Social and Economic Network Analysis Notes

Michael

Economic Department, University of Konstanz

School of Economics, University of Nottingham

1 Introduction

This is the collection of my notes on social and economic network analysis. I do not declare the originality of the content as they are reorganized and rephrased from three books: Easley et al. (2010), Jackson (2010), and Kolaczyk and Csárdi (2014).

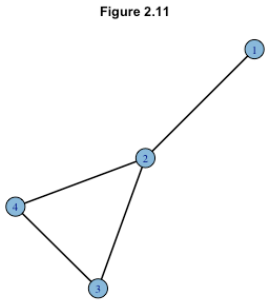
The rise of internet and social medias enables the researchers to access millions of data related to networks. With those datasets, people are very interested on questions like: how highly connected systems operate; how the news (or even disease) spread around the world, etc. In this notes, I will try to answer those questions by combining the usage of mathematics and computer science. This notes also serves for preparing my seminar essay in Konstanz.

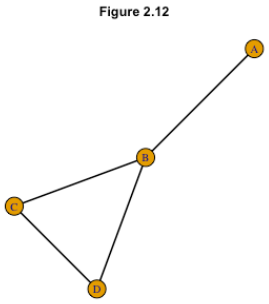
2 Overview and aspects of Networks

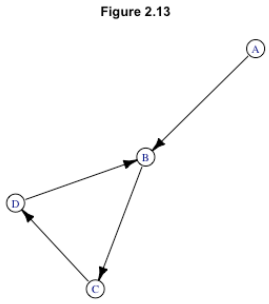
In the most basic sense, a network is any collection of objects in which some pairs of those objects are connected by links. We can see that this definition is very flexible. In the sense of defining networks, we are also on the way to understand networks. Later, you will realize the definition of patterns of networks and measurements of networks can give us astonishing view on Networks.

2.1 Basic Definitions

A *graph* is a way of specifying relationships among a collection of items. A graph consists of a set of objects, called *nodes*, with certain pairs of these objects connected by links called edges. We say that two nodes are neighbors if they are connected by an edge. To express asymmetric relationships - for example, that A points to B but not vice versa. For this purpose, we define a *directed graph* to consist of a set of nodes, as before, together with a set of *directed edges*; each directed edge is a link from one node to another, and the direction is important. Directed graph is very useful to study the transformation of information. Figure 2.13 shows the network structure with direction.







A *path* is simply a sequence of nodes with the property that each consecutive pair in the sequence is connected by an edge. A *connected component* of a graph is a subset of the nodes such that (i) every node in the subset has a path to every other and (ii) the subset is not part of some larger set with the property that every node can reach every other.

There is a stylized fact when one looks across a range of network data sets. Large, complex networks often have what is called a *giant component*, which is deliberately informal term for a connected component that contains a significant fraction of all nodes.

Breadth-First Search. To measure the distance between nodes, we can use compute to find how many paths or steps we need to reach the furthest node. There is a small-world phenomenon in the giant component of networks. The idea is that the world looks "small" when you think of how short a path of friends it takes to get from you to almost anyone else in the network. Figure 2.1 shows that the estimated average distance is 6.6 for all active Microsoft Instant Messengers.

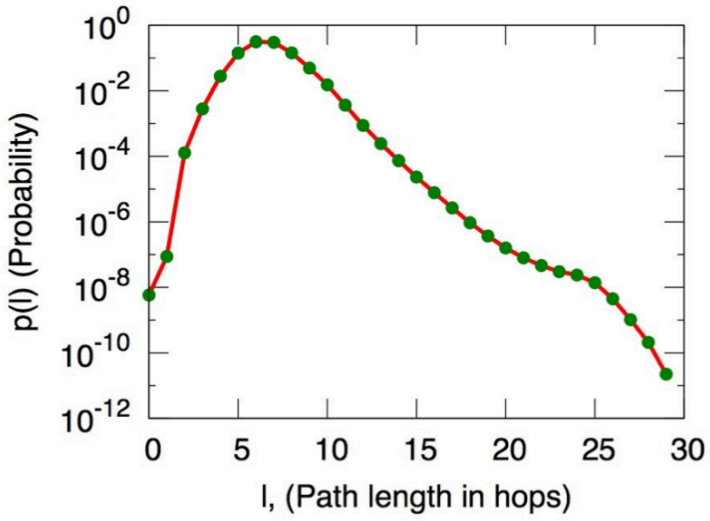


Figure 2.1: The distribution of distances active Microsoft Instant Messenger

2.2 Representing and Measuring Networks

This section presents some of the fundamentals of how network are represented, measured, and characterized. It provides basic concepts and definitions that are the basis for the language of network research.

The set $N = \{1, \cdots, n\}$ is the set of *nodes* that are involved in a network of relationships. Nodes are also referred as "vertices", "individuals", "agents" or "players", depending on the setting. A *graph* (N, g) consists of a set of nodes $N = \{1, \cdots, n\}$ and a real-valued $n \times n$ matrix g, where g_{ij} represents the relation between i and j. This matrix is often referred to as the *adjacency matrix*. It is standard to use the values of either 0 or 1 to represent the unweighted network. In the case in which the entries of g take on more than two values and can track the intensity level of relationships, the graph is referred to as a *weighted* graph.

A network is *directed* if it is possible that $g_{ij} \neq g_{ji}$, and a network is *undirected* $g_{ij} = g_{ji}$ for all nodes i and j. Parts of the literature refer to directed graph as *digraphs*. For instance, if $N = \{1, 2, 3\}$, then

$$g1 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \qquad g2 = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$$

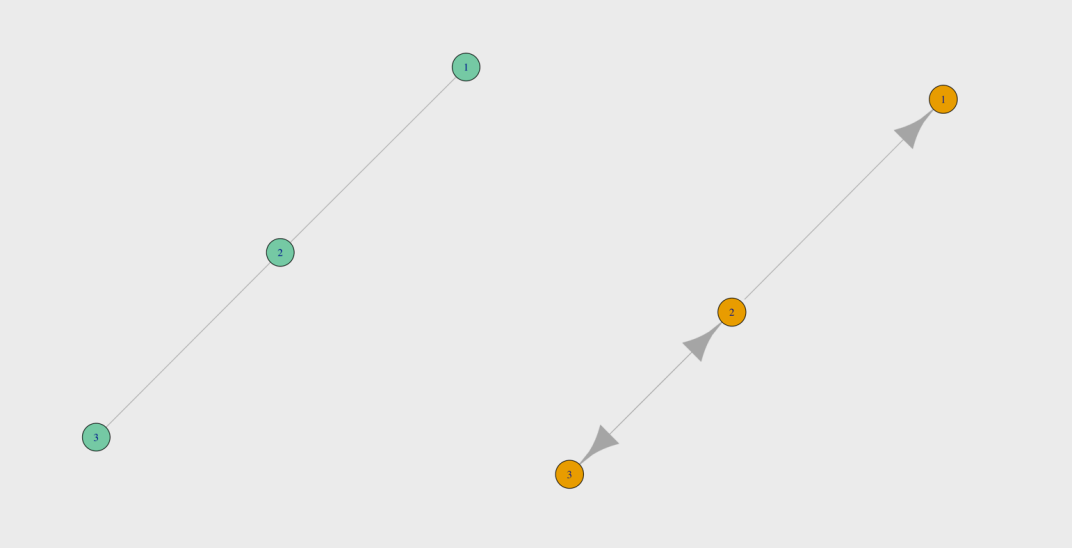


Figure 2.2: Undirected and directed network for g1 and g2

There are equivalent ways of representing a graph. Instead of viewing g as an $n \times n$ matrix, it is sometimes easier to describe a graph by listing all links or edges in the graph. That is, we can view a graph as a pair (N, g), where g is the collection of links that are listed as a subsets of N of size 2. For instance, the network $g1$ can be written as $g1 = \{\{1, 2\}, \{2, 3\}\}$, or $g2 = \{12, 23\}$.

Paths and Cycles. Much of the interest in networked relationships comes from the fact that individual nodes benefit (or suffer) from indirect relationships. A *path* in a network $g \in G(N)$ between nodes i and j is a sequence of links $i_1i_2, i_2i_3, \cdots, i_{K-1}i_K$ such that $i_ki_{k+1} \in g$ for each $k \in \{1, \cdots, K-1\}$, with $i_1 = i$ and $i_K = j$, and such that

each node in the sequence $i_1, \cdots, i_K$ is distinct. A *walk* in a network $g \in G(N)$ between nodes i and j is a sequence of links $i_1i_2, i_2i_3, \cdots, i_{K-1}i_K$ such that $i_ki_{k+1} \in g$ for each $k \in \{1, \cdots, K-1\}$, with $i_1 = i$ and $i_K = j$. The distinction between a path and a walk is whether all involved nodes are distinct. A walk may come back to a given node more than once, whereas a path is a walk that never hits the same node twice. A *cycle* is a walk $i_1i_2, i_2i_3, \cdots, i_{K-1}i_K$ that starts and ends at the same node and such that all other nodes are distinct. A *geodesic* between nodes i and j is a shortest path between these nodes; that is, a path with no more links than any other path between these nodes.

Note that for the convention of setting $g_{ii} = 0$ and $g^2 = g \times g$ tells us how many walks there are of length 2 between any two nodes. For instance for the network

$$g = \begin{pmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{pmatrix} \qquad g^2 = \begin{pmatrix} 2 & 0 & 0 & 2 \\ 0 & 2 & 2 & 0 \\ 0 & 2 & 2 & 0 \\ 2 & 0 & 0 & 2 \end{pmatrix}$$

So, for instance, there are two walks between 1 and 4 of length 2 (passing between 2 and 3, respectively). Figure 2.3 shows the network based on the above matrix.

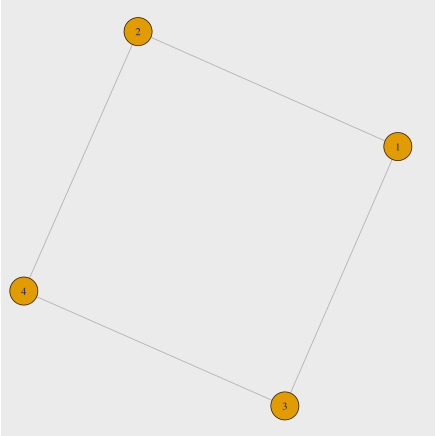


Figure 2.3: Network with four nodes

A *directed walk* in a network $g \in G(N)$ is a sequence of links $i_1i_2, \cdots, i_{K-1}i_K$ such that $i_ki_{k+1} \in g$ (that is, $g_{i_k,i_{k+1}} = 1$) for each $k \in \{1, \cdots, K-1\}$. A *directed path* in a network $g \in G(N)$ is a sequence of links $i_1i_2, \cdots, i_{K-1}i_K$ such that $i_ki_{k+1} \in g$ (that is, $g_{i_k,i_{k+1}} = 1$) for each $k \in \{1, \cdots, K-1\}$, with $i_1 = i$ and $i_K = j$, such that each node in the sequence $i_1, \cdots, i_K$ is distinct. A *directed cycle* in a network $g \in G(N)$ is a sequence of links $i_1i_2, \cdots, i_{K-1}i_K$ such that $i_ki_{k+1} \in g$ (that is, $g_{i_k,i_{k+1}} = 1$) for each $k \in \{1, \cdots, K-1\}$, with $i_1 = i_K$. To be more specific, given a directed network g let $\hat{g}$ denote the undirected network obtained by allowing an undirected link for each directed one present in g. That is, let $\hat{g}_{ij} = \max(g_{ij}, g_{ji})$.

Components and Connected Subgraphs. In many applications it is important to track which nodes can reach which other nodes through paths in the network. This tracking ability plays a critical role in phenomena like contagion, learning and the diffusion of various behaviors through a social network.

A network (N, g) is *connected* if **every** two nodes in the network are connected by some path in the network. That is, (N, g) is connected if for each $i \in N$ and $j \in N$ there

exists a path in (N, g) between i and j. A *component* of a network (N, g) is a nonempty subnetwork (N', g') such that $\emptyset \neq N' \subset N, g' \subset g$,

- (N', g') is connected, and
- if $i \in N'$ and $ij \in g$, then $j \in N'$ and $ij \in g'$.

Thus the components of a network are the distinct maximal connected subgraphs of a network. A *tree* is a connected network that has no cycles. A *forest* is a network such that each component is a tree. A *star* is a network in which there exists some node i such that every link in the network involves node i. A connected network is a tree if and only if it has $n-1$ links. A tree has at least two leaves, where leaves are nodes that have exactly one link. In a tree, there is a unique path between any two nodes.

