Advanced Network Analysis

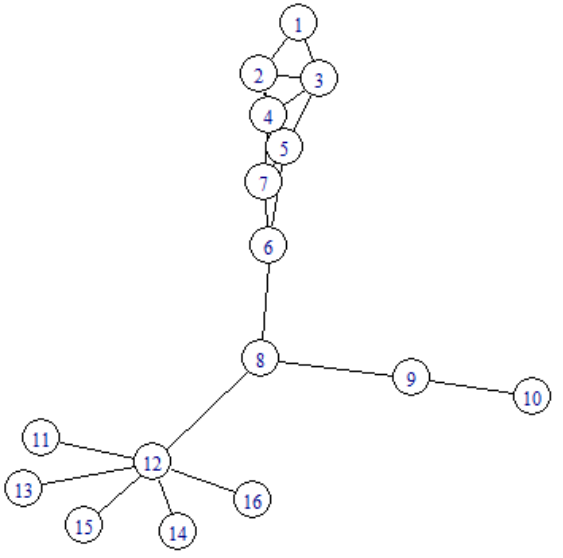
Centrality

Olga Chyzh [www.olgachyzh.com]

Would You Win $1,000,000?

Suppose you are playing the following game:

If you can get all your friends to meet you at the entrance to the ROM in exactly 1 hour, you will win $1,000,000. The catch is that you can only contact one friend, and not all your friends know each other. The picture below is a visualization of your friendship network. Which friend would you call and why?



Discussion

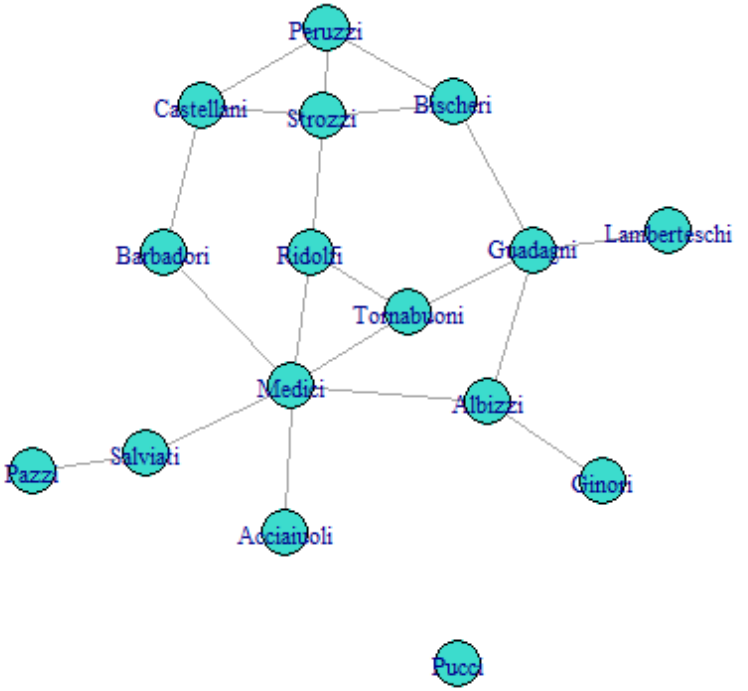
- What criterion helps spread the message in the fewest possible steps?
- What are social science applications of this game?

Network Measures: Centrality

- Centrality measures help understand which node is the most important or central in this network?
  - What do you mean by "important"? "center"?
  - Definition of "center" varies by context/purpose
  - The power a person holds in an organization may be inversely proportional to the number of keys on their keyring
    - A janitor has keys to every office, and no power
    - The CEO does not need a key: people always open the door for her
  - No unanimity on exactly what centrality is or how to measure it.

Florentine Families

Who looks central?



Popular Measures of Centrality

Well ... let's define centrality:

- Degree
- Closeness
- Betweenness
- Eigenvector

Degree centrality

- **Idea**: The nodes with more connections to others are more central
- How to measure:
  - Undirected degree centrality: $\sum_{j:j\neq i} y_{i,j}$
  - Directed outdegree centrality: $\sum_{j:j\neq i} y_{j,i}$
  - Directed indegree centrality: $\sum_{j:j\neq i} y_{i,j}$
- Though simple, degree is often a highly effective measure of the influence or importance of a node
  - In many situations, people with more connections tend to have more power

