

Statistical Analysis of Networks

Social Data Science MSc, Oxford Internet Institute

Gesine Reinert, Department of Statistics,
reinert@stats.ox.ac.uk

HT 2019

0.1

Preliminaries

There will be 8 lectures, two 2-hour practical sessions, and two 2-hour problem sessions.

The lectures will take place Mondays 9:00-11:00 in weeks 1-4, the practical sessions will take place Wednesdays 9:00-11:00 in weeks 2 and 4, and the problem sessions will take place Wednesdays 9:00-11:00 in weeks 3 and 5. The lectures and problem sessions will take place in the Seminar Room at 1 St Giles; the practicals will take place in the IT suite of the Department of Statistics, 24-29 St Giles.

The practicals will use Python. There will be a practical assistant, Javier Pardo Diaz,
javier.pardodiaz@keble.ox.ac.uk.

There will be homework assignments for the problem sessions which will be marked, and some of the assignments require some computer work. Homework assignments are to be handed in Monday of weeks 3 and 5, by 10 am, in the Department of Statistics, 24-29 St Giles'. Your marker will be Lyuba Bozhilova, lyuba.bozhilova@dtc.ox.ac.uk.
0.2

The format of lectures and examination

The lectures will use data projection and the white board. You may find it useful to take notes from the white board even though the lecture notes are available.

Lecture notes may be slightly adapted or updated throughout the course. If there is an update then you will be notified.

The course will be assessed by a mini-project.

Reading

E.D. Kolaczyk. Statistical Analysis of Network Data. Springer, 2009.

E.D. Kolaczyk and G. Csárdi. Statistical Analysis of Network Data with R. Springer, 2014.

M.E.J. Newman. Networks: An Introduction. Oxford University Press, 2010.

0.3

0.4

Outline

- 1 Introduction
- 2 Models for networks
- 3 Summary of distributional results for network statistics
 - The degree of a randomly chosen vertex
 - The number of triangles
 - Shortest paths
- 4 Sampling from networks
- 5 Fitting a model
- 6 Nonparametric methods
- 7 Statistical inference for edges
- 8 Motifs
- 9 Network Centrality Measures
- 10 Modules and communities
- 11 Further topics

Statistical Analysis of Networks
Social Data Science MSc, Oxford Internet Institute

Gesine Reinert, Department of Statistics,
reinert@stats.ox.ac.uk

HT 2019

0.5

1.1

1 Introduction

Networks are just graphs.

Networks can provide a useful representation of interdependencies in data.

Networks are also used to represent statistical models - so-called graphical models - but these lectures do not address graphical models.

Often one would think of a network as a connected graph, but not always. In this lecture course we shall use *network* and *graph* interchangeably.

Here are some examples of networks (graphs).

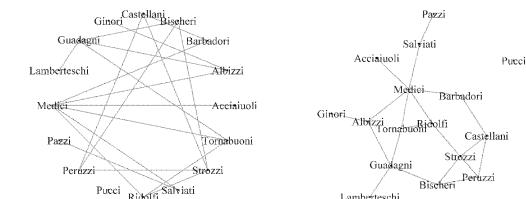


Figure: Marriage relations between Florentine families; two different graphical representations

1.2

The Florentine Families marriage data, collected by Padgett and Ansell (1993), give an undirected network which consists of the marriage ties among 16 families in 15th century Florence, Italy. ^{1.3}

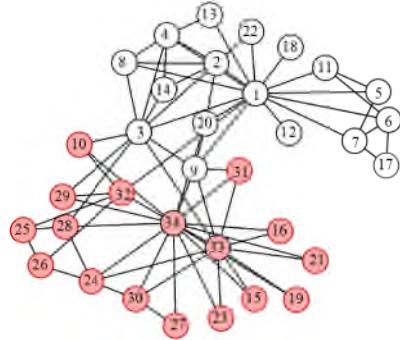


Figure: A friendship network: Zachary's karate club

Zachary's Karate club network (Zachary (1977)) is a social network of friendships between 34 members of a karate club at a US university in the 1970s. The club is known to have split into two different factions as a result of an internal dispute, and the members of each faction are known.