Neighborhood. The *neighborhood* of a node i is the set of nodes that i is linked to

$$N_i(g) = \{j : g_{ij} = 1\}$$

Given some set of nodes S, the neighborhood of S is the union of the neighborhoods of its members. That is

$$N_S(g) = \underset{i \in S}{\cup} N_i(g) = \{j : \exists i \in S, g_{ij} = 1\}$$

We can also talk about extended neighborhoods of a node, for instance of all the nodes that can be reached by walks of length no more than 2, and so on. The two-neighborhood of a node i is

$$N_i^2(g) = N_i(g) \cap \left(\underset{j \in N_i(g)}{\cup} N_j(g) \right)$$

Inductively, all nodes that can be reached from i by walks of length no more than k make up the $k - neighborhood$ of i, which can be defined in a recursive way:

$$N_i^k(g) = N_i(g) \cup \left(\underset{j \in N_i(g)}{\cup} N_j^{k-1}(g) \right)$$

The *degree* of a node is the number of links that involve that node, which is the cardinality of the node's neighborhood. Thus node i's degree in a network g, denoted $d_i(g)$, is

$$d_i(g) = \#\{j : g_{ji} = 1\} = \#N_i(g)$$

In the case of a directed network, the above calculation is the node's *in-degree*. The *out-degree* of node i is the corresponding calculation $\#\{j : g_{ij} = 1\}$. These density of a network keeps track of the relative fraction of links that are presented, and is simply the average degree divided by $n - 1$.

$$undg = \begin{pmatrix} 0 & 1 & 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 \end{pmatrix} \qquad dg = \begin{pmatrix} 0 & 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 \end{pmatrix}$$

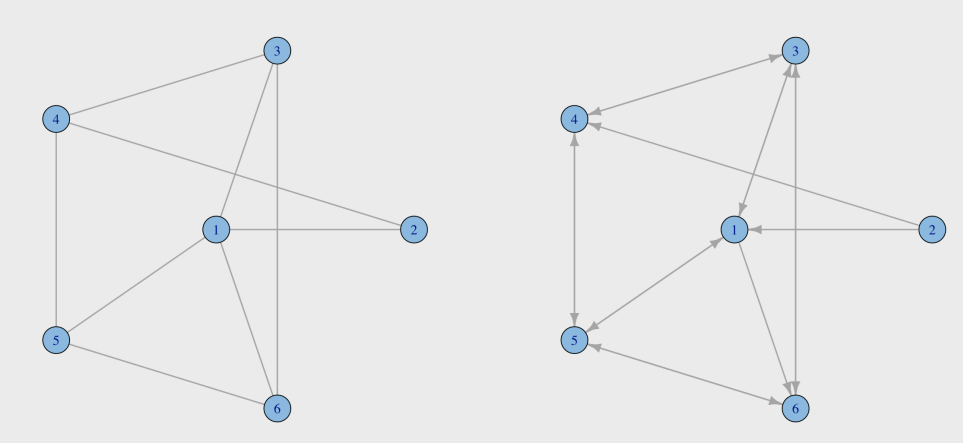


Figure 2.4: Undirected and directed networks

The degree for corresponding nodes in the above undirected network is: $4, 2, 3, 3, 3, 3$. And there is no difference for in-degree and out-degree. For the directed network, the corresponding in-degree is $3, 0, 3, 3, 3, 3$; and out-degree is $3, 2, 3, 2, 3, 2$.

A fundamental characteristic of a network is its degree distribution. The *degree distribution* of a network is a description of the relative frequencies of nodes that have different degrees. That is, $P(d)$ is the fraction of nodes that have degree d under a degree distribution P.

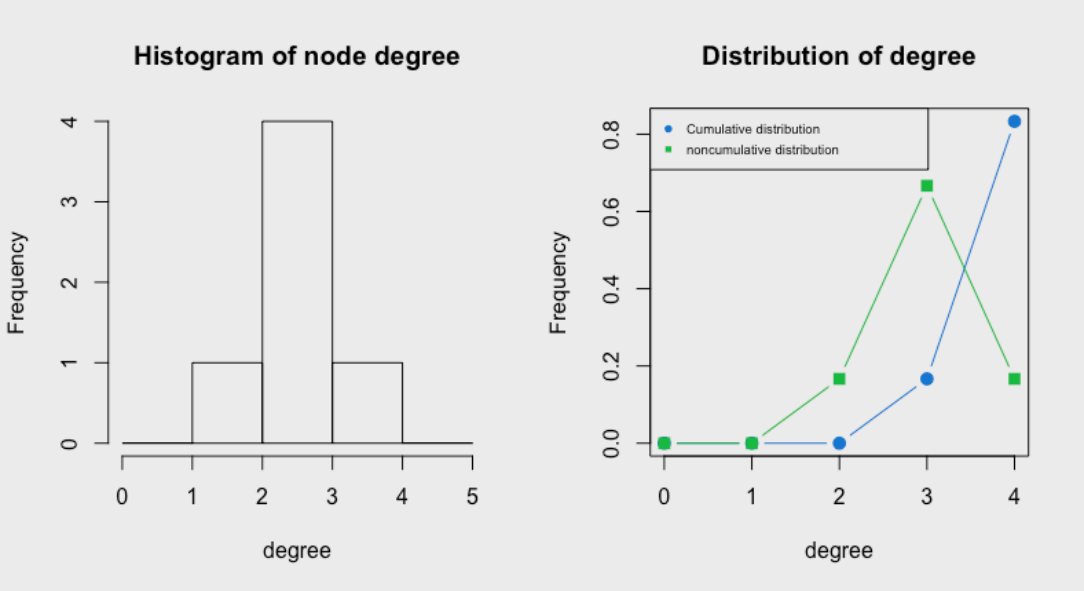


Figure 2.5: Degree Distribution of undirected network

Diameter and Average Path Length. The *distance* between two nodes is the length of (number of links in) the shortest path or *geodesic* between them. This concept leads us to another important characteristic of a network: its *diameter*. The *diameter* of a network is the largest distance between any two nodes in the network. The diameter is one measure of path length, but it only provides an upper bound. *Average path length* between nodes is another measure that captures related properties.

Cliquishness, Cohesiveness, and Clustering. A *clique* is a maximal completely

connected subnetwork of a given network. That is, if some set of nodes $S \subset N$ are such that $g|_S$ is the complete network on the nodes S, and for any $i \in N$ $Sg|_{S\cup\{i\}}$ is not complete, then the nodes S are said to form a clique. The clustering measure is presented by

$$Cl(g) = \frac{\sum_i \#\{jk \in g | k \neq j, j \in N_i(g), k \in N_i(g)\}}{\sum_i \#\{jk | k \neq j, j \in N_i(g), k \in N_i(g)}$$

individual clustering for a node i is defined:

$$Cl_i(g) = \frac{\#\{jk \in g | k \neq j, j \in N_i(g), k \in N_i(g)\}}{\#\{jk | k \neq j, j \in N_i(g), k \in N_i(g)\}}$$

Centrality. Measures of centrality can be categorized into four main groups depending on the types of statistics on which they are based:

1. degree - how connected a node is;
2. closeness - how easily a node can reach other nodes;
3. betweenness - how important a node is in terms of connecting other nodes;
4. neighbors' characteristics - how important, central, or influential a node's neighbors are.

The *degree centrality* of a node is simply

$$dc(i) = \frac{d_i(g)}{(n-1)}$$

One obvious closeness-based measure is just the inverse of the average distance between i and any other node j:

$$\frac{(n-1)}{\sum_{j\neq i} l(i,j)}$$

Let $P_i(kj)$ denote the number of geodesics between k and j that i lies on, and let $P(kj)$ be the total number of geodesics between k and j. We can estimate how important i is in terms of connecting k and j by looking at the ration

$$\frac{P_i(kj)}{P(kj)}$$

If this ratio is close to 1, then i lies on most of the shortest paths connecting k to j, while if it is close to 0, then i is less critical to k and j. Averaging across all pairs of nodes, the *betweenness centrality* of a node i is

$$Ce_i^B(g) = \sum_{k\neq j; i\in\{k,j\}} \frac{P_i(kj)/P(kj)}{(n-1)(n-2)/2}$$

Prestige-, Power-, and Eigenvector-Related Centrality Measures. Bonacich's measure is based on the premise that a node's importance is determined by how important its neighbors are. That is, we might like to account not only for the connectivity or closeness of a node to many other nodes, but also for its proximity to many other "important" nodes. This notion is central to such phenomena as citation rankings and Google page rankings. The difficulty is that such a measure becomes self-referential. There are various approaches to dealing with this issue. The following is a nice application of some basic ideas from matrix algebra and fixed-point theory.

3 Review of Linear Algebra and Basic Graph Theory

3.1 Combination and Permutation

For one set has n elements, the number of k-combinations is equal to binomial coefficients:

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

The number of k-permutations is equal to:

$$P(n,k) = \frac{n!}{(n-k)!}$$

3.2 Eigenvectors and Eigenvalues

Eigenvalues occur naturally in many physical and engineering problems. For instance, the natural frequency of the bridge is an example of an eigenvalue. For Google's **PageRank algorithm** there is a special eigenvector, the entries of which can be used to rank search results.

Definition 3.2.1. Let $A \in M_n(R)$ and let x be a non-zero $n \times 1$ column vector. Then X is called an **eigenvector** of A if

$$Ax = \lambda x$$

where λ is some number. In this situation, λ is called an **eigenvalue** of A, and λ and x are said to **correspond** to each other.

Remarks

Notice that eigenvectors are **not unique**. If x is an eigenvector of A, then any non-zero scalar multiple of x is also an eigenvector (corresponding to the same eigenvalue).

Theorem 3.2.2. Let $A \in M_n(R)$. Then λ is an **eigenvalue** of A if and only if $det(A - \lambda I_n) = 0$. And the **characteristic polynomial** of A is the determinant of the matrix $A - \lambda I_n \in M_n(R)$. The equation

$$det(A - \lambda I_n) = 0$$

is called the **characteristic equation** of A.

Sometime, when we look at polynomials having integer coefficients. The following proposition will be very helpful, in particular, can be applied when looking for eigenvalues of matrices and **having integer** entries.

Proposition 3.2.3. Let

$$p(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0$$

be a polynomial having **integer** coefficients. If $p(x)$ has any integer roots, then they must be factors of its constant term a_0.

Example 3.2.4. Find the eigenvalues of

$$A = \begin{pmatrix} 5 & 6 & 2 \\ 0 & -1 & -8 \\ 1 & 0 & -2 \end{pmatrix}$$

We know now how to find eigenvalues. How do we find the corresponding eigenvectors? From $Ax = \lambda x$, we can find the eigenvectors by looking for **non-zero solutions** of

$$(A - \lambda I_n) \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}$$

Example 3.2.5. Find the eigenvectors of following matrix

$$A = \begin{pmatrix} 5 & 6 & 2 \\ 0 & -1 & -8 \\ 1 & 0 & -2 \end{pmatrix}$$

Example 3.2.6. Find the **unit** eigenvectors of following matrix

$$A = \begin{pmatrix} 8 & -3 & -3 \\ -3 & 8 & -3 \\ -3 & -3 & 8 \end{pmatrix}$$

The matrix A in example 7.2.2 is symmetric. As it turns out, the eigenvalues and eigenvectors of symmetric matrices happen to be very well behaved. This is one of the reasons why symmetric matrices are so useful.

Theorem 3.2.7. Let $A \in M_n(R)$ be symmetric. Then the following statements hold.

1. Every **eigenvalues** of A is a real number; we can list them in descending order as $\lambda_1, \cdots, \lambda_n$, so $\lambda_n \geq \cdots \geq \lambda_n$ (note that they need not necessarily be distinct).
2. Let λ_i and λ_j be **distinct eigenvalues** of A, and let x and y be two eigenvectors of A that correspond to λ_i and λ_j, respectively. Then $x \cdot y = 0$.
3. Furthermore, there exists an **orthogonal matrix** $P \in M_n(R)$, such that
   - $Pe_1, \cdots, Pe_n$ is an basis of R^n;
   - each Pe_i is an eigenvector of A having eigenvalue λ_i, and
   - the matrix product

$$P^T AP = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & 0 & \lambda_n \end{pmatrix}$$

   is a **diagonal matrix**, where the eigenvalues of listed above run along the main diagonal.