Florentine Families: an Adjacency Matrix

How would you calculate Albizzi's degree centrality?

```
##              Acciaiuoli Albizzi Barbadori Bischeri Castellani Ginori
## Acciaiuoli            0       0          0        0           0       0
## Albizzi               0       0          0        0           0       1
## Barbadori             0       0          0        0           1       0
## Bischeri              0       0          0        0           0       0
## Castellani            0       0          1        0           0       0
## Ginori                0       1          0        0           0       0
## Guadagni              0       1          0        1           0       0
## Lamberteschi          0       0          0        0           0       0
## Medici                1       1          1        0           0       0
## Pazzi                 0       0          0        0           0       0
## Peruzzi               0       0          0        1           1       0
## Pucci                 0       0          0        0           0       0
## Ridolfi               0       0          0        0           0       0
## Salviati              0       0          0        0           0       0
## Strozzi               0       0          0        1           1       0
## Tornabuoni            0       0          0        0           0       0
```

Closeness centrality

- **Idea**: If a node is far away from all other nodes, then it should be less central ... or to put it another way, the more central a node, the lower its total distance to all other nodes
- How to measure:
  - (geodesic) distance: $d_{i,j}$ is the minimal path length from i to j
  - closeness centrality: $\frac{1}{\sum_{j:j \neq i} d_{i,j}}$
- Closeness can also be regarded as a measure of how long it will take to spread information from a node to all other nodes sequentially
- This measure won't be useful for disconnected graphs ... why?

Betweenness

- **Idea**: A node is central if it acts as a bridge to other nodes
- How to measure in words:
  - For each pair of nodes, compute the geodesic distance (shortest path between them)
  - Then for each node, determine the fraction of shortest paths that go through the actor in question
  - End by summing this fraction over all pairs of nodes

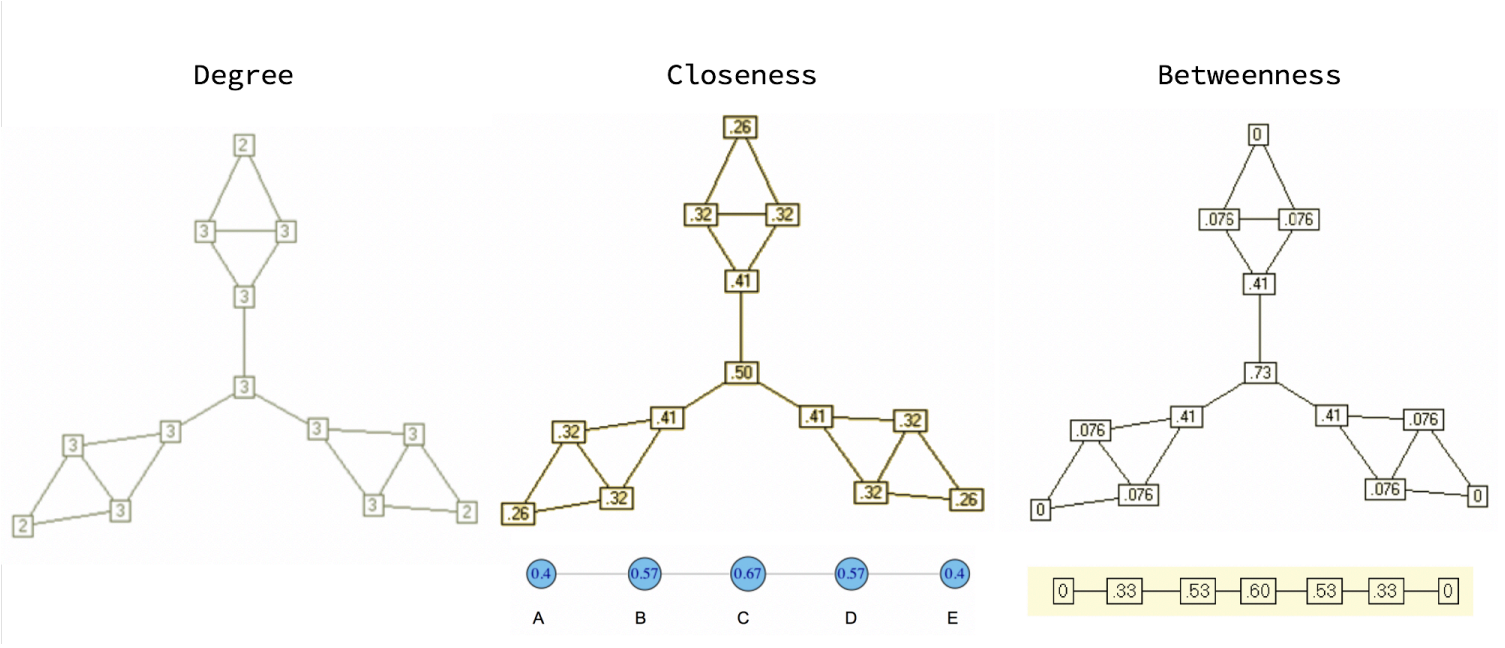
Betweenness

- How to measure a bit more formally:
  - Say $g_{j,k}$ equals the number of geodesics between nodes j and k
  - Say $g_{j,k}(i)$ equals the number of geodesics between nodes j and k going through i
  - Then betweenness centrality for actor i: $\sum_{j<k} \frac{g_{j,k}(i)}{g_{j,k}}$

