Decision Making and Social Networks

Lecture 3: Understanding the structure of a network

Umberto Grandi

Summer 2013
An interesting experiment

In 1967, Stanley Milgram from Harvard wanted to measure the “distance” between two random persons in the United States:

How many acquaintances does it take to connect them?

Two cities selected: Wichita, Kansas and Omaha, Nebraska. Intuitively? Around 100 steps? No...25 % of the letters made it to the recipient, and the median number of steps required was 5.5: the famous six degrees of separation.
A second interesting experiment

“The strength of weak ties”, influential paper by Mark Granovetter in 1973

Reporting on experiments on labour market. How did people get their jobs?

- Mostly through their network of friends (Myers and Shultz, 1951, Rees and Shultz, 1970...)
- Not exactly through friends but through acquaintances

Weak ties are very important in small world networks!
A nice graph
Overview

1. Understanding the structure of a network:
   A lot of definitions to identify important features of networks.

2. How networks form and evolve:
   Networks are not given, how do they grow?

3. How networks influence decision processes:
   Your presentations.
Introduction:

Understanding the structure of a network
Power structure: the Medici family

Power structure: the Medici family

- How **popular** are they? They are connected with 6 families. Strozzi: 4, Guadagni: 4. Not enough as explanation of their rise to power...

- How **connected** are they?

- How **tight** are they?

- How **important** are they?

Let's inspect the power structure...
Power structure: the Medici family

- \( P(i,j) \) number of shortest paths connecting family \( i \) to family \( j \)
- \( P_k(i,j) \) number of shortest paths between \( i \) and \( j \) including \( k \)
- Barbadori-Guadagni: 2, Barbadori-Guadagni including Medici: 2, Barbadori-Guadagni including Strozzi: 0

One possible definition of power

The average fraction of shortest path between two families including the Medici (M) can be expressed as follows:

\[
\sum_{\{i,j| i \neq j, M \not\in \{i,j\}\}} \frac{P_M(i,j)/P(i,j)}{(n-1)(n-2)/2}
\]

Power of the Medici: 0.522. Strozzi: 0.103, Guadagni: 0.255.

Basic Definitions:

Understanding the structure of a network
Basic definitions: Network

Definition

A network is given by a set of nodes (agents, vertices...) \( N = \{1, \ldots, n\} \) and an adjacency matrix \( g \).

- Default is undirected: \( g \) is symmetric
- Default is irreflexive: \( g_{ii} = 0 \)
- Weighted networks are modelled by matrices of real numbers

```
0 1 1 0 0 1 0 0
1 0 0 0 0 0 0 0
1 0 0 1 1 1 0 0
0 0 1 0 1 0 1 0
0 0 1 1 0 0 0 0
1 0 1 0 0 0 1 0
0 0 0 1 0 1 0 0
```
Basic Definitions: Notation

- A subnetwork $g' \subset g$ if $g'_{ij} \Rightarrow g_{ij}$
- $ij \in g$ stands for $g_{ij} = 1$
- $g$ and $g'$ are isomorphic if there exists a relabelling $\rho$ (i.e., bijection) of $N$ that brings $g$ to $g'$, i.e., such that $g'_{\rho(i)\rho(j)} = g_{ij}$
- $g|_S$ is the restriction of $g$ to $S \subseteq N$
- The neighbourhood of $i \in N$ is $N_i(g) = \{j \mid ij \in g\}$.
- The $k + 1$-neighbourhood of $i$ is defined (recursively) as:

$$N_{i}^{k+1}(g) = N_{i}^{k}(g) \cup \bigcup_{\{j \in N_{i}^{k}(g)\}} N_{j}(g)$$
Basic Definitions: Paths and Cycles

- A path between \( i \) and \( j \) is a sequence of links \( i_1i_2 \ldots i_k \) between distinct nodes with \( i_1 = i \) and \( i_k = j \) (formal: such that \( i_ti_{t+1} \in g \) for all \( t < k \)).

- A geodesic between \( i \) and \( j \) is the shortest path connecting \( i \) and \( j \).

- A walk between \( i \) and \( j \) is a sequence of links \( i_1i_2 \ldots i_k \) with \( i_1 = i \) and \( i_k = j \) (nodes can repeat in a walk).

- A cycle is a walk that starts and end at the same node \( i \).

- \( g \) is connected if there is a path between each pair of nodes. A connected component of \( g \) is a maximal connected subgraph of \( g \).

Exercise: how to calculate the number of walks of length \( k \) between \( i \) and \( j \)? What is the length of the shortest path between \( i \) and \( j \)?
Basic Definitions: Last Slide

- A tree is:
  - A connected graph with no cycles.

- A forest is:
  - Such that each connected component is a tree.

- A star is:
  - A network with one node $i \in \mathbb{N}$ such that $g_{ij} = 1$ for all $j$ and $g_{jj} = 0$ otherwise.

- A circle is:
  - A network with a single cycle and such that each node has exactly two neighbours.

Exercise: how many graphs with 30 nodes?
Basic Definitions: Last Slide

- A **tree** is: a connected graph with no cycles.
- A **forest** is:
Basic Definitions: Last Slide

- A **tree** is: a connected graph with no cycles.
- A **forest** is: such that each connected component is a tree.
- A **star** is:
Basic Definitions: Last Slide

- A **tree** is: a connected graph with no cycles.
- A **forest** is: such that each connected component is a tree.
- A **star** is: a network with one $i \in N$ such that $g_{ij} = 1$ for all $j$ and $g_{jj'} = 0$ otherwise.
- A **circle** is:
Basic Definitions: Last Slide

- A tree is: a connected graph with no cycles.
- A forest is: such that each connected component is a tree.
- A star is: a network with one $i \in N$ such that $g_{ij} = 1$ for all $j$ and $g_{jj'} = 0$ otherwise.
- A circle is: a network with a single cycle and such that each node has exactly two neighbours.