1.4

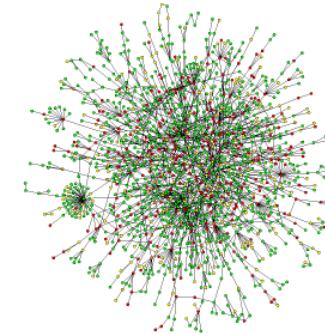


Figure: Yeast protein-protein interactions

In a protein-protein interaction network, vertices are proteins, and edges represent physical interactions. In this network vertices are coloured by lethality.

1.5

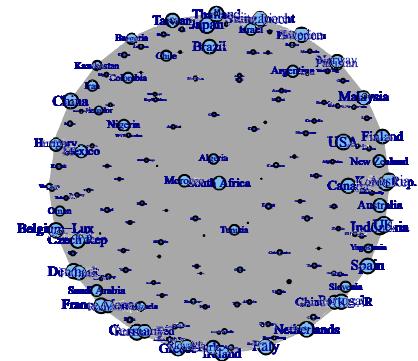


Figure: Trade network 2000

1.6

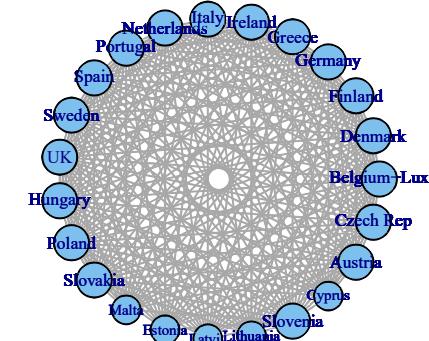


Figure: Trade network 2000: EU countries and those who will join the EU in 2004

1.7

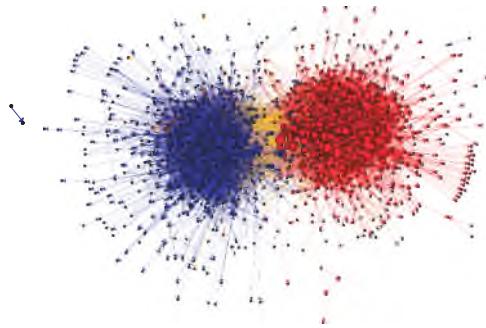


Figure: Political blog data from Adamic and Glance (2005). The colours reflect political orientation, red for conservative, and blue for liberal. Orange links go from liberal to conservative, and purple ones from conservative to liberal. The size of each blog reflects the number of other blogs that link to it.

1.8

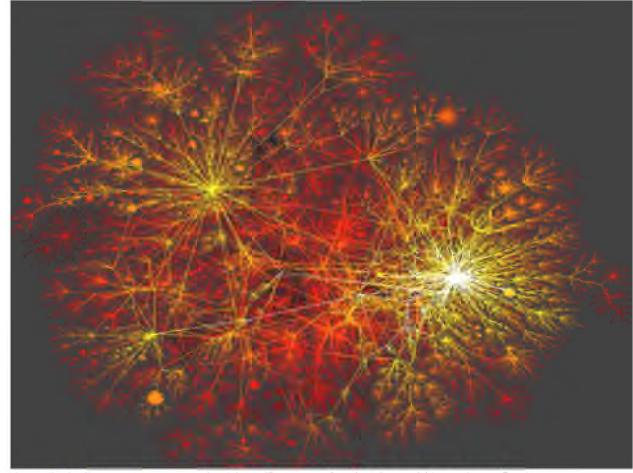


Figure: The internet: vertices are computers and routers, edges represent physical connections between them. From Frieze, Gionis, Tsourakakis, <http://www.math.cmu.edu/~ctsourak/amazing.pdf>

1.9

Networks arise in a multitude of contexts; see for example

- metabolic networks
- spread of epidemics
- brain networks
- collaboration networks (Erdős numbers ...)
- Membership of management boards
- World Wide Web
- power grid of the Western US.

1.10

Unless the network is very small it appears like a hairball, and is difficult to analyse by just looking at it.

