Why Model Networks?

Rogayeh Dastranj Tabrizi
email: rda18@sfu.ca
Office: WMC 3607
Office Hours: Thursdays 12pm-2pm

Department of Economics
Simon Fraser University

Thanks to Matthew Jackson for access to his teaching resources.

10. Februar 2015

Introduction

Many ways in which network structures affect our well-being makes it critical to understand:

- How social network structures impact behaviour.
- Which network structures are likely to emerge in a society.

For example:

- Networks and social interactions in crime:
  - Reiss(1980,1988) - 2/3 of criminals commit crimes with others
  - Glaeser, Sacerdote and Scheinkman (1996) - social interaction important in petty crime, among youths
- Myers and Shultz (1951)- textile workers:
  - 62 % First job from content
  - 23 % by direct application
  - 15 % by agency

The Challenge

How many networks on just 20 nodes?

- Person 1 could have 19 possible links, person 2 could have 18 not counting 1, ... total = 190
- So 190 possible links, each could either be present or not, so $2 \times 2 \times 2 \times ...190$ times $= 2^{190}$ networks
- Atoms in the universe: somewhere between 2^{158} and 2^{246}!!

Simplifying The Complexity

- Global patterns of networks:
  - Degree distributions
  - Path lengths
- Segregation Patterns: node types and homophily
- Local Patterns
  - Clustering
  - Support
- Positions in networks
  - Neighbourhoods
  - Centrality, influence, ...

Florentine Marriages

- The Medici have been called the "godfathers of the Renaissance".
- Their accumulation of power in the early 15th century in Florence, was orchestrated by Cosimo de Medici despite the fact that his family started with less wealth and political power than other families .
- Cosimo consolidated political and economic power by leveraging the central position of the Medici in networks of family inter-marriages and economic/political relationships.

Florentine Marriages

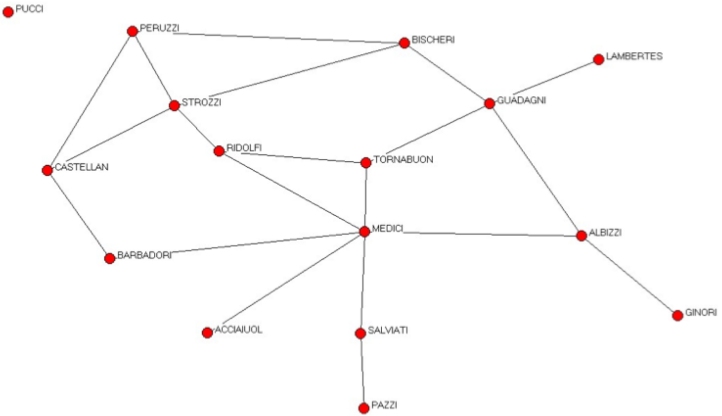


Figure 1.1: 15th Century Florentine Marriges Data from Padgett and Ansell [493] (drawn using UCINET)

Florentine Marriages

- If we do a rough calculation of importance in the network, simply by counting how many families a given family is linked to through marriages, then the Medici do come out on top.
- But, they only edge out the next highest families, the Strozzi and the Guadagni, by a ratio of 3 to 2.
- So maybe we should think about another measure of importance!

Florentine Marriages - Betweenness Centerality

- Let $P(ij)$ denote the number of shortest paths connecting family i to family j.
- Let $P_k(ij)$ denote the number of these paths that family k lies on.
  - For instance, the shortest path between the Barbadori and Guadagni has three links in it.
  - There are two such paths: Barbadori - Medici - Albizzi - Guadagni, and Barbadori - Medici - Tournabouni - Guadagni.
  - If we set $i =$ Barbadori and $j =$ Guadagni, then $P(ij) = 2$.
  - As the Medici lie on both paths, $P_k(ij) = 2$ when we set $k =$Medici, and $i =$ Barbadori and $j =$ Guadagni.
  - In contrast this number is 0 if we set $k =$ Strozzi, and is 1 if we set $k =$ Albizzi.
  - Thus, in a sense, the Medici are the key family in connecting the Barbadori to the Guadagni.

Florentine Marriages - Betweenness Centerality

- In order to get a fuller feel for how central a family is, we can look at an average of this betweenness calculation.
- We can ask for each pair of other families, what **fraction of the total number of shortest paths** between the two, the given family lies on.
    - This would be 1 if we are looking at the fraction of the shortest paths the Medici lie on between the Barbadori and Guadagni.
    - and 1/2 if we examine the corresponding fraction that the Albizzi lie on.

Averaging across all pairs of other families gives us a betweenness or power measure (due to Freeman) for a given family. In particular, we can calculate:

$$\sum_{ij:i\neq j,k\neq\{i,j\}} \frac{P_k(ij)/P(ij)}{(n-1)(n-2)/2}$$

Florentine Marriages - Betweenness Centerality

$$\sum_{ij:i\neq j,k\neq\{i,j\}} \frac{P_k(ij)/P(ij)}{(n-1)(n-2)/2}$$

- where we set $\frac{P_k(ij)}{P_(ij)} = 0$ if there are no paths connecting i andj.
- Denominator captures that a given family could lie on paths between up to $\frac{(n-1)(n-2)}{2}$ pairs of other families.
- This measure of betweenness for the Medici is .522. That means that if we look at all the shortest paths between various families (other than the Medici) in this network, the Medici lie on over half of them!
- In contrast, a similar calculation for the Strozzi comes out at .103, or just over ten percent.