Quadratic forms and matrices. Quadratic forms are special types of function that map vector x in R^n to number in R. They are closely associated with symmetric matrices. Quadratic forms have applications in calculus, number theory, geometry and statistics. They even managed to find their way into the theory behind wireless communications.

Definition 3.2.8. A **quadratic form** is a function $q : \mathbb{R}^n \to \mathbb{R}$ of the form

$$q(x) = \sum_{i=1}^{n} a_i x_i^2 + \sum_{1 \le i < j \le n} b_{ij} x_i x_j$$

where $x = (x_1, \cdots, x_n)$ and where a_i, $1 \le i \le n$, and b_{ij}, $1 \le i < j \le n$.

Theorem 3.2.9. Let $q : R^n \to R$ be a **quadratic form**. Then there exists a unique symmetric matrix $A \in M_n(R)$, such that

$$q(x) = x \cdot (Ax) \text{ or } q(x) = x'Ax \text{ or } x^T Ax$$

conversely, given a symmetric matrix $A \in M_n(R)$, the function q defined by the expression above is a quadratic form.

Diagonalization. In many cases, the eigenvalue-eigenvector information contained within a matrix A can be displayed in a useful factorization of the form $A = PDP^{-1}$ where D is a diagonal matrix.

Theorem 3.2.10. An $n \times n$ matrix A is diagonalizable if and only if A has n linearly independent eigenvectors.

In other words, A is diagonalizable if and only if there are enough eigenvectors to form a basis of $\mathbb{R}^n$. We call such a basis an eigenvectors basis of $\mathbb{R}^n$.

Example 3.2.11. Diagonalize the following matrix, if possible

$$A = \begin{pmatrix} 1 & 3 & 3 \\ -3 & -5 & -3 \\ 3 & 3 & 1 \end{pmatrix}$$

That is, find an invertible matrix P and a diagonal matrix D such that $A = PDP^{-1}$.

Solution: There are four steps to diagonalize the above matrix.

1. **Find the eigenvalues of** A, based on the characteristic equation

$$\begin{aligned} &det(A - \lambda I_3) = 0 \\ &det \begin{pmatrix} 1-\lambda & 3 & 3 \\ -3 & -5-\lambda & -3 \\ 3 & 3 & 1-\lambda \end{pmatrix} = 0 \\ &(1-\lambda)[(-5-\lambda)+9] - 3[-3(1-\lambda)+9] + 3[-9+3(5+\lambda)] = 0 \\ &-\lambda^3 - 3\lambda^2 + 4 = 0 \\ &-(\lambda-1)(\lambda+2)^2 = 0 \end{aligned}$$

The eigenvalues are $\lambda = 1$ and $\lambda = -2$.

2. **Find three linearly independent eigenvectors of** A.

   For $\lambda = 1$

$$(A - I_3) = \begin{pmatrix} 0 & 3 & 3 \\ -3 & -6 & -3 \\ 3 & 3 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = 0$$

   Find the reduced row-echelon form

$$\begin{pmatrix} 1 & 1 & 0 \\ 0 & -1 & -1 \\ 0 & 0 & 0 \end{pmatrix} \Rightarrow \begin{pmatrix} x_3 \\ -x_3 \\ x_3 \end{pmatrix} \Rightarrow \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}$$

   as the third variable is free.

   For $\lambda = -2$

$$(A + 2I_3) = \begin{pmatrix} 3 & 3 & 3 \\ -3 & -3 & -3 \\ 3 & 3 & 3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = 0$$

   Find the reduced row-echelon form

$$\begin{pmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \Rightarrow \begin{pmatrix} -(x_2 + x_3) \\ x_2 \\ x_3 \end{pmatrix} \Rightarrow \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} \quad or \quad \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}$$

   You can check that $[v_1, v_2, v_3]$ is linearly independent set.

3. **Construct** P **from the vectors in step 2**. The order of the vectors is unimportant.

$$P = [v_1, v_2, v_3] = \begin{pmatrix} 1 & -1 & -1 \\ -1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}$$

4. **Construct** D **from the corresponding eigenvalues**. In this step, it is essential that the **order** of the eigenvalues matches the order chosen for the column of P.

$$D = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & -2 \end{pmatrix}$$

In the end we should have $A = PDP^{-1}$, when P is inverible:

$$A = \begin{pmatrix} 1 & -1 & -1 \\ -1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & -2 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 1 \\ -1 & -1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 3 & 3 \\ -3 & -5 & -3 \\ 3 & 3 & 1 \end{pmatrix}$$

When the matrix is symmetric, the diagonalization becomes more interesting.

Example 3.2.12. If possible, diagonalize the matrix

$$A = \begin{pmatrix} 6 & -2 & -1 \\ -2 & 6 & -1 \\ -1 & -1 & 5 \end{pmatrix}$$

Solution: The characteristic equation of A is

$$\begin{aligned} -\lambda^3 + 17^2 - 90\lambda + 144 = 0 \\ -(\lambda - 8)(\lambda - 6)(\lambda - 3) = 0 \\ \lambda_1 = 8 \quad \lambda_2 = 6 \quad \lambda_3 = 3 \end{aligned}$$

For the above eigenvalues, we can find the corresponding eigenvectors. For $\lambda_1 = 8$,

$$\begin{pmatrix} -2 & -2 & -1 \\ -2 & -2 & -1 \\ -1 & -1 & -3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = 0$$

Find the reduced row-echelon form

$$\begin{pmatrix} -1 & -1 & -3 \\ 0 & 0 & 5 \\ 0 & 0 & 0 \end{pmatrix} \quad \Rightarrow \quad x_3 = 0;\ x_1 = -x_2 \quad \Rightarrow \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}$$

In sum, we have

$$\lambda = 8:\ v_1 = \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix};\quad \lambda = 6:\ v_2 = \begin{pmatrix} -1 \\ -1 \\ 2 \end{pmatrix};\quad \lambda = 3:\ v_3 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

These three vectors from a basis for $\mathbb{R}^3$. In fact, it is easy to check that $[v_1, v_2, v_3]$ is an orthogonal basis for $\mathbb{R}^3$. Now, if we normalize the eigenvectors

$$u_1 = \begin{pmatrix} \frac{-1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \end{pmatrix}, u_2 = \begin{pmatrix} \frac{-1}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} \end{pmatrix}, u_3 = \begin{pmatrix} \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \end{pmatrix}; P = \begin{pmatrix} -1/\sqrt{2} & -1/\sqrt{6} & 1/\sqrt{3} \\ 1/\sqrt{2} & 1/\sqrt{6} & 1/\sqrt{3} \\ 0 & 2/\sqrt{6} & 1/\sqrt{3} \end{pmatrix};\ D = \begin{pmatrix} 8 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 3 \end{pmatrix}$$

Then $A = PDP^{-1}$, as usual. But this time, since P is square and has orthonormal columns, P is an **orthogonal** matrix. and P^{-1} is simply P^T.

Diagonalization of symmetric matrices

Theorem 3.2.13. An $m \times n$ matrix A has orthonormal columns if and only if $U^T U = I$.

Theorem 3.2.14. If A is symmetric, then any two eigenvectors from **different** eigenspaces are orthogonal.

Theorem 3.2.15. An $n \times n$ matrix A is orthogonally diagonalizable if and only if A is a symmetric.

You can find the proof of above theorems from the book - *Linear Algebra and It's Applications* by Lay (2012).

Remark: be careful on the diagonalization of the matrix:

- A is diagonalizable and $A = PDP^{-1}$ if and only if A has n linearly independent eigenvectors.
- A is diagonalizable and $A = PDP^{-1} = PDP^T$ if and only if A is symmetric matrix.
- Never diagonalize a matrix $A = P^{-1}DP$!

The Spectral Theorem. The set of eigenvalues of a matrix A is sometimes called the **spectrum** of A, and the following description of the eigenvalues is called a **spectral theorem**

The spectral Theorem for Symmetric Matrices

An $n \times n$ symmetric matrix A has the following properties:

1. A has n real eigenvalues, counting multiplies.
2. The dimension of the eigenspaces for each eigenvalue λ equals the multiplicity of λ as a root of the characteristic equation.
3. The eigenvalues are mutually orthogonal, in the sense that the eigenvectors corresponding to different eigenvalues are orthogonal.
4. A is orthogonally diagonalizable.

Suppose $A = PDP^{-1}$, where the column of P are orthonormal eigenvectors $u_1, \cdots, u_n$ of A and the corresponding eigenvalues $\lambda_1, \cdots, \lambda_n$ are in the diagonal matrix D. Then, since $P^{-1} = P^T$,

$$A = PDP^T = [u_1, \ \cdots \ u_n] \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix} \begin{bmatrix} u_1^T \\ \vdots \\ u_n^T \end{bmatrix}$$

$$= [\lambda_1 u_1 \ \cdots \ \lambda_n u_n] \begin{bmatrix} u_1^T \\ \vdots \\ u_n^T \end{bmatrix}$$

Using the column-row expansion of a product, we can write

$$A = \lambda_1 u_1 u_1^T + \lambda_2 u_2 u_2^T + \cdots + \lambda_n u_n u_n^T$$

This representation of A is called a spectral decomposition of A because it breaks up A into pieces of determined by the spectrum (eigenvalues) of A. Make sure you can understand this, it is very important to build up the higher level of understanding for time series analysis.

Example 3.2.16. Construct a spectral decomposition of the matrix A that has the orthogonal diagonalization

$$A = \begin{pmatrix} 7 & 2 \\ 2 & 4 \end{pmatrix} = \begin{pmatrix} 2/\sqrt{5} & -1/\sqrt{5} \\ 1/\sqrt{5} & 2/\sqrt{5} \end{pmatrix} \begin{pmatrix} 8 & 0 \\ 0 & 3 \end{pmatrix} \begin{pmatrix} 2/\sqrt{5} & 1/\sqrt{5} \\ -1/\sqrt{5} & 2/\sqrt{5} \end{pmatrix}$$

Solution: Denote the columns of P by u_1 and u_2. Then

$$A = 8u_1u_1^T + 3u_2u_2^T = \begin{pmatrix} 32/5 & 16/5 \\ 16/5 & 8/5 \end{pmatrix} + \begin{pmatrix} 3/5 & -6/5 \\ -6/5 & 12/5 \end{pmatrix} = \begin{pmatrix} 7 & 2 \\ 2 & 4 \end{pmatrix}$$

Constrained Optimization. Engineers, economists, scientists, and mathematicians often need to find the maximum or minimum value of a quadratic from $Q(x)$ for x in some specified set. Typically, the problem can be arranged so that x variables over the set of unit vectors. This *constrained optimization problem* has an interesting and elegant solution.

Theorem for constrained optimization

Theorem 3.2.17. Let $q : R^n \to R$ be a quadratic form, and let $\lambda_1 \geq \cdots \geq \lambda_n$ be the ordered eigenvalues of related symmetric matrix A. The maximum and minimum values that $q(x)$ can take, subject to the constraint $||x|| = 1$, equal λ_1 and λ_n, respectively.

The reason of the above theorem is true because we can change the variable through the diagonalization of the symmetric matrix. If x represents a variable vector in R^n, then the **change of variable** is an equation of the form

$$x = Py$$

If the change of variable is made in q quadratic form x^TAx, then

$$x^TAx = (Py)^TA(Py) = yP^TAPy$$

From the diagonalization part, we have

$$\begin{aligned} A = PDP^{-1} &\Leftrightarrow A \text{ has n linearly independent eigenvectors} \\ A = PDP^{-1} = PDP^T &\Leftrightarrow A \text{ is symmetric matrix} \\ \Rightarrow\ P^TAP = P^TPDP^TP = D & \end{aligned}$$

Now, let use an example to go through all those concepts.

Example 3.2.18. Consider the quadratic from $q : R^3 \to R$, given by

$$q(x) = 4x_1^2 + 2x_2^2 + 3x_3^2 - 4x_1x_3 + 4x_2x_4, \quad x = (x_1, x_2, x_3)$$

Determine

1. the maximum and minimum values a and b of $q(x)$, respectively, subject to the condition $||x|| = 1$, and
2. unit vectors u and v in R^3, such that $q(u) = a$ and $q(v) = b$.

