Basic of networks

Indiana University,

Emerson Melo

- There is a set of nodes, $N = \{1, \ldots, n\}$, where n is a finite number.
- Relationships between nodes are conceptualized in terms of binary variables, so that a relationship either exists or does not exist.
- Denote by $g_{ij} \in \{0, 1\}$ a relationship between two nodes i and j.

- The variable g_{ij} takes on a value of 1 if there exists a link between i and j and 0 otherwise.
- The set of nodes taken along with the links between them defines the network; this network is denoted by G and the collection of all possible networks on n nodes is denoted by $\mathscr{G}$.

Definition

The network G is undirected if $g_{ij} = g_{ji}$ for all pair i, j

Definition

The network G is undirected if $g_{ij} = g_{ji}$ for all pair i, j

Definition

The network G is directed if the links have a direction

Definition

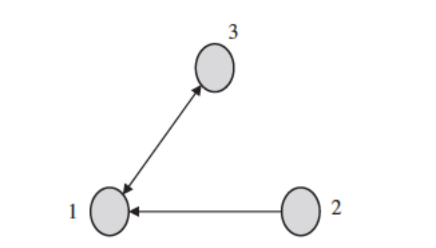
The network G is undirected if $g_{ij} = g_{ji}$ for all pair i, j

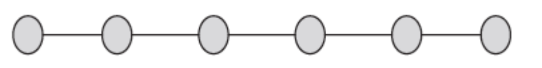
Definition

The network G is directed if the links have a direction

Definition

The network G is weighted (directed or undirected) if $g_{ij} \in \mathbb{R}_+$ for all pair of nodes


3
1
2

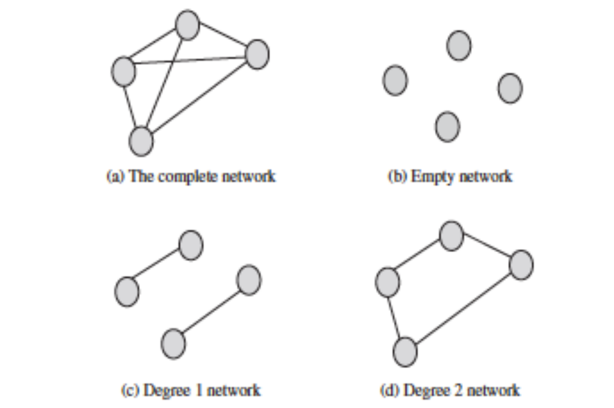


- The network G may be interpreted as a connection matrix
- Given a network G, $G + g_{ij}$ and $G - g_{ij}$ have the natural interpretation
- When $g_{ij} = 0$ in G, $G + g_{ij}$ adds the link $g_{ij} = 1$, while if $g_{ij} = 0$ in G, then $G + g_{ij} = G$.

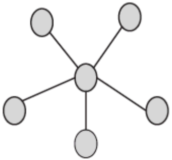
- Similarly, if $g_{ij} = 1$ in G, $G - g_{ij}$ deletes the link g_{ij}, while if $g_{ij} = 0$ in G, then $G - g_{ij} = G$
- Let $N_i(G) = \{j \in N | g_{ij} = 1\}$ denotes the node with which node i has a link; this set will be referred to as the neighbors of i
- Let $\eta_i(G) = |N_i(G)|$ denote the number of neighbors of node i in network G

- Let $\mathbf{N}_1(G),\ldots,\mathbf{N}_{n-1}(G)$ be a division of nodes into distinct groups, where nodes belong to the same group if and only if they have the same number of links, i.e., $i,j \in \mathbf{N}_k(G)$, $k = 1,\ldots,n-1$, if and only if $\eta_i(G) = \eta_j(G)$
- With this notation in hand we can now describe a number of well-known networks.