The study of networks has a long tradition in social science, where it is called *Social Network Analysis*. The networks under consideration are typically fairly small. In contrast, starting at around 1997, statistical physicists have turned their attention to large-scale properties of networks. Our lectures will try to get a glimpse on both approaches.

1.11

Research questions include

- ▶ How do these networks work? Where could we best manipulate a network in order to prevent, say, tumor growth?
- ▶ How did these networks evolve?
- ▶ How similar are these networks?
- ▶ How are these networks interlinked?
- ▶ What are the building principles of these networks? How is resilience achieved, and how is flexibility achieved?

1.12

From a statistical viewpoint, questions include

- ▶ How to best describe networks?
- ▶ How to infer characteristics of vertices in the network?
- ▶ How to infer missing links, and how to check whether existing links are not false positives?
- ▶ How to compare networks?
- ▶ How to predict functions from networks?
- ▶ How to find relevant sub-structures of a network?

Statistical inference relies on the assumption that there is some randomness in the data.

1.13

1.1. What are networks?

Mathematically, we abbreviate a graph \mathcal{G} as $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where \mathcal{V} is the set of vertices (nodes) and \mathcal{E} is the set of edges (links). We use the notation $|S|$ to denote the number of elements in the set S . Then $|\mathcal{V}|$ is the number of vertices, and $|\mathcal{E}|$ is the number of edges in the graph \mathcal{G} . If u and v are two vertices and there is an edge from u to v , then we write that $(u, v) \in \mathcal{E}$, and we say that v is a *neighbour* of u .

Edges may be *directed* or *undirected*. A *directed graph*, or *digraph*, is a graph where all edges are directed. The *underlying* graph of a digraph is the graph that results from turning all directed edges into undirected edges. Here we shall mainly deal with undirected graphs.

If both endpoints of an edge are the same, then the edge is a *loop*. Here we shall mainly exclude self-loops, as well as multiple edges between two vertices.

1.14

Two vertices are called *adjacent* if they are joined by an edge. A graph can be described by its $|\mathcal{V}| \times |\mathcal{V}|$ *adjacency matrix* $A = (a_{u,v})$;

$$a_{u,v} = 1 \text{ if and only if } (u, v) \in \mathcal{E}.$$

If there are no self-loops, all elements on the diagonal of the adjacency matrix are 0. If the edges of the graph are undirected, then the adjacency matrix will be symmetric.

The adjacency matrix entries tell us for every vertex v which vertices are within (graph) distance 1 of v . If we take the matrix product $A^2 = A \times A$, the entry for (u, v) with $u \neq v$ would be

$$a^{(2)}(u, v) = \sum_{w \in \mathcal{V}} a_{u,w} a_{w,v}.$$

If $a^{(2)}(u, v) \neq 0$ then u can be reached from v within two steps; u is within distance 2 of v . Higher powers can be interpreted similarly.

1.15

Example: Adjacency matrix for Florentine marriages

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}$$

1.16

A *complete* graph is a graph, without self-loops, such that every pair of vertices is joined by an edge. The adjacency matrix has entry 0 on the diagonal, and 1 everywhere else.

A *bipartite* graph is a graph where the vertex set \mathcal{V} is decomposed into two disjoint subsets, \mathcal{U} and \mathcal{W} , say, such that there are no edges between any two vertices in \mathcal{U} , and also there are no edges between any two vertices in \mathcal{W} ; all edges have one endpoint in \mathcal{U} and the other endpoint in \mathcal{W} .

The adjacency matrix A can then be arranged such that it is of the form

$$\begin{bmatrix} 0 & A_1 \\ A_2 & 0 \end{bmatrix}.$$

1.17

1.2. Network summaries

To analyse and to compare networks, often low-dimensional summaries are used.

Some summaries concentrate on local features, such as local clustering, whereas other summaries concentrate on global features.

1.2.1. Degrees and the degree distribution

The *degree* $d(v)$ of a vertex v is the number of edges which involve v as an endpoint. The degree is easily calculated from the adjacency matrix A ;

$$d(v) = \sum_u a_{u,v}.$$

The *average degree* of a graph is the average of its vertex degrees;

$$\bar{d} = \frac{1}{|\mathcal{V}|} \sum_{v \in \mathcal{V}} d(v).$$

1.18

1.19

The *degree sequence* of a given network on a vertex set \mathcal{V} with n elements is the unordered n -element set of degrees $\{d(v), v \in \mathcal{V}\}$. For example, the degree sequence of a triangle on three vertices is $\{2, 2, 2\}$.