Betweenness

- Simple way to think of $\frac{g_{j,k}(i)}{g_{j,k}}$ is the probability that a "message" from j to k goes through i
  - j and k have $g_{j,k}$ routes of communication
  - i is on $g_{j,k}(i)$ of these routes
  - a randomly selected path contains i with probability $\frac{g_{j,k}(i)}{g_{j,k}}$
- Examples where this might be useful?

Comparison of these measures (Thanks to Arifuzzaman & Bhuiyan)



Comparison of these measures (Thanks to Arifuzzaman & Bhuiyan)

	Low Degree	Low Closeness	Low Betweenness
High Degree		Embedded in cluster that is far from the rest of the network	Ego's connections are redundant - communication bypasses him/her
High Closeness	Key player tied to important/active alters		Probably multiple paths in the network, ego is near many people, but so are many others
High Betweenness	Ego's few ties are crucial for network flow	Very rare cell. Would mean that ego monopolizes the ties from a small number of people to many others.	

One more ... Eigenvector

- **Idea**: An actor is more central if it is connected to other more central actors
- Eigenvector centrality: centrality of each node is proportional to the sum of the centralities of its neighbors (let c_i^e denote the eigenvector centrality of actor i):
  - $c_i^e = \frac{1}{\lambda} \sum_{j:j \neq i} y_{ij} c_j^e$
- Based on some matrix algebra:
  - $Yc^e = \lambda c^e$
  - Vector c^e satisfying the above equation is an **eigenvector** of Y
- Generally, there are multiple eigenvectors, centrality is taken to be the one corresponding to the largest value of λ
- Examples of where this might be useful?

Trillion dollar application

`Google Describing PageRank`: PageRank relies on the uniquely democratic nature of the web by using its vast link structure as an indicator of an individual page's value. In essence, Google interprets a link from page A to page B as a vote, by page A, for page B. But, Google looks at more than the sheer volume of votes, or links a page receives; it also analyzes the page that casts the vote. Votes cast by pages that are themselves "important" weigh more heavily and help to make other pages "important."

Your Turn

For each of the following networks, think of the best measure of centrality to measure the amount of influence in different contexts.

- countries connected by trade relations
- a network of student friendships on a university campus
- a network of legislators connected by co-sponsorships of bills
- a network of CEOs connected based on their undergraduate institutions

Lab: Calculate Centrality

Working example

Florentine marriages (Padgett & Ansell 1993)

```
# load datasets
data(flo) #this dataset is available from the -network- package
flo[1:5,1:5]
```

```
##             Acciaiuoli Albizzi Barbadori Bischeri Castellani
## Acciaiuoli          0       0          0        0           0
## Albizzi             0       0          0        0           0
## Barbadori           0       0          0        0           1
## Bischeri            0       0          0        0           0
## Castellani          0       0          1        0           0
```

Florentine families

```
# convert to igraph object
library(igraph)
g = graph_from_adjacency_matrix(flo, mode='undirected', diag=FALSE)
g<-  delete_vertices(g, V(g)[degree(g) == 0])
plot(g)
```

Centrality

Lets calculate each centrality measure in R (using the `igraph` package)

```
sort(igraph::degree(g), decreasing=TRUE)[1:6]
```

```
##     Medici   Guadagni    Strozzi    Albizzi   Bischeri Castellani
##          6          4          4          3          3          3
```

```
sort(igraph::closeness(g), decreasing=TRUE)[1:6]
```

```
##     Medici    Ridolfi    Albizzi Tornabuoni   Guadagni  Barbadori
## 0.04000000 0.03571429 0.03448276 0.03448276 0.03333333 0.03125000
```

```
sort(igraph::betweenness(g), decreasing=TRUE)[1:6]
```

```
##   Medici Guadagni  Albizzi Salviati  Ridolfi Bischeri
## 47.50000 23.16667 19.33333 13.00000 10.33333  9.50000
```

```
sort(igraph::eigen_centrality(g)$vector, decreasing=TRUE)[1:6]
```

```
##     Medici    Strozzi    Ridolfi Tornabuoni   Guadagni   Bischeri
##  1.0000000  0.8272688  0.7937398  0.7572302  0.6718805  0.6572037
```

Your Turn

- Plot the Florentine network, such that node size is proportionate to each centrality measure. How do they compare?