konno Naoki Masuda and Kazuyuki Aihara. Transmission of severe acute respiratory syndrome in dynamical small-world networks. *Physical Review E*, 69, 2004.

[19] Jonathan M. Read and Matt J. Keeling. Disease evolution on networks: the role of contact structure. *Proc. Roy. Soc. of London Series B*, 270:699 – 708, 2003.

[20] Ian Sample. Six degrees of copulation. Website: http://www.newscientist.com/news/news.jsp?id=ns9999913, Jun 2001.

[21] Andreas Wagner and David Fell. The small world of metabolism. *Nature Biotechnology*, 18:1121 – 1122, 2000.

[22] Andreas Wagner and David Fell. The small world inside large metabolic networks. *Proc. Roy. Soc. London Series B*, 268:1803 – 1810, 2001.

[23] Duncan J. Watts and Steven H. Strogatz. Collective dynamics of 'small-world' networks. *Nature 393*, 1998.