Solution: First, we need find the symmetric matrix A to give the quadratic form $q(x)$:

$$\begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix} \begin{bmatrix} 4 & a & b \\ a & 2 & c \\ b & c & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix} \begin{bmatrix} 4x_1 + ax_2 + bx_3 \\ ax_1 + 2x_2 + cx_3 \\ bx_1 + cx_2 + 3x_3 \end{bmatrix}$$

$$= 4x_1^2 + ax_1x_2 + bx_1x_3 + ax_1x_2 + 2x_2^2 + cx_2x_3 + bx_1x_3 + cx_2x_3 + 3x_3^2$$

$$4x_1^2 + 2x_2^2 + 3x_3^2 + 2ax_1x_2 + 2bx_1x_3 + 2cx_2x_3$$

$$\Rightarrow a = 0 \quad b = -2 \quad c = 2$$

$$\Rightarrow A = \begin{pmatrix} 4 & 0 & -2 \\ 0 & 2 & 2 \\ -2 & 2 & 3 \end{pmatrix}$$

Now, let's find the eigenvalues and eigenvectors of this matrix:

$$det(A - \lambda I_3) = det \begin{pmatrix} 4-\lambda & 0 & -2 \\ 0 & 2-\lambda & 2 \\ -2 & 2 & 3-\lambda \end{pmatrix} = -\lambda(\lambda - 3)(\lambda - 6) = 0$$

$$\Rightarrow \text{eigenvalues are: } \lambda_1 = 0, \lambda_2 = 3, \lambda_3 = 6$$

For $\lambda = 0, 3, 6$, we can find the corresponding vectors

$$v_1 = \pm \begin{bmatrix} 1/3 \\ -2/3 \\ 2/3 \end{bmatrix}; \quad v_2 = \pm \begin{bmatrix} 2/3 \\ 2/3 \\ 1/3 \end{bmatrix}; \quad v_3 = \pm \begin{bmatrix} 2/3 \\ -1/3 \\ -2/3 \end{bmatrix}; \quad P = \begin{pmatrix} 1/3 & 2/3 & 2/3 \\ -2/3 & 2/3 & -1/3 \\ 2/3 & 1/3 & -2/3 \end{pmatrix}$$

P is the corresponding orthogonal matrix, and

$$P^T AP = \begin{pmatrix} 6 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

Finally, setting $y = P^T x$, where $x = (x_1, x_2, x_3)$ and $y = (y_1, y_2, y_3)$, yields the diagonalisation

$$q(x) = 6y_1^2 + 3y_2^2$$

where $q(x)$ has maximum and minimum values 24 and 4, respectively, subject to the condition $||x|| = 1$.

Matrix norms. It happens that matrices have norms. In fact, there are many different ways in which one can impose a norm on a matrix. in our brief treatment, we will look at just one.

Definition 3.2.19. Let $A \in M_n(R)$. The **norm** of A, written $||A||$, is defined to be the maximum of $||Ax||$, subject to the constraint that $||x|| = 1$.

At first glance, it may not be obvious that there should be such a maximum number at all. But there, courtesy of a particular quadratic form. If we square $||Ax||$, we obtain

$$||Ax||^2 = (Ax) \cdot (Ax) = x \cdot (A^T Ax)$$

This defines a quadratic form, because $A^T A$ is symmetric, no matter what the original matrix A is up to. Therefore, by the theorem 9.6.1, the maximum value of $||Ax||^2$, subject to the constraint that $||x|| = 1$, is equal to λ, where λ is the **largest eigenvalues of the symmetric matrix** $A^T A$.

Example 3.2.20. Compute $||A||$, where $A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$.

Solution: We see that

$$A^T = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \quad and \quad A^T A = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$$

The eigenvalues for the above matrix is

$$\lambda = \frac{3}{2} \pm \frac{\sqrt{5}}{2}$$

Then the **larger eigenvalue** would be $\frac{3}{2} + \frac{\sqrt{5}}{2}$, giving $||A|| = \sqrt{\frac{3}{2} + \frac{\sqrt{5}}{2}}$.

Permutation Matrix. A permutation matrix is a square matrix obtained from the same size *identity* matrix by a permutation of rows. Such a matrix is always row equivalent to an identity. Every row and every column of a permutation matrix contain exactly one nonzero entry, which is 1. There are two 2×2 permutation matrices. There are six 3×3 permutation matrices. There are $n!$ permutation matrices of size n.

Every permutation matrix is a product of elementary row-interchange matrices. The elementary matrix factors may be chosen to only involve adjacent rows. For instance,

$$\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Since interchanging ith and jth rows of an identity is equivalent to interchanging its ith and jth columns, every *elementary* permutation matrix is symmetric, $P^T = P$. A general permutation matrix is not symmetric.

3.3 Basic Graph Theory

Since the adjacency matrix A of an undirected graph G is real and symmetric, the eigenvalues of A are real. In addition, there exists an orthogonal matrix S such that $S^T S = SS^T = I_n$ and $S^T AS = D$, where D is the diagonal matrix of eigenvalues of A and the columns of S are the corresponding eigenvectors and I_n is the identity matrix.

Perron-Frobenius theorem. The *Perron-Frobenius eigenvalues* $\lambda_{PF}(G)$ is the largest real eigenvalue of A satisfy $|\lambda_i| \leq \lambda_{PF}(G)$ for $i = 1, \cdots, n$ and there exists an associated nongegative eigenvector $V_{PF} \geq 0$ such that

$$AV_{PF} = \lambda_{PF}(G)V_{PF}$$

For a connected graph G the adjacency matrix A has a unique largest real eigenvalue $\lambda_{PF}(G)$ and a positive associated eigenvector $V_{PF} > 0$, called the Perron-Frobenius eigenvalue and eigenvector, respectively.

The number of walks in a graph can be enumerated with powers of the adjacency matrix A. There exists a relation between the number of walks in a graph and its eigenvalues. The number of closed walks of length k from a node i in G to itself is given by $(A^k)_{ii}$ and the total number of closed walks of length k in G is $tr(A^k) = \sum_{i=1}^{n} (A^k)_{ii} = \sum_{i=1}^{n} = \lambda_i^n$. We further have:

- $tr(A) = 0$
- $tr(A^2)$ gives twice the number of links in G, and
- $tr(A^3)$ gives six times the number of triangles in G.

Connectivity and Irreducability. If a graph G is not connected then its adjacency matrix $A(G)$ can be decomposed in blocks, each block correspond to a *connected* component. An $n \times n$ matrix A is said to be a reducible matrix if and only if for some permutation matrix P, the matrix $P^T AP$ is block upper triangular. If a square matrix is not reducible, it is said to be an *irreducible* matrix. If a graph is connected then there exists a path from every node to every node in the graph. The adjacency matrix of a *connected graph* is irreducible and in particular it cannot be decomposed in blocks.

Irreducible matrices can be *primitive* or *cyclic*. This distinction is relevant for several results on the convergence of linear systems. A non-negative matrix A is primitive if $A^k > 0$ for some positive integer $k \leq (n-1)n^n$. This means that, A is primitive if, for some k, there is a walk of length k from every node to every other node. A graph G is said to be primitive if its associated adjacency matrix $A(G)$ is primitive.

Theorem 3.3.1. Let A be a non-negative matrix, then the spectral radius is an eigenvalue, (called λ_{PF}) and all other eigenvalues are smaller or equal in absolute value. λ_{PF} is associated to one or more non-negative eigenvectors and it is bounded from below and above as follows:

$$\min_i \sum_j a_{ij} \leq \lambda_{PF} \leq \max_i \sum_j a_{ij}$$

If, in addition, A is an irreducible matrix, then λ_{PF} has multiplicity 1 and the associated eigenvector is positive. If, in addition, A is a primitive matrix, then λ_{PF} is strictly greater in absolute value than all other eigenvalues.

The Perron-Frobenius theorem implies that the largest eigenvalues of any nonnegative matrix is real valued, and its corresponding eigenvector is nonnegative.

Bipartite Graphs. A *bipartite* network (N, g) is one for which N can be partitioned into two sets A and B such that if a link ij is in g, then one of the nodes comes from A and the other comes from B. Settings with two classes of nodes are often referred to as *matching* settings (and in some cases marriage markets), where one group is referred to as "women" and the other as "men". It has applications to markets, where for instance one of the sets consists of buyers and the other of sellers. One interpretation of a bipartite graph in a matching setting is that it represents the *potential relationships* that might occur.

Theorem 3.3.2. Consider a bipartite graph (N, g) with an associated bipartition of nodes $\{A, B\}$. There exists a matching of a set $S \subset A$ if and only if $|N_{S'}(g)| \geq |S'|$ for all $S' \subset S$.

Mathematically the definition of bipartite graph can be stated as follows:

Definition 3.3.3. $G(U, S, E)$ will be called a bipartite graph if $V(G) = U(G) \cup S(G)$ and $U(G) \cap S(G) = \emptyset$ and for each edge $(uv) \in E(G)$ either $u \in U(G)$, $v \in S(G)$ or $v \in U(G)$, $u \in S(G)$. G will be a complete bipartite graph if $\forall u \in U(G)$ and $\forall v \in S(G)$, $(uv) \in E(G)$.

For a bipartite graph it is interesting to notice that for any vertex $u_i \in U, N(u_i) \subset S$ and $\forall s_j \in S\ N(s_j) \subset U$. Normally a bipartite graph is represented by a $|U(G)| \times |S(G)|$ matrix known as Bi-adjacency matrix B and the content of the matrix is as follows:

$$(B)_{ij} = \begin{cases} 1, & \text{if } (u_i s_j) \in E(G) \\ 0, & \text{otherwise} \end{cases}$$

Below a bi-adjacency matrix of order 6×4 is given and its corresponding bipartite graph is shown in Fig.3.1:

$$M = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

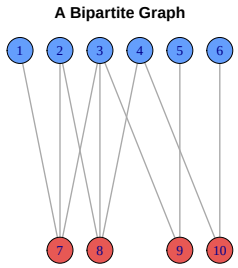


Figure 3.1: Bi-adjacency matrix and corresponding matrix

A bipartite graph can be "projected" into two one-mode networks. Assume that each node denoted with a number represents a project (e.g. a research paper), while each node denoted with a letter represents an individual.

Assume that each node denoted with a number represents a project (e.g. a research paper), while each node denoted with a letter represents an individual. The *one-mode projection* on the individuals is a new graph in which there is a link between two individuals if they work together in one or more projects. In doing this projection some information is lost. For instance, three individuals connected in a triangle, The links do not specify whether each pair of individuals work in a different project or whether the three individuals work all in the same project. This means the projection between the original network and new projected-network is not 'one-to-one' corresponding projection. However, bipartite graph does not have this issue, which means that there is no information loss for one-method projection.

Denote C the incidence matrix of our network of projects and individuals, i.e.

$$C_{ik} = \begin{cases} 1 & \text{if } i\text{works on project } k \\ 0 & \text{otherwise} \end{cases}$$

C is an $n \times m$ matrix, n counting the number of individuals, and m counting the number of projects. This is a binary matrix, and in general it is neither square, nor symmetric. For the one-mode projection relative to the projects, we should take into account that the number of individuals working on projects k and l, is equivalent to the number of paths of length 2 connecting k and l in the bipartite graph.

Therefore, this number can be assigned as the weight of the connection between k and l, and result in a natural way from the following operation on the adjacency matrix.

If we define the adjacency matrix of the project network as

$$B_{kl} = \begin{cases} w_{kl} & \text{if } k \text{ and } l \text{ are connected with weight } w_{kl} \\ 0 & \text{if } k \text{ and } l \text{ are not connected} \end{cases}$$

then it holds that

$$B_{kl} = \sum_{i=1}^{n} C_{ik} C_{il}$$

In terms of a matrix production this can be written as $B = C^T C$. In analogous way, the adjacency matrix of the network of individuals is related to the initial project-individual network as follows

$$D_{ij} = \sum_{k=1}^{m} C_{ik} C_{jk}$$

which is equivalent to $D = CC^T$. While the off-diagonal entries correspond to the edge weights, the diagonal entries, are, respectively, the size B_{kk} of project k (the number of individuals working on it), and the number D_{ii} of projects which individual i works on.

With the same example in Fig.3.1, we will show how the one-mode projection works. Figure 3.2 gives the projection network.

