Lecture 01 — 01/15/13

Lecturer: Ravi Srinivasan *Scribe: N/A*

1.1 Introduction

Welcome to M375T/M396C. In this course we will consider mathematical concepts that underlie the study of complex networks and, in particular, social networks. A list of topics to be covered, the class syllabus, and any course material (lecture notes, problem sets, links to articles, etc.) is available on the course website:

`http://www.ma.utexas.edu/users/rav/ComplexNetworks/`

1.2 Social networks and small-world property

In order to study mathematical models that have some basis in reality, we need to have a sense of the properties that real-world networks share. One particularly important one is the *small-world property*.

Most of us have had the experience that one of our friends knows—or is friends with someone who knows—a person we believed to be socially distant, like a celebrity or a famous politician. This property has garnered fame under the notion of "six degrees of separation," which hypothesizes that we are each connected to any other person on the planet by only six links, on average. The small-world effect is necessary for many important features of a social network to take hold: fast and widespread propagation of information, tipping point phenomena where small changes have large effects, and network robustness and resiliency.

1.2.1 Milgram's experiment

The first evidence of a small-world property was discovered in a series of groundbreaking experiments by the sociologist Stanley Milgram in the 1960's. At the time there was no notion of what a social graph looked like—remember, Facebook didn't exist yet! To infer the underlying network structure, Milgram picked a "target" in Boston and sent folders with a description of the target to an arbitrary sample of people across the country. These participants were asked to add their name to the folder and send it to a friend or acquaintance who would be most likely to know the target or someone who does (i.e., *greedy routing*).

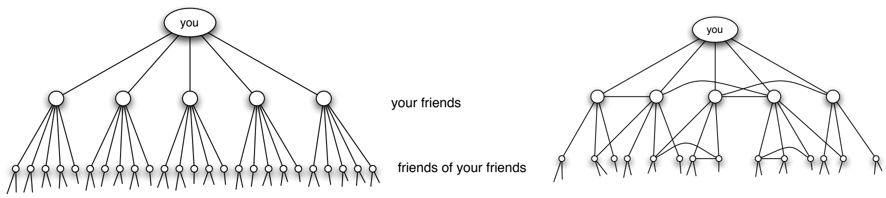

Figure 1.1. Which one looks more realistic?

Two experiments showed that of the $\sim 25\%$ of folders that were actually received by the target, their paths had an average length of 6.2 links—hence, "six degrees of separation." It was noted that these paths were not random, with local links correlated to gender and friends/family and global links typically related to hometown or work. Furthermore, several intermediaries received multiple folders, ensuring that they somehow play a central role in this routing network.

Milgram's observations have been confirmed in modern settings, most notably in an online experiment by Dodds and Watts using e-mail forwarding. The small-world property is also evident in networks where the social graph is known, like when computing one's Erdős number (distance to Paul Erdős in a co-authorship network) or Bacon number (distance to Kevin Bacon in an acting collaboration network). In an analysis of the largest social graph to date, a 2011 study of Facebook's 721 million users found an average number of 4.74 links between users. These experiments provide good evidence that it is, in fact, "a small world, after all."

1.2.2 A first guess at a mathematical model

To build a nave model of a social graph, consider the following reasoning. Suppose you have 500 direct acquaintances, each of which has 500 direct acquaintances, and so forth. If each person's set of acquaintances is disjoint, then the resulting graph is a tree. The result is that 500 people (0.00017% of US population) are one link away from you, 250K (0.083% of US population) are two links away, and 125 million (42% of US population) are three links away. Notice that we have a small-world effect in this model.

More generally, suppose each person has λ acquaintances. The total number of people d links away equals $n = \lambda^d$. Then $d = \ln n / \ln \lambda$, so that the diameter of this simple network scales logarithmically with the number of people. This is a tell-tale sign of a small-world effect which we will observe in a variety of settings throughout the course. Is this model sufficient to understand a real-world network?