The *degree distribution* (d_0, d_1, d_2, \dots) of a graph on n vertices is the vector of fraction of vertices with given degree;

$$d_k = \frac{1}{n} \times \text{number of vertices of degree } k.$$

For directed graphs we define the *in-degree* as the number of edges directed at the vertex, and the *out-degree* as the number of edges that go out from that vertex.

1.20

Note that

$$\frac{1}{2} \sum_{u \neq v \in \mathcal{V}; w \neq u, v \in \mathcal{V}} a_{u,v} a_{w,v} a_{u,w}$$

is the number of triangles involving v in the graph. Similarly,

$$\frac{1}{2} \sum_{u \neq w \in \mathcal{V}} a_{u,v} a_{w,v}$$

is the number of *2-stars* centred around v in the graph. The local clustering coefficient describes how "locally dense" a graph is.

1.22

1.2.2. The local clustering coefficient

The *local clustering coefficient* of a vertex v is, intuitively, the proportion of its "friends" who are friends themselves. Mathematically, it is the proportion of neighbours of v which are neighbours themselves. In adjacency matrix notation,

$$C(v) = \frac{\sum_{u,w \in \mathcal{V}} a_{u,v} a_{w,v} a_{u,w}}{\sum_{u,w \in \mathcal{V}; u \neq w} a_{u,v} a_{w,v}}.$$

Here $0/0 := 0$.

The *average clustering coefficient* is defined as

$$\bar{C} = \frac{1}{|\mathcal{V}|} \sum_{v \in \mathcal{V}} C(v).$$

1.21

1.2.3. The global clustering coefficient

The *global clustering coefficient* or *transitivity* is defined as

$$C = \frac{3 \times \text{number of triangles}}{\text{number of connected triples}}.$$

By a connected triple we mean three vertices a, b, c with edges (a, b) and (b, c) present, while the edge (a, c) may or may not be present.

Equivalently,

$$C = \frac{6 \times \text{number of triangles}}{\text{number of paths of length two}}$$

as each triangle in the network abc gives rise to six paths of length 2: $a - b - c, b - c - a, c - a - b$ and $a - c - b, c - b - a, b - a - c$.

1.23

1.2.4. The expected clustering coefficient

Note that $\bar{C} \neq C$ in general. Indeed \bar{C} tends to be dominated by vertices with low degree, since they tend to have small denominators in the local clustering coefficient.

The global clustering coefficient in the Florentine family example is 0.1914894; the global clustering coefficient in the Yeast data is 0.1023149. The average clustering coefficient in the Florentine family example is 0.1395833.

For models of random networks often we consider the *expected clustering coefficient*

$$E(C) = \frac{3 \times \mathbb{E}(\text{number of triangles})}{\mathbb{E}(\text{number of connected triples})}.$$

Unfortunately all of the average clustering coefficient, the global clustering coefficient, and the expected clustering coefficient are often just called *the clustering coefficient* in the literature. Here we mean by *clustering coefficient* the global clustering coefficient, unless otherwise stated.

1.24

1.25

1.2.5. The average shortest path

In a graph a *path* from vertex v_0 to vertex v_n is an alternating sequence of vertices and edges, $(v_0, e_1, v_1, e_2, \dots, v_{n-1}, e_n, v_n)$ such that the endpoints of e_i are v_{i-1} and v_i , for $i = 1, \dots, n$. The *distance* $\ell(u, v)$ between two vertices u and v is the length of the shortest path joining them. This path does not have to be unique.

We can calculate the distance $\ell(u, v)$ from the adjacency matrix A as the smallest power p of A such that the (u, v) -element of A^p is not zero.

A graph is called *connected* if there is a walk between any pair of vertices in the graph, otherwise it is called *disconnected*.