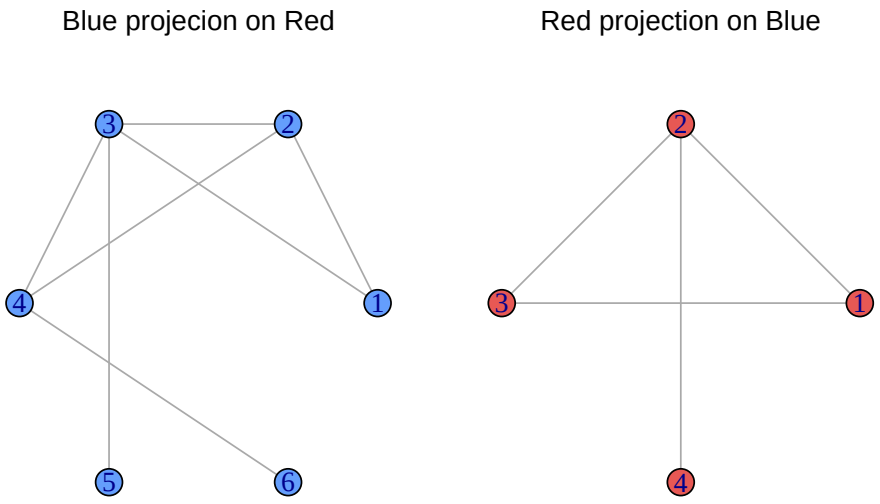


Figure 3.2: Projection of bipartite graph

Now, Let's write down the incidence matrix for project red (we have four sub-projects) and project blue (6 sub-projects):

$$C_r = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \qquad C_b = \begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 \end{pmatrix}$$

Then we can get the adjacency matrix by C_rC_b and C_bC_r:

$$B_{rb} = \begin{pmatrix} 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 2 & 1 & 0 & 0 \\ 1 & 2 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix} rowsum = \begin{bmatrix} 2 \\ 4 \\ 5 \\ 3 \\ 1 \\ 1 \end{bmatrix} \quad B_{br} = \begin{pmatrix} 0 & 2 & 1 & 0 \\ 2 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} rowsum = \begin{bmatrix} 3 \\ 4 \\ 2 \\ 1 \end{bmatrix}$$

The above matrix tells us who works together and how many projects they work together for blue team. For example, the group member 1 works with member 2 and 3, and she works with each of them on one project. For group member 3, she works with member 1 on project red 7 and she works with member 2 on two projects red 7 and 8. It is amazing that how this matrix can tell us the situation of corporation (or partnership). The sum of row shows the number of project participation for each member.

Nested Split Graphs. Consider a nested split graph (NSG) $G = (N, g)$ and let $D = (D_0, D_1, \cdots, D_k)$ be its degree partition. Then the nodes N can be partitioned in *independent sets* D_i, $i = 1, \cdots, \lfloor \frac{k}{2} \rfloor$ and a *dominating set* $\cup_{i=\lfloor \frac{k}{2} \rfloor + 1}^{k} D_i$ in the graph $G' = (N \backslash D_0, g)$.

Comment: for bipartite graph, we need binary indices to split the whole group into two sub-group; or we can treat row and column of bi-adjacency matrix as just two dimension depends on the interest of research.

4 Network Measurements

In this section, we will try to measure different graphs. We will also use one example from `igraph` package and one example from my own research to illustrate every measurement.

The *average path length* l is the mean geodesic (i.e. shortest) distance between node pairs in a graph G

$$l = \frac{1}{\frac{1}{n(n-1)}} \sum_{i \geq j}^{n} d_{ij}(G)$$

where $d_{ij}(G)$ is the geodesic distance from node i to node j in G. The average path length has been analyzed for example in the context of information exchange in networks. The average path length is a macro-indicator of network. Different types of network have different patterns of average path. Now, we will give several examples and present their average path length.

We are now familiar with undirected and directed graph, we will introduce more types of graph. A graph having no edges is called a *Null* graph. A graph with only one vertex is called a *Trivial* graph. A graph with no loops and no parallel edges is called a *simple graph*. A *connected* graph is said to be connected if there exists a path between every pair of vertices. A *regular* graph is the one which all its vertices have the same degree. A simple graph with n mutual vertices is called a *complete* graph and in the graph, a vertex should have edges with all other vertices. A simple graph with n vertices and n edges is called a *cycle* graph if all its edges form a cycle of length n. Figure 4.1 gives the examples of network with 6 nodes.

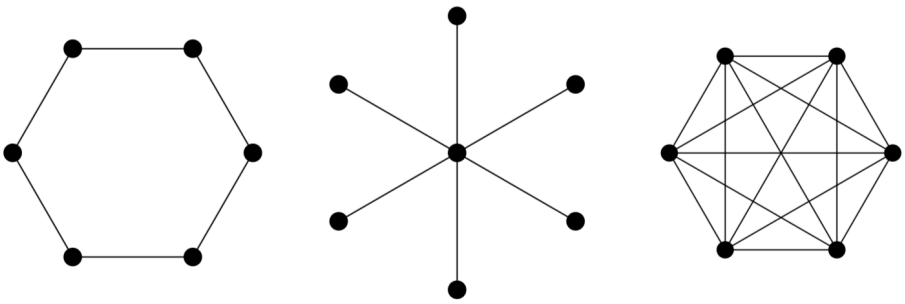


Figure: Illustration of a cycle C_6, a star $K_{1,6}$ and a complete graph, K_6.

Figure 4.1: Different types of graphs

For each node i, the *local clustering coefficient*, $C_l(i)$, is simply defined as the fraction of pairs of neighbors of i that are themselves neighbors. The number of possible links between the neighbors of node i is simple $d_i(d_i - 1)/2$, where d_i denotes the degree of agent i in G:

$$C_l(i) = \frac{|jk \in g(g) : ij \in g(G) \cap ik \in g(G)}{d_i(d_i - 1)/2}$$

The *global clustering coefficient* C_g is then given by

$$C_l = \frac{1}{n} \sum_{i=1}^{n} C_l(i)$$

A high clustering coefficient C_l mean (in the language of social networks), that two of your friends are likely to be also friends of each other. A high clustering coefficient also indicates a high redundancy of the network.

There exists a measure of degree correlation called *average nearest neighbor connectivity*. More precisely, the average nearest neighbor connectivity $d_{nn}(u)$ is the average degree of the neighbors of a node with degree d_u. It is defined by

$$d_{nn}(u) = \frac{1}{d_u} \sum_{v \in N_u} d_v$$

When the average nearest neighbor connectivity is a monotonic increasing function of the degree d, then the network is *assortative*, while, if it is monotonic degreasing with d, it is *dissortative*.

The *degree centrality* of node i is just the number of links, i.e. the degree d_i. We have that $d_i = \sum_{j=1}^{n} a_{ij} = \sum_{j=1}^{n} a_{ji}$ when A is symmetric. If we consider the degree of an agent as a measure of centrality then her centrality depends on the size of the network (with maximum centrality given by $n - 1$). In order to overcome this *bias* one can consider the normalized degree centrality that divides the degree by $n - 1$, yielding a measure in $[0, 1]$. There are several applications of degree centrality, for example the popularity in friendship networks, the diffusion of information and the spread of infections.

The *closeness* $C_C(i)$ of i is the reciprocal of the sum of geodesic distances to all other nodes in the graph, that is

$$C_C(i) = \sum_{v \neq i} \frac{1}{d_{iv}}$$

If an agent has high closeness centrality she can quickly interact with *all* other agents and gather information from them since she has short communication paths to the others.

The *betweenness centrality* of agent i, denoted by $C_B(i)$ is defined as follows

$$C_B(i) = \sum_{u,v \neq i} \frac{g_{uv}(i)}{g_{uv}}$$

More precisely, if g_{uv} is the number of geodesic paths d_{uv} from u to v and $g_{uv}(i)$ is the number of paths from u to v that pass through i, then $\frac{g_{uv}(i)}{g_{uv}}$ is the fraction of geodesic paths from u to v that pass through i. Normalized betweenness divides simple betweenness by its maximum value. If an agent lies on many such path connecting different components in a network then she has a high betweenness centrality.

Eigenvector Centrality. Let's assume that the importance of a node i is measured by x_i. Then the *eigenvector centrality* of all nodes which which are connected to i:

$$x_i = \frac{1}{\lambda} \sum_{j \in N_i} x_j = \frac{1}{\lambda} \sum_{j=1}^{n} a_{ij} x_j$$

where N_i is the set of nodes that are connected to node i, n is the total number of nodes and λ is a constant. In matrix-vector notation we can write

$$Ax = \lambda x$$

which is the eigenvector equation.

If the proportionality factor λ is given by the largest eigenvalue of λ_{PF} of the adjacency matrix A then all the elements in the eigenvector must be positive (by the Perron-Frobenium theorem) and we get a proper measure of centrality. Even if a node is only connected to a few others (low degree) its neighbors may be important, and therefore the node is important too, giving it a high eigenvector centrality.

Now, we introduce a network measure capturing the centrality of an agent in the network due to (Katz, 1953) and later extended by (Bonacich, 1987). Let A be the symmetric $n \times n$ adjacency matrix of the network G and λ_{PF} its largest real eigenvalue. The matrix $M(G, \phi) = (I - \phi A)^{-1}$ exists and is non-negative if and only if $\phi < 1/\lambda_{PF}$. Then:

$$M(G, \phi) = \sum_{k=0}^{\infty} \phi^k A^k$$

The *Katz-Bonacich centrality* vector is given by

$$b_u(G, \phi) = M(G, \phi) \cdot u,$$

where $u = (1, \cdots, 1)^T$. We can write the Katz-Bonacich centrality vector as

$$b_u(G, \phi) = \sum_{k=0}^{\infty} \phi^k A^k \cdot u = (I - \phi A)^{-1} \cdot u$$

Page rank. *page rank* x_i of node i satisfies the following equation:

$$x_i = 1 + \alpha \sum_{k=1}^{n} a_{ik} \frac{1}{d_k^+} x_k$$

Denoting by D the diagonal matrix with diagonal entries given by the out-degrees of the nodes (in a directed network G), we can write

$$x = u + \alpha A D^{-1} x$$

with the solution:

$$x = (I_n - \alpha A D^{-1})^{-1} u = D(D - \alpha A)^{-1} u$$

4.1 Cases studies

In this section, we will just use two dataset `UKFaculty` and `Cologne Gymnasium Students (KGP)` to explore the network measurements.

5 Random Graphs and Network Formation

When it comes to the formation network, there are two broad kind of models: static and dynamic models.

5.1 Static Model (Random Networks)

The term *static* refers to a typed model in which all nodes are established at the same time and then links are drawn between them according to some probabilistic rule.

Consider a set of nodes $N = \{1, \cdots, n\}$, and let a link between any two nodes, i and j, be formed with probability p, where $0 < p < 1$. The formation of links is independent. This is a binomial model of link formation, which gives rise to a manageable set of calculations regarding the resulting network structure. For instance, if $n = 3$, then a complete network forms with probability p^3, any given network with two links (there are three such networks) forms with probability $p^2(1-p)$, any given network with one link forms with probability $p(1-p)^2$.