Friendships Among High School Students

This example comes from the The National Longitudinal Adolescent Health Data Set, known as "**Add Health**". These data provide detailed social network information for over ninety thousand high school students from U.S. high schools interviewed during the mid 1990s; together with various data on the students' socio-economic background, behaviours and opinions.

Next Figure shows the network of romantic relationships. The students were asked to list the romantic liaisons that they had during the six months previous to the survey.

Friendships Among High School Students

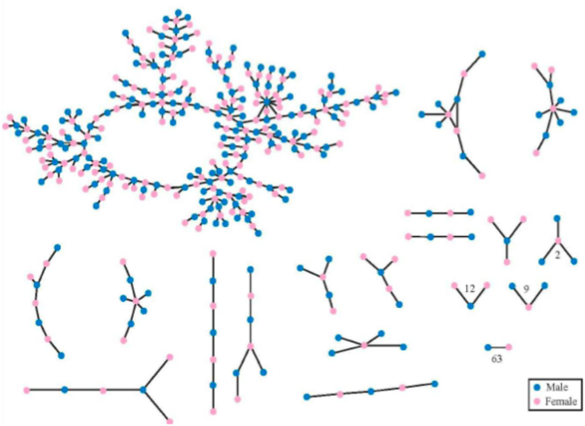


Figure 1.2: A Figure from Bearman, Moody and Stovel [47] based the Add Health Data Set. A Link Denotes a Romantic Relationship, and the Numbers by Some Components Indicate How Many Such Components Appear.

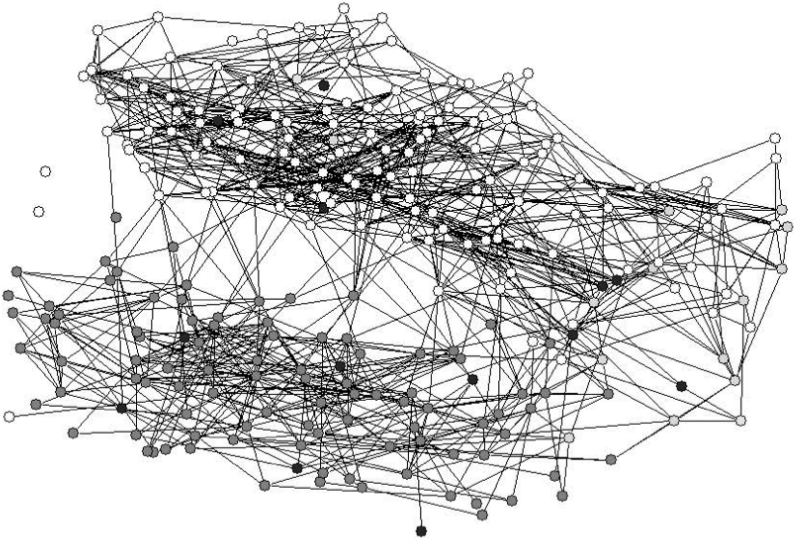
Friendships Among High School Students

- The network is nearly a bipartite network, meaning that the nodes can be divide into two groups, male and female, so that links only lie between groups.
- Despite its nearly bipartite nature, the distribution of the degrees of the nodes (number of links each node has) turns out to closely match a network where links are formed **uniformly at random**.
  - e.g. we see a giant component , where over one hundred of the students are connected via sequences of links in the network.
  - The next largest component (maximal set of students who are each linked to one another via sequences of links) only has ten students in it.
  - This component structure has important implications for the diffusion of disease, information, and behaviours.

Friendships Among High School Students

- Network is quite "tree-like" in that there are very few loops or cycles in the network.
- There is a very large cycle visible in the giant component, and then a couple of smaller cycles present, but very few overall.
- The absence of many cycles means that as one walks along the links of the network until hitting a dead-end, most of the nodes that are met are new ones that have not been encountered before.
  - This is important in navigation of networks.
  - This feature is found in many random networks in cases where there are **enough links** so that a giant component is present, but there are also **few enough links** so that the network is not fully connected.

Friendships Among High School Students



Friendships Among High School Students

- This network is also from the Add Health data set and connects a population of high school students.
- Here the nodes are coded by their race rather than sex, and the relationships are friendships rather than romantic relationships.
- This is a much denser network than the romance network.
- A strong feature in this network is Homophily: there is a bias towards similar individuals.

Friendships Among High School Students

- The bias due to Homophily is above what one would expect due to the makeup of the population.
  - 52 percent of the students are white, and yet 86 percent of whites-friendships are with other whites.
  - 38 percent of the students are black and yet 85 percent of blacks-friendships are with other blacks.
  - Hispanics are more integrated in this school, comprising 5 percent of the population, but having only 2 percent of their friendships with Hispanics.

Friendships Among High School Students

- If friendships were formed without race being a factor, then whites would have roughly 52 % of their friendships with other whites rather than 85 %. This bias is referred to as "inbreeding homophily" and has strong consequences.
- It means that the students end up somewhat segregated by race.
- This will impact the spread of information, learning, and the speed with which things propagate through the network;