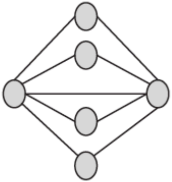
- A network G is said to be *regular* if every node has the same number of links, i.e., $\eta_i(G) = \eta$ for all $i \in N$
- The *complete* network, G^c is a regular network in which $\eta = n - 1$
- The *empty* network, G^e, is a regular network in which $\eta = 0$


(a) The complete network
(b) Empty network
(c) Degree 1 network
(d) Degree 2 network

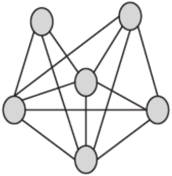
- A *core-periphery* network structure describes the following situation. There are two groups of nodes, $\mathbf{N}_1(G)$ and $\mathbf{N}_k(G)$ with $k > |\mathbf{N}_k(G)|$. Nodes in $\mathbf{N}_1(G)$ constitute the periphery and have a single link each and this link is with a node in $\mathbf{N}_k(G)$
- Nodes in the set $\mathbf{N}_k(G)$ constitute the core and are fully linked with each other and with a subset of nodes in $\mathbf{N}_1(G)$



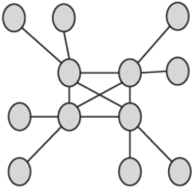
(a) Star network



(b) Interlinked star (two centers)



(c) Interlinked star (three centers)



(d) Core–periphery network

- An *interlinked stars* network consists of two groups $\mathbf{N}_1(G)$ and $\mathbf{N}_{n-1}(G)$ which satisfy the following condition: $N_i(G) = \mathbf{N}_{n-1}(G)$ for $i \in \mathbf{N}_k(G)$
- The star network is again a special case of such an architecture with $|\mathbf{N}_{n-1}(G)| = 1$ and $|\mathbf{N}_1(G)| = n-1$
- In an interlinked star network, nodes which have $n-1$ links are referred to as central nodes or as hubs, while the complementary set of nodes are referred to as peripheral nodes or as spokes
- A *line* network consists of two groups of nodes $\mathbf{N}_1(G)$ and $\mathbf{N}_2(G)$, with $|\mathbf{N}_1(G)| = 2$ and $|\mathbf{N}_2(G)| = n-2$

- The *degree* of node i is the number of i's direct connections; so $\eta_i(G) = N_i(G)$ denotes the degree of node i in network G
- The *degree distribution* in a network is a vector P, where

$$P(k) = \frac{|\mathbf{N}_k(G)|}{n}$$

  is the frequency/fraction of nodes with degree k
- Thus $P(k)$ for each k, and $\sum_{k=0}^{n-1} P(k) = 1$

- This degree distribution has support on $\mathscr{D} = \{1, \dots, n-1\}$
- The average degree in network G is defined as

$$\hat{\eta}(G) = \sum_{k=0}^{n-1} P(k)k = \sum_{i \in N} \frac{\eta_i(G)}{n} \tag{1}$$

- In a star network the degree distribution has support on degrees 1 and $n-1$, with $n-1$ nodes having degree 1, and 1 node having degree $n-1$
- The average degree in a star is $2 - \frac{2}{n}$

- An important concern in the study of networks is the variation in the degrees
- This variation has an important interpretation: degree may be related to node behavior and well-being
- One of the primary motivations for the study of networks in economics is the issue of how nodes extract advantages on account of their connections
- The variance in the degree distribution is defined as

$$Var(G) = \sum_{k=0}^{n-1} P(k)[\hat{\eta}(G) - k]^2 \tag{2}$$

- The degree variance in a star grows with n, while the variance in any regular network is 0 for all n
- The range of degrees in network G is

$$R(G) = \max_{i \in N} \eta_i(G) - \min_{j \in N} \eta_j(G) \tag{3}$$

- The range has a maximum value of $n-2$ and a minimum value of 0
- The range in a star is $n-2$, while the range in any regular network is 0

- The description of a network in terms of a degree distribution allows for an elegant way to study the addition and the redistribution of links
- The idea of adding links is captured in the relation of *first-order stochastic domination*
- Similarly, the idea of redistributing links is captured in the relation of *mean-preserving spreads* and *second-order stochastic domination*
- Given a degree distribution P, let the cumulative distribution function be denoted by $\mathscr{P} : \{0, 1, \ldots, n\} \longrightarrow [0, 1]$, where

$$\mathscr{P}(\eta) = \sum_{x=0}^{\eta} P(x) \tag{4}$$

- Let P and P' be two degree distributions defined on $\{0, 1, \ldots, n\}$ and $\mathscr{P}$ and $\mathscr{P}'$ the two corresponding cumulative distribution functions, respectively