For n nodes, the all possible combination (or links) are:

$$\binom{n}{2} = \frac{n!}{2!(n-2)!} = \frac{n(n-1)}{2}$$

Then, any given network that has m links on n nodes has a probability of

$$p^m (1-p)^{\frac{n(n-1)}{2} - m}$$

of forming under this process. There is a distinction between the probability of some specific network forming and some network architecture forming. With four nodes the chance that a network forms with a link between nodes 1 and 2 and one between nodes 2 and 3 is

$$p^2(1-p)^4$$

However, the chance that a network forms that contains two links involving three nodes is

$$\binom{4}{3} \cdot \binom{3}{2} p^2 (1-p)^4 = 12 p^2 (1-p)^4$$

The probability that any given node i has exactly d links is

$$\binom{n-1}{d} p^d (1-p)^{n-1-d}$$

5.1.1 Number of Links

The probability that a random network has exactly L links is the product of three terms:

- The probability that L of the attempts to connect the $N(N-1)/2$ pairs of nodes have resulted in a link, which is p^L.
- The probability that the remaining $N(N-2)/2 - L$ attempts have not resulted in a link, which is $(1-p)^{N(N-1)/2-L}$.
- A combination factor

$$\binom{\frac{N(N-1)}{2}}{L}$$

We can therefore write the probability that a particular realization of a random network has exactly L links as

$$p_L = \binom{\frac{N(N-1)}{2}}{L} p^L (1-p)^{N(N-1)/2-L} \tag{5.1}$$

We know for the binomial distribution

$$p_x = \binom{N}{x} p^x (1-p)^{N-x}$$

The mean of the distribution is

$$E[x] = \sum_{x=0}^{N} x p_x = Np$$

Its second moment is

$$E[x^2] = p(1-p)N + p^2 N^2$$

Then the expected number of links in a random graph is

$$E[L] = \frac{pN(N-1)}{2} = \frac{NE[k]}{2} \tag{5.2}$$

Using 5.2, we can obtain the average degree of a random network

$$E[k] = \frac{2E[L]}{N} = p(N-1) \tag{5.3}$$

5.1.2 Poisson Distribution

To derive the Poisson form of the degree distribution we start from the exact binomial distribution:

$$p_k = \binom{N-1}{k} p^k (1-p)^{N-1-k} \tag{5.4}$$

We rewrite the first term on the r.h.s as

$$\binom{N-1}{k} = \frac{(N-1)(N-1-1)(N-1-2)\cdots(N-1-k+1)}{k!} \approx \frac{(N-1)^k}{k!}$$

Now, define the mean of the binomial distribution as λ:

$$\lambda = E[k] = \sum_{k=0}^{N-1} k p_k = (N-1)p \quad \Rightarrow \quad p = \frac{\lambda}{N-1}$$

Take the log of last term of (5.4)

$$\ln[(1-p)^{N-1-k}] = (N-1-k)\ln\big(1 - \frac{\lambda}{N-1}\big)$$

Using the series expansion

$$\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \cdots, \quad \textit{it converges } \forall\ |x| \leq 1$$

This can gives us

$$\begin{aligned} &\ln[(1-p)^{N-1-k}] \approx (N-1-k)\frac{-\lambda}{N-1} = -\lambda\big(1 - \frac{k}{N-1}\big) \approx -\lambda \ \textit{when} \ \ N \to \infty \\ \Rightarrow &(1-p)^{N-1-k} = e^{-\lambda} \end{aligned}$$

Now, substitute p and above expression into (5.4), we have the Poisson distribution

$$\begin{aligned} p_k &= \binom{N-1}{k} p^k (1-p)^{N-1-k} = \frac{(N-1)^k}{k!}\big(\frac{\lambda}{N-1}\big)^k e^{-\lambda} \\ &= e^{-\lambda}\frac{\lambda^k}{k!} \\ mean &= E[k] = \lambda \\ variance &= \lambda \end{aligned} \tag{5.5}$$

Equation (5.5) describes the probability of events. An event can occur $0, 1, 2, \cdots$ times in an interval. Suppose we know the average number of events in an interval $\lambda = E[k]$, which is called the rate parameter. The probability of observing k events in an interval is given by the equation (5.5). Poisson distribution is very powerful, for instance if there are twelve cars crossing a bridge per minute on average, the probability of having seventeen or more cars crossing the bridge in a particular minute is around 0.89871, which is calculated by the following R code: `ppois(16, lambda = 12)`. Figure 5.1 gives the Poisson density plots with varies λ values.

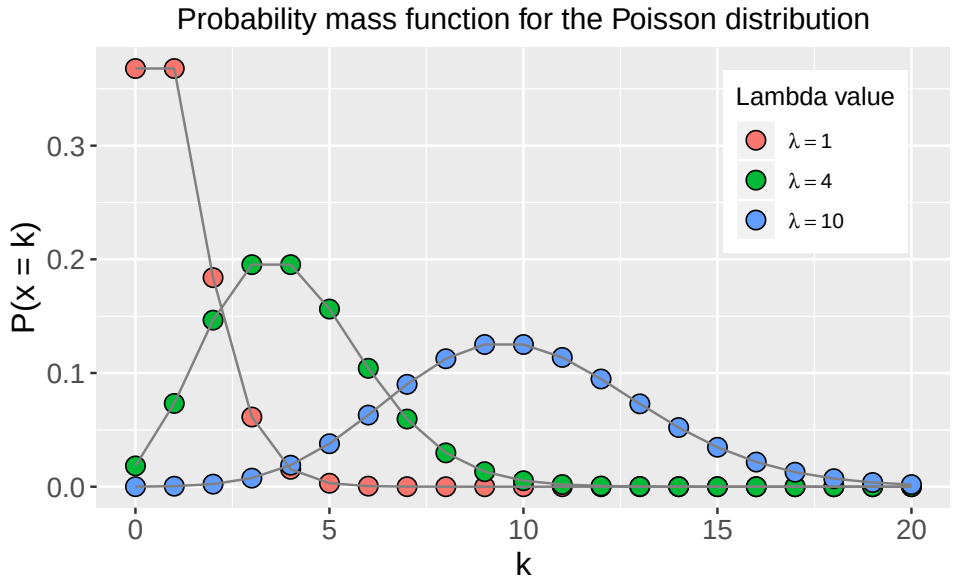


Figure 5.1: Poisson Density plots with varies λ

The binomial and the Poisson distribution describe the same quantity, hence they have similar properties:

- Both distributions have a peak around $E(k)$. If we increase p the network becomes denser.
- The width of the distribution (dispersion) is also controlled by p or $E(k)$. The denser the network, the wider is the distribution, hence the larger are differences in the degree.

Throughout this notes, unless noted otherwise, we will refer to the Poisson form as the degree distribution of a random network. Its key feature is that its properties are independent of the network size and depend on a sing parameter, the average degree λ.

5.1.3 Maximum and Minimum Degrees

To determine the expected degree of the largest node in a random network, called the network's *upper natural cutoff*, we define the degree k_{max} such that in a network of N nodes we have at most one node with degree higher than k_{max}. Mathematically this means that the area behind the Poisson distribution p_k for $k \geq k_{max}$ should be approximately one. Now, define $P(k)$ as the cumulative degree distribution of p_k, the network's largest node satisfies:

$$N[1 - P(k_{max})] \approx 1 \tag{5.6}$$

We use approximate because k_{max} is an integer, so in integer the exact equation does not have a solution. For a Poisson distribution

$$1 - P(k_{max}) = 1 - e^{-\lambda} \sum_{k=0}^{k_{max}} \frac{\lambda^k}{k!} = e^{-\lambda} \sum_{k=k_{max}+1}^{\infty} \frac{\lambda^k}{k!} \approx e^{-\lambda} \frac{\lambda^{k_{max}+1}}{(k_{max}+1)!} \tag{5.7}$$

Combine (5.6) and (5.7), we have

$$N\left[1 - e^{-\lambda}\frac{\lambda^{k_{max}+1}}{(k_{max}+1)!}\right] \approx 1 \tag{5.8}$$

For $N = 10^9$ and $\lambda = 1000$, roughly the size and the average degree of the globe's social network, (5.8) predict $k_{max} = 1185$, indicating that a random network lacks extremely popular individuals, or hubs. We can use the similar trick to get the formula for the expected degree of the smallest node k_{min}:

$$N\left[e^{-\lambda}\sum_{k=0}^{k_{min}-1}\frac{\lambda^k}{k!}\right] \approx 1 \tag{5.9}$$

Solving (5.9) with $N = 10^9$ and $\lambda = 1000$ we obtain $k_{min} = 816$.

Taken together, in a ransom society all individuals are expected to have a comparable number of friends. Hence if people are randomly connected to each other, we lack outliers: there are no highly popular individuals, and no one is left behind as even $k_{min} = 816$. This prediction blatantly conflicts with reality.

5.1.4 Giant Component and Small Worlds

According to Erdős and Rényi (1960), the condition for the emergence of the giant component is

$$\lambda = 1$$

In other words, we have a giant component if and only if each node has one average more than one link. Figure 5.2 gives the evolution of random graph varying with λ.

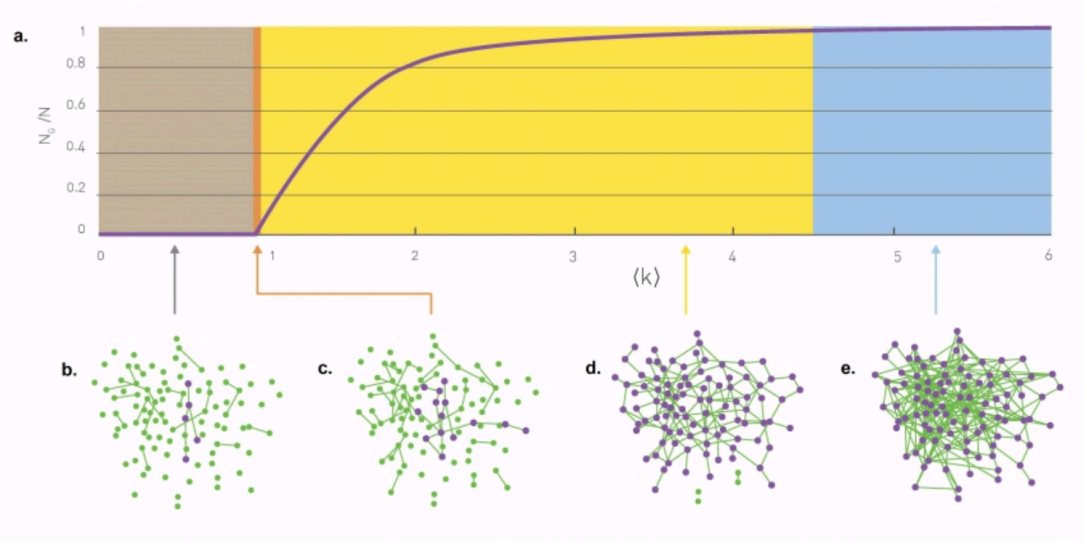


Figure 5.2: Evolution of Random Graph

For sufficiently large p the giant component absorbs all nodes and components. In the absence of isolated nodes the network becomes connected. The average degree at which this happens depends on N as

$$\lambda = \ln N$$

Therefore, for $\lambda > \ln N$ all components are absorbed by the giant component, resulting in a single connected network. Do real networks satisfy the criteria for the existence of a giant component ($\lambda > 1$)? And will this giant component contain all nodes for $\lambda > \ln N$? Table 5.1 gives a snapshot on real world networks. Except for actor network, we didn't see the phenomenon of one single giant component.

Network	**N**	**L**	$\boldsymbol{\lambda}$	$\boldsymbol{\ln N}$
Internet	192,244	609,066	6.34	12.17
Power Grid	4,941	6,594	2.67	8.51
Science Collaboration	23,133	94,437	8.08	10.05
Actor network	702,388	29,397,908	83.71	13.46
Protein Interactions	2.018	2.930	2.90	7.61

Table 5.1: Network Examples from Real Worlds

In summary, we find that most real networks are in the supercritical regime. Therefore these networks are expected to have a giant component, which is in agreement with the observations. Yet, this giant component should coexist with many disconnected components, a prediction that fails for several real networks.

In the language of network science the small world phenomenon implies that the *distance* between two randomly chosen nodes in a network is short. Consider a random network with average degree λ. A node in this network has on average:

- λ nodes at distance one ($d = 1$)
- λ^2 nodes at distance two ($d = 2$)
- λ^3 nodes at distance two ($d = 3$)
- $\cdots$
- λ^d nodes at distance d.

5.2 Scale-free Networks

The poisson form offers a poor fit for the degree distribution of networks in real world based on the analysis of last section. The degree distribution of most real world networks are well approximated with

$$p_k \sim k^{-\gamma} \tag{5.10}$$

Equation (5.10) is called a *power law distribution* and the exponent γ is its *degree exponent*. If we take a logarithm of (5.10), we obtain

$$\log p_k \sim -\gamma \log k \tag{5.11}$$

If (5.10) holds, $\log p_k$ is expected to depend linearly on $\log k$, the slope of this line being the degree exponent γ.

A *scale-free network* is a network whose degree distribution follows a power law. To better understand the scale-free property, we have to define the power-law distribution in more precise terms. Therefore next we discuss the discrete and the continuum formalisms.