In a connected graph, the *average shortest path length* is defined as

$$\ell = \frac{1}{|\mathcal{V}|(|\mathcal{V}| - 1)} \sum_{u \neq v \in \mathcal{V}} \ell(u, v).$$

The average shortest path length describes how "globally connected" a graph is.

1.26

1.27

1.2.6. Small graphs and motifs

In addition to considering these general summary statistics, it has proven fruitful to summarise networks in terms of the small graphs which are contained in the network. Such small subgraphs can be viewed as building-block patterns of networks. By *small* we mean graphs on a small number of vertices, such as 3 - 5 vertices.

Often a small graph is called a *motif* when it is over-represented in the network. Over-representation is judged using a probabilistic model for the network.

Here we think of a motif as a small graph with a fixed number of vertices and with a given topology, and we use the term interchangeably with *small graph*.

In biological networks, it turns out that motifs seem to be conserved across species. They seem to reflect functional units which combine to regulate the cellular behaviour as a whole.

1.28

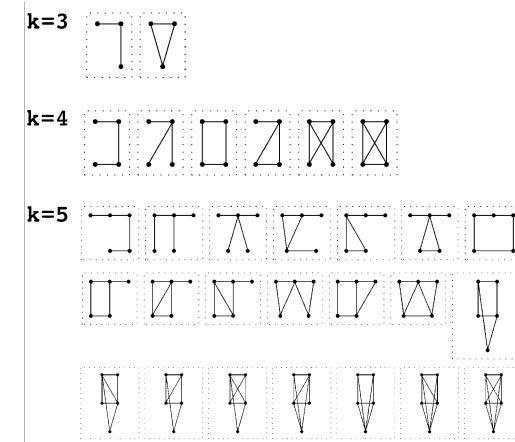


Figure: Some small graphs (motifs)

1.29

1.2.7. Network Centrality Measures

Network centrality measures attempt to quantify the importance of vertices, edges, or other network structures in various ways.

The choice depends on the problem and data under study, as what it means to be "most central" (and most important) is obviously context-dependent. The perhaps simplest centrality measure is degree centrality (or just: degree), but there are many others, such as closeness centrality (the average distance to the other vertices in the network), betweenness centrality (the number of shortest path passing through the vertex or edge), and centrality measures based on the spectral properties of the adjacency matrix (eigenvectors, eigenvalues) such as eigenvector centrality, Katz centrality, and the Page Rank centrality which was used in early versions of Google.

Centralities are often not robust to small perturbations either of their definition or of network structure; hence caution is advised when interpreting the results of centrality calculations.

1.30

Different measures of centrality

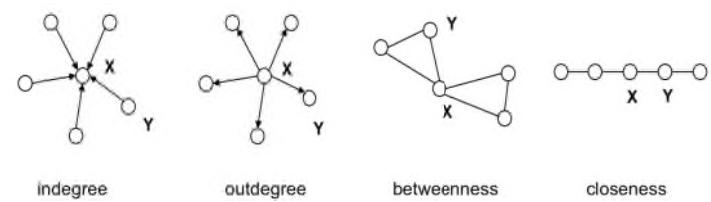


Figure: In each of the following networks, X has higher centrality than Y according to a particular measure.

1.31

Other summaries

The above network summaries provide an initial go at networks. Specific networks may require specific concepts. For example, in fungal networks, there are hardly any triangles, so the clustering coefficient does not make much sense for these networks.

When there is spatial structure in the network, then geometrical considerations may enter; we do not cover this situation in these lectures.

1.32

Excursion: Milgram and the *small world effect*

In 1967 the American sociologist Milgram reported a series of experiments of the following type. A number of people from a remote US state (Nebraska, say) are asked to have a letter (or package) delivered to a certain person in Boston, Massachusetts (such as the wife of a divinity student). The catch is that the letter can only be sent to someone whom the current holder knew on a first-name basis. Milgram kept track of how many intermediaries were required until the letters arrived.

1.33

Milgram reported a median of six. This made him coin the notion of *six degrees of separation*, often interpreted as everyone being six handshakes away from the President. While the experiments were somewhat flawed (in the first experiment only 3 letters arrived), the concept of *six degrees of separation* has stuck.