Definition

P first-order stochastically dominates (FOSD) P' if and only if

$$\mathscr{P}(k) \leq \mathscr{P}'(k),$$

for every $k \in \{0, 1, 2, \ldots, n-1\}$

Definition

P second-order stochastically dominates P' if and only if

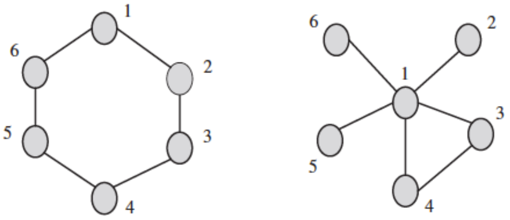
$$\sum_{k=0}^{x} \mathscr{P}(x) \leq \sum_{k=0}^{x} \mathscr{P}'(x),$$

for every x

Definition

P' is a mean-preserving spread of P if and only if P and P' have the same mean and P second-order stochastically dominates P'

- A simple example of first-order shift in degree distribution arises when we move from a regular network with degree k to a regular network with degree $k+1$
- A simple example of a second-order shift arises when we move from a cycle with $n=6$ nodes to a network in which node 1 is linked to all nodes, nodes $2,3$, and 4 have just this one link with node 1, while nodes 5 and 6 have two links each, a link with node 1 and a link with each other



(a) Cycle (b) Relocated links

- A major concern throughout the course will be the ways in which a node can be reached from another node in the network
- The first step in understanding this issue is the notion of walk: a walk is a sequence of nodes in which two nodes have a link between them in the network (they are neighbors)

- A node or a link may appear more than once in a walk
- A walk is the most general sequence of nodes and links possible in a network, subject to the constraint that any two consecutive nodes must have a link in the network
- The length of a walk is simply the number of links it crosses; this is simply equal to the number of nodes involved minus one
- A walk in which all links are distinct is called a *trail*

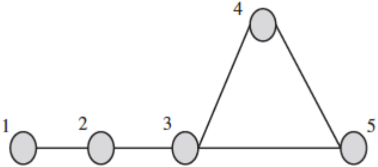
- A trail in which there are three or more nodes and the initial and the end node are the same is called a *cycle*
- A trail in which every node is distinct is called a *path*
- Formally, there is a *path* between two distinct nodes i and j either if $g_{ij} = 1$ or if there is a set of distinct intermediate nodes $j_1, \ldots, j_n$ such that $g_{ij_1} = g_{j_1 j_2} = \cdots g_{j_n j} = 1$

- Next example represents a network with $n = 5$
- A possible walk in this network is $2, 3, 4, 3, 2$
- This walk contains the links g_{23} and g_{34} twice and the nodes 2 and 3 also appear twice each
- This walk is therefore not a trail
- A possible trail in the network is $3, 4, 5, 3$
- However, since node 3 appears twice this trail is not a path
- A possible path in this network is 2, 3, 4, 5

Walk: 2, 3, 4, 3, 2.

Trail/cycle: 3, 4, 5, 3.

Path: 2, 3, 4, 5.



- Two nodes belong to the same *component* if and only if there exists a path between them
- A network is *connected* if there exists a path between any pair of nodes $i, j \in N$
- It follows that there exists only one component in a connected network

- In the case of an unconnected network, the components can be ordered in terms of their size, and the network has a *giant component* if, informally speaking, the largest component covers a relatively large fraction of the nodes while all other components are small

- The notion of minimality plays an important role in networks

- Intuitively speaking, minimality of a network reflects the idea that no link is “superfluous”
- A component is said to be minimal if the deletion of any single link in the component breaks the component into two components
- A network is said to be minimal if the deletion of any single link in the network increases the number of components by 1

- The *geodesic distance* between two nodes i and j in network G is the length of the shortest path between them, and will be denoted by $d(i, j : G)$
- If there is no path between i and j in network G, then by convention set $d(i, j : G) = \infty$
- When G is connected , the average distance between nodes of a network G is

$$d(G) = \frac{\sum_{i \in N} \sum_{j \in N} d(i, j; G)}{n(n-1)} \tag{5}$$

- The centrality of a node in a network captures a number of ideas relating to the *prominence* of a node in a network
- Degree centrality captures the relative prominence of a node vis-a-vis other nodes in terms of the degree
- The (standardized) degree centrality of a node i in network G is simply the degree of this node divided by the maximum possible degree:

$$C_d(i;G) = \frac{\eta_i(G)}{n-1} \tag{6}$$

- We now turn to a measure of centrality which is based on proximity

- The total distance from node i to all other nodes in the network G is

$$\sum_{j \neq i} d(i, j; G)$$

- This distance will be related to the number of nodes in a network and to facilitate comparison across networks of different size, it is useful to normalize the measure by multiplying with the minimum possible total distance, which is $n - 1$