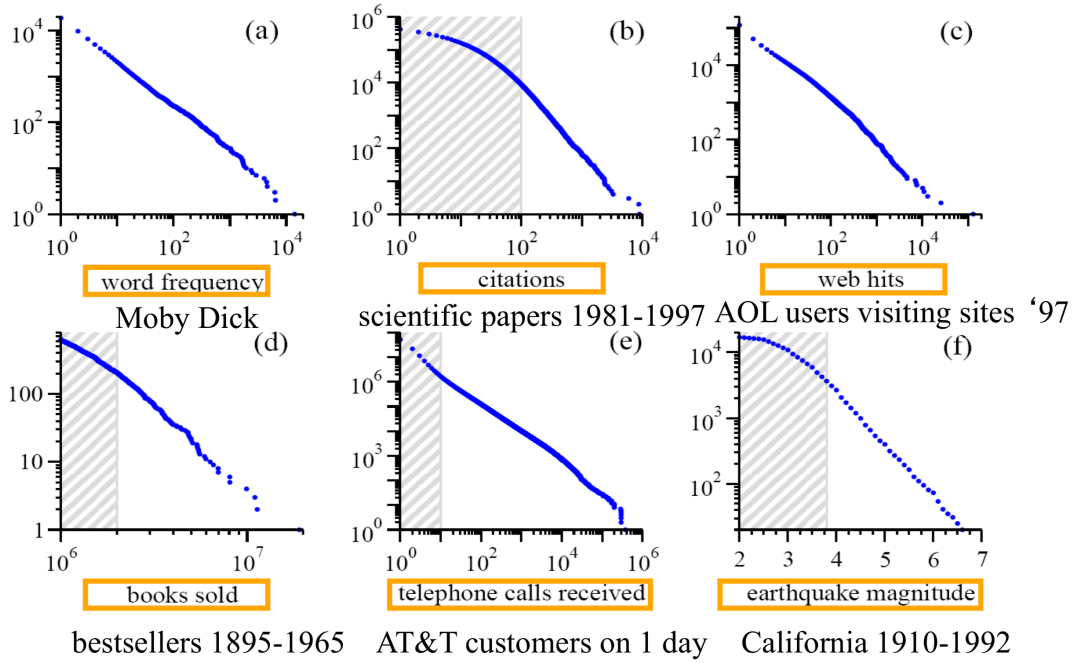


Figure 5.3: Everywhere Power Law

Discrete formalism. As node degrees are positive integers, $k = 0, 1, 2, \cdots$, the discrete formalism provides the probability p_k that a node has exactly k links

$$p_k = Ck^{-\gamma} \tag{5.12}$$

The constant C is determined by the normalization condition.

$$\sum_{k=1}^{\infty} p_k = 1 \quad \Rightarrow \quad C \sum_{k=1}^{\infty} k^{-\gamma} = 1$$

hence

$$C = \frac{1}{\sum_{k=1}^{\infty} k^{-\gamma}} = \frac{1}{\zeta(\gamma)} \tag{5.13}$$

where $\zeta(\gamma)$ is the Riemann-zeta function, then the discrete power-law distribution has the form

$$p_k = \frac{k^{-\gamma}}{\zeta(\gamma)} \tag{5.14}$$

Continuum Formalism. In analytical calculations it is often convenient to assume that the degrees can have any positive real value. In this case we write the power-law degree distribution as

$$p(k) = Ck^{-\gamma} \tag{5.15}$$

Using the normalization condition

$$\int_{k_{min}}^{\infty} p(k) dK = 1 \tag{5.16}$$

we obtain

$$C = \frac{1}{\int_{k_{min}}^{\infty} k^{-\gamma} dk} = (\gamma - 1) k_{min}^{\gamma - 1} \tag{5.17}$$

Therefore in the continuum formalism the degree distribution has the form

$$p(k) = (\gamma - 1) k_{min}^{\gamma - 1} k^{-\gamma} \tag{5.18}$$

Here k_{min} is the smallest degree for which the power law holds.

In summary, networks whose degree distribution follows a power law are called scale-free networks. If a network is directed, the scale-free property applies separately to the in- and the out-degrees. To mathematically study the properties of scale-free networks, we can use either the discrete or the continuum formalism. The scale-free property is independent of the formalism we use.

5.3 Barabasi-Albert Model

The recognition that growth and preferential attachment coexists in real networks has inspired a minimal model called the *Barabasi-Albert* model, which can generate scale-free networks.

We start with m_0 nodes, the links between which are chosen arbitrarily, as long as each node has at least one link. The network develops following two steps:

(a) Growth: at each timestep we add a new node with m ($\leq m_0$) links that connect the new node to m modes already in the network.

(b) The probability $\Pi(k)$ that a link of the new node connects to node i depends on the degree k_i as

$$\Pi(k_i) = \frac{k_i}{\sum_j k_j} \tag{5.19}$$

Preferential attachment is a probabilistic mechanism: A new node is free to connect to any node in the network, whether it is a hub or has a single link. Equation (5.19) implies, however, that if a new nodes has a choice between a degree-two and a degree-four node, it is twice as likely that it connects to the degree-four node.

5.3.1 Degree Dynamics

To understand the emergence of the scale-free property, we need to focus on the time evolution of the Barabasi-Albert model. Let us approximate the degree k_i with a continuous real variable, representing its expectation value over many realization of the growth process. The rate at which an existing node i acquires links as a result of new nodes connecting to it is

$$\frac{dk_i}{dt} = m\Pi(k_i) = m\frac{k_i}{\sum_{j=1}^{N-1} k_j} \tag{5.20}$$

The coefficient m describes that each new node arrives with m links. Hence, node i has m chances to be chosen. The sum in the denominator of (5.20) goes over all nodes in the

network except the newly added node, thus

$$\sum_{j=1}^{N-1} k_j = 2mt - m \tag{5.21}$$

Therefore, (5.20) becomes

$$\frac{dk_i}{dt} = \frac{k_i}{2t-1} \tag{5.22}$$

For large t the (-1) term can be neglected in the denominator, obtaining

$$\frac{dk_i}{k_i} = \frac{1}{2}\frac{dt}{t} \tag{5.23}$$

By integrating (5.23) and using the fact that $k_i(t_i) = m$, meaning that node i joins the network at time t_i with m links, we obtain

$$k_i(t) = m\big(\frac{t}{t_i}\big)^{\beta} \tag{5.24}$$

We call β the *dynamic exponent* and has the value

$$\beta = \frac{1}{2}$$

Equation (5.24) offers a number of predictions:

- The degree of each node increases following a power-law with the same dynamical exponent $\beta = 1/2$. Hence all nodes follow the same dynamic law.
- The growth in the degree is sublinear. This is a consequence of the growing nature of the Barabasi-Albert model.
- The earlier node i was added, the higher is its degree $k_t(t)$. Hence, hubs are large because they arrived earlier, a phenomenon called first-mover advantage.

The coexistence of growth and preferential attachment in the Baraba- si-Albert model raises an important question: Are they both necessary for the emergence of the scale-free property? In other words, could we generate a scale-free network with only one of the two ingredients?

In summary, the absence of preferential attachment leads to a growing network with a stationary but exponential degree distribution. In contrast the absence of growth leads to the loss of stationarity, forcing the network to converge to a complete graph. This failure of Models A and B to reproduce the empirically observed scale-free distribution indicates that growth and preferential attachment are simultaneously needed for the emergence of the scale-free property.

5.3.2 Measuring Preferential Attachment

In this section, we show how to detect preferential attachment by measuring the $\Pi(k)$ function in real networks.

Preference attachment relies on two distinct hypothesis:

- **Hypothesis 1**: The likelihood to connect to a node depends on that node's degree k.
- **Hypothesis 2**: The function form of $\Pi(k)$ is linear in k.

Both hypothesis can be tested by measuring $\Pi(k)$. We can determine $\Pi(k)$ for systems for which we know the time at which each node joined the network, or we have at least two networks maps collected at not too distinct moments in time.

Consider a network for which we have two different maps, the first taken at time t and the second at time $t + \Delta t$. For nodes that changed their degree during the Δt time frame we measure $\Delta k_i = k_t(t + \Delta t) - k_t(t)$. The relative change $\Delta k_t / \Delta t$ should follow

$$\frac{\Delta k_i}{\Delta t} \sim \Pi(k_i) \tag{5.25}$$

In practice the obtained $\Delta k_t / \Delta t$ curve can be noisy. To reduce this noise we measure the *cumulative preferential attachment function*

$$\pi(k) = \sum_{k_i=0}^{k} \Pi(k_i) \tag{5.26}$$

In the absence of preferential attachment we have $Pi(k_i) = constant$, hence, $\pi(k) \sim k$ according to (5.26). If linear preferential attachment is present, i.e. if $\Pi(k_i) = k_i$, we expect $\pi(k) \sim k^2$.

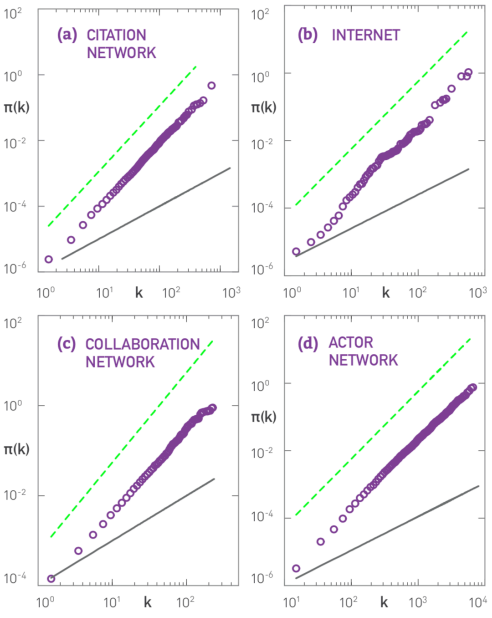


Figure 5.10
Evidence of Preferential Attachment

The figure shows the cumulative preferential attachment function $\pi(k)$, defined in (5.21), for several real systems:

(a) Citation network.

(b) Internet.

(c) Scientific collaboration network (neuroscience).

(d) Actor network.

In each panel we have two lines to guide the eye: The dashed line corresponds to linear preferential attachment ($\pi(k)$-k^2) and the continuous line indicates the absence of preferential attachment ($\pi(k)$-k). In line with Hypothesis 1 we detect a k-dependence in each dataset. Yet, in (c) and (d) $\pi(k)$ grows slower than k^2, indicating that for these systems preferential attachment is sublinear, violating Hypothesis 2. Note that these measurements only consider links added through the arrival of new nodes, ignoring the addition of internal links. After [14].

Figure 5.4: Measuring the cumulative likelihood

Figure 5.4 shows the measured $pi(k)$ for four real networks. For each system we observe a faster than linear increase in $\pi(k)$, indicating the presence of preferential attachment. It also suggests that $\Pi(k)$ can be approximate with

$$\Pi(k) \sim k^{\alpha} \tag{5.27}$$

6 Econometrics of Interactions in Networks

In this section, we will present some econometric models to study the interactions in networks.

Spatial Autoregressive (SAR) Model. Now, consider an $n \times 1$ vector of outcome variables, $y = (y_1, \cdots, y_n)'$. An $n \times n$ adjacency matrix, $A = [a_{ij}]$, with $1 \leq i, j \leq n$. Then, the *Spatial Autoregressive (SAR)* model is defined as

$$y = \lambda Ay + X\beta + \varepsilon \tag{6.1}$$

Ay is the spatial lag, and λ is the spatial autoregressive parameter. X is an $n \times k$ matrix of observations with k exogenous variables. and we assume the error vector $\varepsilon_i | A, X \sim N(0, \sigma^2)$.

Linear Quadratic Utility. Consider a game in which players decide how much time or effort to exert in some activity $y_i \in R_+$. The payoff to player i as a function of the action profile and network, $\pi_i : R_+ \times \mathcal{G}^n$, given by

$$\pi_i(y) = \alpha_i y_i - \frac{1}{2} y_i^2 + \lambda \sum_{j=1}^{n} a_{ij} y_i y_j \tag{6.2}$$

The first two terms $\alpha_i y_i - \frac{1}{2} y_i^2$ give the benefits and cost of providing the action level y_i. The last term $\lambda \sum_{j=1}^{n} a_{ij} y_i y_j$ reflects the strategic complementarity between friends' and acquaintances' actions and own action.

The first-order necessary condition (FOC) for each player i's choice of action to maximize his or her payoff is:

$$\frac{\partial \pi_i(y)}{\partial y_i} = \alpha_i - y_i + \lambda \sum_{j=1}^{n} a_{ij} y_j = 0 \tag{6.3}$$

which leads to:

$$y_i = \alpha_i + \lambda \sum_{j=1}^{n} a_{ij} y_j \tag{6.4}$$

When $\alpha = X\beta + \varepsilon$, then Eq. 6.4 is equivalent to Eq. 6.1.