For more details see for example the report by Judith Kleinfeld at http://www.columbia.edu/itc/sociology/watts/w3233/client_edit/big_world.html.

1.34

2 Models for Networks

Below is an example from scientific collaboration networks (*N. Boccara, Modeling Complex Systems, Springer 2004, p.283*).

Network	n	ave degree	C
Los Alamos	52,909	9.7	0.43
MEDLINE	1,520,251	18.1	0.066
NCSTRL	11,994	3.59	0.496

What do we learn from these summaries?

In order to judge whether a network summary is "unusual" or whether a motif is "frequent", there is an underlying assumption of randomness in the network. The randomness can be intrinsic to the network, and/or may stem from errors in the data.

To understand the randomness, mathematical models have been suggested.

2.1

2.1. Bernoulli (Erdös-Renyi) random graphs

The most standard random graph model is that of Erdös and Renyi (1959). The vertex set \mathcal{V} of finite size n is given, and an edge between two vertices is present with probability p , independently of all other edges.

The expected number of edges is

$$\binom{n}{2}p.$$

Each vertex has $n - 1$ potential neighbours, and each of these $n - 1$ edges is present with probability p , and so the degree of a randomly picked vertex is $\text{Bin}(n - 1, p)$ -distributed.

2.2

The expected number of triangles in the graph is

$$\binom{n}{3}p^3 = \frac{n(n-1)(n-2)}{6}p^3.$$

The expected clustering coefficient is p .

In a Bernoulli random graphs, your friends are no more likely to be friends themselves than would be a two complete strangers.

2.3

In the example from scientific collaboration networks (*N. Boccara, Modeling Complex Systems, Springer 2004, p.283*) we estimate p as the proportion of possible edges which are present in the network; this estimate is directly proportional to the average degree (see Problem Sheet 1) and would also be an estimate of the expected clustering coefficient in a Bernoulli random graph.

Network	n	ave degree	C	$EC_{\text{Bernoulli}}$
Los Alamos	52,909	9.7	0.43	0.00018
MEDLINE	1,520,251	18.1	0.066	0.000011
NCSTRL	11,994	3.59	0.496	0.0003

The average clustering coefficient is much larger than what we would expect from a Bernoulli random graph!

Beware: EC is not $\mathbb{E}(C)$ so the interpretation is not completely clear.

2.4

The Small World phenomenon

Also in real-world graphs often the shortest path length is much shorter than expected from a Bernoulli random graph with the same average vertex degree.

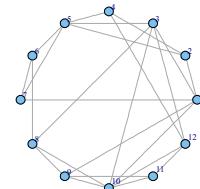
The phenomenon of short paths, often coupled with high clustering coefficient, is called the *small world phenomenon*. Remember the Milgram experiments!

2.5

2.2. The Watts-Strogatz small world model

Arrange n vertices on a ring. Hard-wire each vertex to its k nearest neighbours on each side on the ring.

Choose a vertex and the edge that connects this vertex to its clockwise nearest neighbour. With probability ϕ this edge is reconnected to a vertex chosen uniformly at random over the ring, with duplicate edges excluded; otherwise the edge is left in place. This “rewiring” process is repeated by moving clockwise around the ring. Next the edges are considered which connect vertices to their second-nearest neighbours, again clockwise, and we repeat the rewiring process. With nk edges in the network the process stops after k laps. The number of edges remains nk .



2.6

If there are no shortcuts, then the average distance between two randomly chosen vertices is of the order n , the number of vertices. But we shall see that as soon as there are just a few shortcuts, then the average distance between two randomly chosen vertices has an expectation of order $\log n$. Thinking of an epidemic on a graph - just a few shortcuts dramatically increase the speed at which the disease is spread.

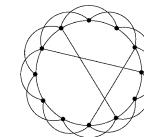
While the Watts-Strogatz model is able to replicate a wide range of clustering coefficient and shortest path length simultaneously, it falls short of producing the observed types of vertex degree distributions. It is often observed that vertices tend to attach to “popular” vertices; popularity is attractive.