Exercise: how many graphs with 30 nodes?
Understanding the structure of a network: Questions

Let \((N, g)\) be a network, and \(i\) a node:

- How popular is \(i\)? \(\rightarrow\) Degree
- How large is a network? \(\rightarrow\) Diameter
- How tightly connected is \(i\)? \(\rightarrow\) Clustering
- How important is \(i\)? \(\rightarrow\) Centrality
Degree and Degree Distribution

Definition

The degree of node \( i \) is the number of nodes that are connected to \( i \)

\[
d_i(g) = |N_i(g)|
\]

Exercise: how to count this on the adjacency matrix?

Definition

The degree distribution \( p(d) \) of a network \((N, g)\) is the frequency (listed values or probability distribution) of nodes with degree \( d \).

This is a very important description of the network!
The Poisson Distribution

Assume links form randomly with probability $p$.

Degree distribution for large $|N|$ approximates:

$$p(d) = \frac{e^{-(n-1)p}((n-1)p^d)}{d!}$$


Image from wikipedia.
The Scale-free Distribution

Degree is \( p(d) = cd^{-\gamma} \)

- Scale-free because \( \frac{p(kd)}{p(d)} = \frac{p(kc)}{p(c)} \).
- Linear if plotted on a log-log scale.
- Examples: WWW, collaboration networks...(\(2 < \gamma < 3\) usually)
- Typically organised into hubs!


http://www.macs.hw.ac.uk/ pdw/topology/ScaleFree.html
Understanding the structure of a network: Questions

Let \((N, g)\) be a network, and \(i\) a node:

- How popular is \(i\)? \(\rightarrow\) Degree
- How large is a network? \(\rightarrow\) Diameter
- How tightly connected is \(i\)? \(\rightarrow\) Clustering
- How important is \(i\)? \(\rightarrow\) Centrality
The distance between $i$ and $j$ is the length of the geodesic between $i$ and $j$.

**Definition**

The diameter of a network $(N, g)$ (also applies to connected components or subgraphs) is the maximum distance between two nodes in $N$.

*Exercise*: what is the diameter of a tree? of a circle?

Other interesting measures: average path length, minimal path length...

*Exercise*: how to compute minimal path length between $i$ and $j$?
Historically interesting question! Remember the 6 degrees of separation...

Here are some observations:

- Friendship network (Milgram, 1967, letter experiment): median 5.5 on 25% of letters that made it

- Math collaborations network (Grossmann, 2002): mean 7.6 max 27

- The internet (Adamic, Pitkow, 1999): mean 3.1

- Facebook (Backstrom Et Al, 2012): mean 4.74
Understanding the structure of a network: Questions

Let \((N, g)\) be a network, and \(i\) a node:

- **How popular is \(i\)?** → Degree
- **How large is a network?** → Diameter
- **How tightly connected is \(i\)?** → Clustering
- **How important is \(i\)?** → Centrality
Clustering: possible definitions

Are my connections connected between each other? And in average?

Definition

The clustering of node \( i \) is the ratio of pairs of nodes connected to \( i \) that are also connected between each other:

\[
Cl(g) = \frac{\sum_{i} |\{jk \in g \mid k \neq j, j, k \in N_i(g)\}|}{\sum |\{jk \mid k \neq j, j.k \in N_i(g)\}|}
\]

- average clustering coefficient can be measured (two different formulas).
- cliques and transitive triples are the main example of clusters.
- useful to detect small worlds networks

Clustering: example
Let \((N, g)\) be a network, and \(i\) a node:

- How popular is \(i\)? \(\rightarrow\) Degree
- How large is a network? \(\rightarrow\) Diameter
- How tightly connected is \(i\)? \(\rightarrow\) Clustering
- How important is \(i\)? \(\rightarrow\) Centrality
Measures of Centrality I

There are many notions to measure how central a node is (i.e., in a decision network, how powerful):

Degree Centrality

*The degree centrality of node $i$ can be measured by its discounted degree:*

$$C^D_i (g) = \frac{d_i(g)}{|n - 1|}$$

Closeness Centrality

*The closeness centrality of node $i$ is the inverse of the average shortest distance between $i$ and any other node:*

$$C^C_i (g) = \frac{n - 1}{\sum_{i \neq j} \ell(i, j)}$$

Measures of Centrality II

Betweenness Centrality

The average fraction of shortest path between two arbitrary nodes including $i$:

$$C^B_i(g) = \sum_{\{k,j|k\neq j, i\notin \{i,j\}} \frac{P_M(k,j)/P(k,j)}{(n-1)(n-2)/2}$$

An elegant centrality measure: the Katz measure of prestige of node $i$ is the sum of the prestige of the nodes connected to $i$ discounted by their degree.

Exercise 1: show that the vector of Katz prestige of all nodes is an eigenvector of the adjacency matrix discounted by the degrees.

Exercise 2: show that this is equivalent up to a scalar to degree centrality.

Eigenvector Centrality

The eigenvector centrality of node $i$ is the $i$-th coordinate of the eigenvector associated to the largest eigenvalue of the adjacency matrix $g$.

In this lecture we have defined interesting features of networks:

- Networks are represented as adjacency matrices, and this representation is very useful to compute the basic features of a network.
- There are several notions to characterise a node and describe features of a network: how popular (degree), how tightly connected (clustering), how important (centrality) ...

In the next lecture we are going to study how networks form and grow:

- Erdös-Rényi random graphs
- Preferential attachment and scale-free networks
- Strategic network formation