Now, rewrite Eq. 6.1 as:

$$(I - \lambda A) y = X\beta_{\varepsilon}$$

Suppose, $I_n - \lambda A$ is nonsingular, then

$$y = (I - \lambda A)^{-1} X\beta + (I - \lambda A)^{-1} \varepsilon \tag{6.5}$$

This gives us the **reduced form equation**. The average *direct* marginal effect of x_{ik} on y_i is

$$\begin{aligned} AME_k^d &= \frac{1}{n}\sum_{i=1}^{n}\frac{\partial y_i}{\partial x_{ik}} \\ &= \frac{1}{n}\sum_{i=1}^{n}[(I_n - \lambda A)^{-1}]_{ii}\beta_k \\ &= \frac{1}{n}tr[(I_n - \lambda A)^{-1}]\beta_k \end{aligned}$$

The average *indirect* marginal effect of x_{ik} on y_i is

$$\begin{aligned} AME_k^{id} &= \frac{1}{n}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\frac{\partial y_j}{\partial x_{ik}} \\ &= \frac{1}{n}\sum_{i=1}^{n}\sum_{j\neq i}^{n}[(I_n - \lambda A)^{-1}]_{ji}\beta_k \end{aligned}$$

where

$$AME_k^{total} = \frac{1}{m}u_m^T(I_n - \lambda A)^{-1}u_n\beta_K$$

Hence, λ captures a *network multiplier* effects: if $\lambda = 0$ then $AME_k^d = \beta_k$ and $AME_k^{id} = 0$.

Endogeneity of the Spatial Lag. From the reduced form Eq 6.5, we can have:

$$Ay = A(I_n - \lambda A)^{-1}X\beta + A(I_n - \lambda A)^{-1}\varepsilon \tag{6.6}$$

Since

$$\begin{aligned} E[\varepsilon' Ay] &= E[\varepsilon' A(I_n - \lambda A)^{-1}X\beta] + E[\varepsilon' A(I_n - \lambda A)^{-1}\varepsilon] \\ &= E[\varepsilon' A(I_n - \lambda A)^{-1}\varepsilon] \\ &= \sigma^2 tr[(I_n - \lambda A)^{-1}] \neq 0 \end{aligned}$$

Hence, the spatial lag Ay is endogenous. However, from Eq. 6.6, notice:

$$E[Ay] = A(I_n - \lambda A)^{-1}X\beta$$

If $|\lambda| < 1/||A||$, then:

$$(I_n - \lambda A)^{-1} = I_n + \lambda A + \lambda^2 A^2 + \cdots$$

Therefore,

$$E[Ay] = AX\beta + \lambda A^2 X\beta + \lambda^2 A^3 X\beta + \cdots$$

then, $[AX, A^2X, A^3X, \cdots]$ can be used as IVs for Ay. The intuition is that the characteristics of my friends, my friends' friends, etc., influence their actions and hence me, but do not directly influence me.

Comment: you might feel excited when you reach the above part, thinking that you found the IV to solve the endogenous issue. But, wait a moment, do we? Let's check the assumption:

$$\varepsilon_i | A, X \sim N(0, \sigma^2)$$

The assumption here implies that the error term is *iid* in terms of first order relationship in the network. By first order relationship, all nodes are connected to i by path 1. This means the network forming with more than 2 paths is note defined yet. Although we have shown that $AX, A^2X, \cdots$ can be used as IVs, the *omitted variable* issue or endogenous issue is still not solved. Suppose there exist an unobservable variable z affect the characteristics of me and my friends, my friends' friends, etc., how do we identify the peer effects in network?

Example, Now suppose we have the following network:

$$A = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \qquad A^2 = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 2 & 0 \\ 1 & 0 & 1 \end{bmatrix}$$

Then,

$$Ay = \begin{bmatrix} y_1 \\ y_1 + y_3 \\ y_2 \end{bmatrix}, \qquad AX = \begin{bmatrix} x_2 \\ x_1 + x_3 \\ x_2 \end{bmatrix}, \qquad A^2X = \begin{bmatrix} x_1 + x_3 \\ 2x_2 \\ x_1 + x_3 \end{bmatrix}$$

Let $H = [X, AX, A^2X, \cdots, A^pX]$ be the $n \times h$ IV matrix. Then,

$$E(H'\varepsilon) = 0$$

Now, let $\varepsilon(\delta) = y - Z\delta$ where $Z = [Ay, X]$ and $\delta = (\lambda, \beta')'$. Then the empirical counterpart is:

$$H'\varepsilon(\delta) = 0$$

Now, if $h = 1 + k$, which means we have enough instrumental variables, then $g(\delta) = 0$ has a unique solution given by

$$\hat{\delta}_{IV} = (H'Z)^{-1}H'y$$

If $h > 1 + k$, then $g(\delta) = 0$ may not have a solution. In this case, δ can be estimated by the generalized method of moments (GMM)

$$\hat{\delta} = \text{argmin}_\delta \; g(\delta)' \wedge g(\delta)$$

The optimal choice of the weighting matrix $\wedge$ is the inverse of

$$Var[g(\delta)] = \sigma^2 H'H$$

The optimal GMM estimator

$$\begin{aligned} \hat{\delta}_{2SLS} &= \text{argmin}_\delta g(\delta)'(H'H)^{-1}g(\delta) \\ &= (Z'P_HZ)^{-1}Z'P_Hy \end{aligned}$$

where $P_H = H(H'H)^{-1}H'$, is the two stages least squares estimator.

Maximum Likelihood Estimation (MLE). Suppose $\varepsilon|A, X \sim N(), \sigma^2 I_n)$, then we have

$$y \sim N(\mu_y, \Sigma_y)$$

where

$$\begin{aligned}\mu_y =&(I_n - \lambda A)^{-1} X\beta \\ \Sigma_y =&\sigma^2 (I_n - \lambda A)^{-1}(I_n - \lambda A')^{-1}\end{aligned}$$

The joint density of y is then

$$f(y) = (2\pi)^{-\frac{n}{2}}(\det\Sigma_y)^{-1/2}\exp\Big\{-\frac{1}{2}(y-\mu_y)'\Sigma_y^{-1}(y-\mu_y)\Big\}$$

The log-likelihood is given by

$$\mathcal{L}(\theta) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\sigma^2 + \ln|\det(I_n - \lambda A)| - \frac{1}{2\sigma^2}\varepsilon(\delta)'\varepsilon(\sigma)$$

where

$$\varepsilon(\delta) = y - Z\delta = y - \lambda A y - X\beta$$

The MEL are given by

$$\begin{aligned}\hat{\lambda}_{MLE} &= \text{argmin}\mathcal{L}(\lambda) \\ \hat{\beta}_{MLE} &= (X'X)^{-1}X'(I_n - \hat{\lambda}_{MLE}A)y \\ \hat{\sigma}^2_{MLE} &= \frac{1}{n}y'(I_n - \hat{\lambda}A)'M_X(I_n - \hat{\lambda}A)y\end{aligned}$$

where

$$M_X = I_n - X(X'X)^{-1}X'$$

Table 6.1 gives the summary of comparison of two methods.

Methods	Pros	Cons
IV-based estimator	consistent under spatially correlated errors and heteroskedasticiy; easy to implement	identification issues potential weak-IV problem
MLE	identification, asymptotic efficiency	computationally costly inconsistent under heteroskedasticity

Table 6.1: Comparison of 2SLS and MLE

Identification Issues. Recall the SAR model is given by

$$y = \lambda A y + X\beta + AX\gamma + \varepsilon \tag{6.7}$$

In this model, we have the *endogenous effect*: when the propensity of an individual to behave in some way varies with the behavior of the peer group (λ). We also have the *exogenous contextual effect*: when the propensity of an individual to behave in some way varies with the exogenous characteristics of the peer group (γ). Clearly there is a correlation (sometimes very strong) between endogenous and exogenous effects as individuals in the same group tend to behave similarly with the similar individual characteristics or facing similar institutional environments.

To solve this problem, we need disentangle the process of network formation. Now, assume agents self-select into different networks (components) in a first step, and that link formation takes place within groups in a second step. If a link formation is uncorrelated with the observable variables, this two-step model of link formation generates *network fixed effects*.

Suppose there are r networks (components), with n_s agents in the s-th network, with $n = \sum_{s=1}^{n} n_s$. For the s-th network

$$y_s = \lambda A_s y_s + X_s\beta + A_s X_s\gamma + \alpha_s u_{n_s} + \varepsilon_s$$

where α_s is the network-specific fixed effect. For all the r networks in the data

$$y = \lambda A y + X\beta + AX\gamma + U\alpha + \varepsilon$$

where $y = (y_1', \cdots, y_r')'$, and $X = (X_1', \cdots, X_r')'$, $\alpha = (\alpha_1, \cdots, \alpha_r)'$, $\varepsilon = (\varepsilon_1', \cdots, \varepsilon_r')'$.

$$A = \begin{bmatrix} A_1 & & \\ & \ddots & \\ & & A_r \end{bmatrix}, \qquad U = \begin{bmatrix} u_1 & & \\ & \ddots & \\ & & u_r \end{bmatrix}$$

Now let $J = \text{diag}\{J_s\}_{s=1}^{r}$ where $J_s = I_{n_s} - \frac{1}{n_s} u_{n_s}' u_{n_s}$. As $JU = 0$, the "within" transformation eliminates the network fixed effect, s.t.

$$Jy = \lambda JAy + JX\beta + JAX\gamma + J\varepsilon$$

The reduced-form is

$$y = (I_n - \lambda A)^{-1}(X\beta + AX\gamma) + (I_n - \lambda A)^{-1}U\alpha$$

which implies that

$$\begin{aligned} E(JAy) = {} & JAX\beta \\ & + (JA^2X\lambda JA^3X + \cdots)(\lambda\beta + \gamma) \\ & + (JAU + \lambda JA^2U + \cdots)\alpha \end{aligned}$$

If $\alpha \neq 0$ and A has non-constant row sums, then $[JAU, JA^2U, \cdots]$ can be used as IVs for JAy.

Now, let's check one example with following matrix

$$A = \begin{bmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} \qquad A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

Reflection Problem. In regular networks, simultaneity in the behavior of interacting agents introduces a perfect collinearity between the expected outcome of the group and its characteristics. Therefore, it is difficult to differentiate between the effect of peers' choice and peers' characteristics. This problem is called *reflection problem*. However, in social networks the preference group is individual-specific unless the network is complete or has a very regular structure. Thus, there is no reflection problem.

Network Formation. One can have an explicit model of network formation and estimate the outcome equation and the bilateral choice of links simultaneously (*control function* approach). This approach allows for the presence of unobservable factors that vary by link-type. Consider a network model given by

$$y = \lambda Ay + X\beta + AX\gamma + \alpha u_n + \varepsilon$$

where $A = [a_{ij}]$. Now, considering the following logistic regression model

$$P(a_{ij} = 1) = \frac{\exp(\psi_{ij})}{1 + \exp(\psi_{ij})}$$

where

$$\psi_{ij} = z_{ij}\delta + \phi|u_i - u_j|$$

ψ_{ij} can be interpreted as the difference in the deterministic part of the random utility for i to form a link with j. We also assume that z_{ij} is uncorrelated with $(\varepsilon_i, \varepsilon_j)$. With the estimated formation, we can have the new Ivs:

- Run a logistic regression of a_{ij} on z_{ij} to obtain $\hat{A} = [\hat{a}_{ij}]$.
- Estimate the network model using the IVs

$$[u_n, X, \hat{A}X, \hat{A}^2X]$$

Multiple Spatial Weight Matrices. Now, consider a SAR model with multiple spatial lags in both y and X

$$y = \sum_{r=1}^{p} \lambda_r A_r y + X\beta + \sum_{r=1}^{p} A_r X\gamma_r + \nu$$

where the error term is defined as

$$\mu = \sum_{r=1}^{q} \rho_r W_r \nu + \varepsilon + \sum_{r=q+1}^{p} \rho_r W_r \varepsilon$$

Appendices

We put the mathematical tools in the appendix, so readers can review some fundamental concepts used in graph theory.

References

Bonacich, P. (1987): "Power and centrality: A family of measures," *American journal of sociology*, 92, 1170–1182.

Easley, D., J. Kleinberg, et al. (2010): *Networks, crowds, and markets*, vol. 8, Cambridge University Press Cambridge.

Erdős, P. and A. Rényi (1960): "On the evolution of random graphs," *Publ. Math. Inst. Hung. Acad. Sci*, 5, 17–60.

Jackson, M. O. (2010): *Social and economic networks*, Princeton university press.

Katz, L. (1953): "A new status index derived from sociometric analysis," *Psychometrika*, 18, 39–43.

Kolaczyk, E. D. and G. Csárdi (2014): *Statistical analysis of network data with R*, vol. 65, Springer.

Lay, D. (2012): *Linear Algebra and Its Applications*, Addison-Wesley.