2.8

The Newman-Moore-Watts model

The version of the Watts-Strogatz model which is currently most used is often called the *Newman-Moore-Watts model*, also known as the great circle model, *Ball, Mollison and Scalia-Tomba 1997*: Again arrange the n vertices of \mathcal{V} on a lattice and hard-wire each vertex to its k nearest neighbours on each side on the lattice, where k is small. Now do not rewire edges but instead introduce random shortcuts between vertices which are not hard-wired; the shortcuts are chosen independently, all with the same probability p . Thus the number of edges is no longer constant.

The degree of a randomly chosen vertex has distribution $2k + \text{Bin}(n - 2k - 1, p)$.



2.7

2.3. "The" Barabasi-Albert model

In 1999, Barabasi and Albert noticed that the actor collaboration graph and the World Wide Web had degree distributions that were of the type

$$d_k \sim Ck^{-\gamma}$$

for $k \rightarrow \infty$. Such behaviour is called *power-law behaviour*; the constant γ is called the *power-law exponent*. Subsequently a number of networks have been identified which show this type of behaviour. They are also called *scale-free random graphs*.

2.9

To explain this behaviour, Barabasi and Albert introduced the *preferential attachment* model for network growth. Suppose that the process starts at time 1 with 2 vertices linked by m (parallel) edges. At every time $t \geq 2$ we add a new vertex with m edges that link the new vertex to vertices already present in the network. We assume that the probability π_i that the new vertex will be connected to a vertex i depends on the degree $d(i)$ of i so that

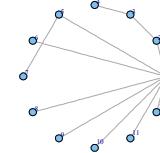
$$\pi_i = \frac{d(i)}{\sum_j d(j)}.$$

To be precise, when we add a new vertex we will add edges one at a time, with the second and subsequent edges doing preferential attachment using the updated degrees.

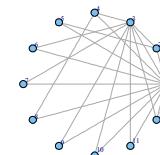
Other modifications vary the probability for choosing a given vertex.

2.10

For $m = 1$ the above construction will not result in any triangles at all.



For $m = 2$ it is possible to modify the construction so that *any* distribution of triangles can be achieved while keeping the marginal distribution π_i the same. [Note that in the figure the multiple edge between 1 and 2 is merged into one edge.]



2.11

2.4. The stochastic blockmodel

The stochastic blockmodel, also called *Erdős-Renyi mixture model* and *latent block models* (*Holland, Laskey and Leinhardt (1983), Nowicki and Snijders (2001)*), assumes that vertices are of different types, say, there are L different types. Then edges are constructed independently, such that the probability for an edge varies only depending on the type of the vertices at the endpoints of the edge;

$$p_{i,j} = \mathbb{P}((u, v) \in \mathcal{E} | u \text{ is of type } i, v \text{ is of type } j).$$

If $\alpha_1, \dots, \alpha_L$ denote the proportion of the vertices of different types, $\sum_\ell \alpha_\ell = 1$, then for a vertex V picked uniformly at random from \mathcal{V} ,

$$\mathbb{E}(d(V)) = \sum_\ell \alpha_\ell \left((|\mathcal{V}| \alpha_\ell - 1) p_{\ell,\ell} + \sum_{k \neq \ell} |\mathcal{V}| \alpha_k p_{k,\ell} \right).$$

Often the type allocation itself is not known.

2.12

2.5. The configuration model

The configuration model is a model which gives a graph which is uniformly chosen among all graphs with a given degree sequence.

Given a feasible degree sequence, $(d(i), i = 1, \dots, n)$ and $m = \frac{1}{2} \sum_i d(i)$, construct a graph as follows.

1. Give each vertex i a total of $d(i)$ “stubs” of edges.
2. Choose two of the stubs uniformly at random and create an edge by connecting them to one another.
3. Choose another pair from the remaining $2m - 2$ stubs and connect these.
4. Continue until all the stubs are used up.

In the resulting network every vertex has exactly the desired degree.

Here we assume that all degrees are at least 1.

2.13

We note that more than one set of matchings may result in the same network; if we labelled the stubs there would be typically many ways we can join up pairs of labelled stubs to create the same final configuration.