- The *closeness centrality* of node i in network G is defined as

$$C_c(i;G) = \frac{n-1}{\sum_{j\neq i} d(i,j;G)} \tag{7}$$

- This measure of centrality has a natural analogue at the aggregate network level

- This measure is built upon differences across nodes in a network and is normalized to account for maximum attainable differences

- The measure of closeness centrality is based solely on the length of the shortest paths between nodes in a network
- In some contexts it is quite possible that links are not perfectly reliable and so the number of paths of different lengths may all matter
- More generally, it is possible that actions of a person may have implications for the actions of her neighbors, which may in turn feedback on the initial individual, and so on

- These considerations motivate the study of a notion of centrality which allows for a richer range of direct and indirect influences in a network
- Bonacich (1972) developed such a measure of centrality and we now turn to it
- **Bonacich's measure is one of the most popular measures in network economic**

- Consider the adjacency (connection) matrix $\mathbf{G}$ of network G
- In this matrix an entry in a square corresponding to a pair $\{i, j\}$ signifies the presence or absence of a link
- Let $\mathbf{G}^k$ be the kth power of the matrix
- The 0 power matrix $\mathbf{G}^0 = \mathbf{I}$, the $n \times n$ identity matrix
- In $\mathbf{G}^k$, an entry g_{ij}^k measures the “number” of walks of length k that exist between players i and j in network

Example

Consider a network with three players, $1, 2$, and 3. Suppose links take on values of 0 and 1, and let the network consist of two links, $g_{12} = g_{23} = 1$. We can find $\mathbf{G}$ and $\mathbf{G}^2$

	1	2	3
1	0	1	0
2	1	0	1
3	0	1	0

	1	2	3
1	1	0	1
2	0	2	0
3	1	0	1

- Thus there is one walk of length 2 between 1 and 1 and between 3 and 3, but two walks of length 2 between 2 and 2
- There are no other walks of length 2 in this network

- Let $a \geq 0$ be a scalar and let $\mathbf{I}$ be the identity matrix
- Define the matrix $\mathbf{M}(G,a)$ as follows:

$$\mathbf{M}(G,a) = [\mathbf{I} - a\mathbf{G}]^{-1} = \sum_{k=0}^{\infty} (a\mathbf{G})^k \tag{8}$$

- This expression is well-defined so long as a is sufficiently small
- The entry $m_{ij}(G,a) = \sum_{k=0}^{\infty} a^k g_{ij}^k$ counts the total number of walks in G from i to j, where walks of length k are weighted by a factor a^k

- Given parameter a, the Bonacich centrality vector is defined as

$$C_B(G,a) = [\mathbf{I} - a\mathbf{G}]^{-1} \cdot \mathbf{1} = \mathbf{M}(G,a) \cdot \mathbf{1}, \tag{9}$$

where $\mathbf{1}$ is the (column) vector of 1s

- In particular, the Bonacich centrality of player i is

$$C_B(i;G,a) = \sum_{j=1}^{n} m_{ij}(G,a) \tag{10}$$

- This measure of centrality counts the total number of (suitably weighted) walks of different lengths starting from i in network G.

- To see this note that (10) can be rewritten as follows:

$$C_B(i;G,a) = m_{ii}(G,a) + \sum_{j \neq i} m_{ij}(G,a) \tag{11}$$

- Since $\mathbf{G}^0 = \mathbf{I}$, it follows that $m_{ii}(G,a) \geq 1$ and so for every player i in any network, $C_B(i;G,a) \geq 1$. It is exactly equal to 1 when $a = 0$

- The Bonacich centrality of a node can also be expressed as a function of the centrality of its neighbors
- Let $\rho(G)$ be the (largest) eigenvalue of the adjacency matrix $\mathbf{G}$
- The Bonacich centrality of a node can then be defined as

$$C_B(i;G,a) = \frac{1}{\rho(G)} \sum_{j \in N} g_{ij} C_B(j;G,a) \tag{12}$$