The answer, of course, is no. The unrealistic assumption of disjoint sets of direct contacts rules out the phenomenon of *clustering*. In any social context, it is reasonable to expect that that your friends are likely to be friends with each other. This leads to cross-connections

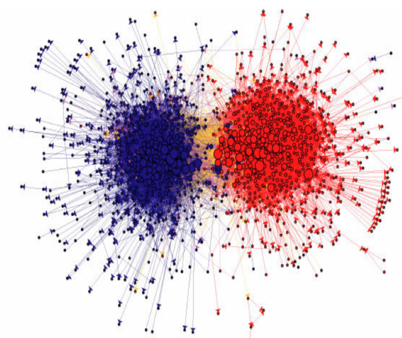

Figure 1.2. A typical real-world network. Note the existence of two dense clusters that are well-separated, and that some nodes may have many more connections than others.

across levels of the simple tree model, i.e., triads (see Figure 1.1). A quick calculation demonstrates that if each acquaintance only has a fraction $m \in (1/\lambda, 1]$ of "new" acquaintances then the maximum distance from you is $d = (\ln m + \ln n)/(\ln m + \ln \lambda) \approx \ln n/(\ln m + \ln \lambda)$ for large n. Thus, it is possible to drop the disjointness assumption and still have the diameter still grow logarithmically, albeit at a faster rate.

Another reason that the simple model is incorrect is that social acquaintances are biased, and often lead to people who are significantly more well-connected than others. The appearance of *power-law distributions* in the number of contacts per person, etc., is something we will look at in greater detail later.

To summarize, there are several criterion that we would like our models to satisfy in order to be in accordance with reality:

- *Small-world effect*: small diameter of social graph
- *Clustering*: prevalence of triads
- *Algorithmic small-world*: efficiency of greedy routing
- *Power-laws*: some people are significantly better-connected than others

1.3 Landau symbols

We will make frequent use of *Landau notation* (i.e., *big-O notation*) throughout this course. Landau symbols are used to describe the asymptotic behavior of a function relative to another function, as we define precisely below.

Let $f : \mathbb{N} \to \mathbb{R}$ and $g : \mathbb{N} \to \mathbb{R}$. We say that

$$f = O(g) \quad \Longleftrightarrow \quad \limsup_{n \to \infty} \left| \frac{f(n)}{g(n)} \right| < +\infty.$$

Equivalently, $f = O(g)$ if and only if there exist constants C and N such that $|f(n)| \leq C|g(x)|$ for all $n > N$. This implies that f grows no faster (or decays no slower) than g as $n \to \infty$.

Next, we say that

$$f = o(g) \quad \Longleftrightarrow \quad \lim_{n\to\infty} \left|\frac{f(n)}{g(n)}\right| = 0.$$

In this case, f grows strictly slower (or decays strictly faster) than g as $n \to \infty$.

We can use big-O and little-O notation to define all the other Landau symbols. In particular,

$$\begin{aligned} f = \Omega(g) &\iff g = O(f) \\ f = \omega(g) &\iff g = o(f) \\ f = \Theta(g) &\iff f = O(g) \text{ and } g = O(f). \end{aligned}$$

A handy table of analogies to inequalities is given below:

Notation	Analogy
$f = O(g)$	$f \leq g$
$f = o(g)$	$f < g$
$f = \Omega(g)$	$f \geq g$
$f = \omega(g)$	$f > g$
$f = \Theta(g)$	$f = g$

References

1. Acemoglu, D., & Ozdaglar A. (2009). 6.207/14.15, Networks, Fall 2009 [Lecture notes]. Retrieved from:
`http://economics.mit.edu/faculty/acemoglu/courses/`

2. Backstrom, L., Boldi, P., Rosa, M., Ugander, J., & Vigna, S. (2011). Four degrees of separation. arXiv preprint arXiv:1111.4570.

3. Chaintreau, A. (2012). COMS 4995-1, Introduction to Social Networks, Fall 2012 [Lecture notes]. Retrieved from:
`http://www.cs.columbia.edu/~augustin/`

4. Dodds, P. S., Muhamad, R., & Watts, D. J. (2003). An experimental study of search in global social networks. Science, 301(5634), 827-829.

5. Easley, D., & Kleinberg, J. (2010). Networks, crowds, and markets. Cambridge University Press.

6. Jackson, M. O. (2010). Social and economic networks. Princeton University Press.
7. Milgram, S. (1967). The small world problem. Psychology today, 2(1), 60-67.
8. Travers, J., & Milgram, S. (1969). An experimental study of the small world problem. Sociometry, 425-443.
9. Ugander, J., Karrer, B., Backstrom, L., & Marlow, C. (2011). The anatomy of the Facebook social graph. arXiv preprint arXiv:1111.4503.

Last edited: March 21, 2013