The model lends itself to analytical considerations through noting that the expected number of edges between vertices i and $j \neq i$ is

$$\frac{d(i)d(j)}{2m-1}.$$

The construction may create self-loops and multiple edges. But the density of self-loops and multiple edges tends to zero as the number of vertices tends to infinity. Let's see why this is so, for the density of self-loops:

2.14

Self-loops

The expected number of self-loops from i to i is

$$\frac{d(i)(d(i)-1)}{2(2m-1)}.$$

Hence the expected number of self-loops is

$$\sum_{i=1}^n \frac{d(i)(d(i)-1)}{2(2m-1)} = \frac{\mathbb{E}\{d(V)^2\} - \mathbb{E}d(V)}{2(\mathbb{E}d(V) - 1/n)}$$

where $d(V)$ is the degree of a randomly chosen vertex; $\mathbb{E}d(V) = \frac{1}{n} \sum_{i=1}^n d(i) = \frac{2m}{n}$. As long as $\mathbb{E}d(V)^2$ remains bounded, the expected number of self-loops will remain bounded, whereas the number of edges will grow with the number of vertices in the network. Hence self-loops will be rare when the network is large.

A similar argument applies to multiple edges.

2.15

2.6. The Chung-Lu model

A variant of the configuration model is a random graph model with a given *expected* degree sequence, fixing the probability of edges but not their actual number. The *Chung-Lu* model (also called *Newman-Girvan model*) is as follows: given a sequence of non-negative integers $\{w_1, \dots, w_n\}$ a graph on n vertices is constructed by letting

$$\theta_i = \frac{w_i}{\sqrt{\sum_j w_j}}$$

and an edge between vertices i and j is created with probability $\min(1, \theta_i \theta_j)$, independently of all other edges.

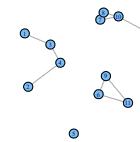
Often we assume that $\max_i w_i^2 < \sum_k w_k$. Self-loops are allowed.

In this model the expected degree of vertex i is w_i .

2.16

2.7. Geometric random graphs

In geometric random graphs, introduced by Gilbert in 1961, n points X_1, \dots, X_n are chosen independently at random according to a density $f(\cdot)$ on \mathbb{R}^d . A value $r > 0$ is chosen and we put an edge (i, j) if $d(X_i, X_j) \leq r$. Here d is a distance on \mathbb{R}^d ; usually we choose the Euclidean distance.



The 2003 book by Penrose contains many more details.

2.17

2.8. Exponential random graph (p^*) models

Networks have been analysed for a long time in the social science literature, see for example the book by Wasserman and Faust. Here usually directed graphs are studied. Typical research questions could be

- ▶ Is there a tendency in friendship towards transitivity; are friends of friends my friends?
- ▶ What is the role of explanatory variables such as income on the position in the network?
- ▶ What is the role of friendship in creating behaviour (such as smoking)?
- ▶ Is there a hierarchy in the network?
- ▶ Is the network influenced by other networks for which the membership overlaps?

2.18

Exponential random graph (p^*) models model the whole adjacency matrix \mathbf{X} of a graph simultaneously, making it easy to incorporate dependence. The general form of the model is

$$\mathbb{P}(\mathbf{X} = \mathbf{x}) = \frac{1}{\kappa} \exp \{ \theta' \mathbf{z}(\mathbf{x}) \},$$

where θ is a vector of model parameters and $\mathbf{z}(\mathbf{x})$ is a vector of network statistics; κ is a normalising quantity so that the probabilities sum to 1.

The simplest such model is that the probability of any edge is constant across all possible edges, i.e. the Bernoulli graph, for which

$$\mathbb{P}(\mathbf{X} = \mathbf{x}) = \frac{1}{\kappa} \exp \{ \lambda L(\mathbf{x}) \},$$

where $L(\mathbf{x})$ is the number of edges in the network \mathbf{x} and λ is a parameter.

2.19

2.9. Specific models for specific networks

Depending on the research question, it may make sense to build a specific network model. For protein-protein interaction networks gene duplication models have been proposed as models for network growth.

When thinking of flows through networks, it may be a good idea to use weighted networks; the weights could themselves be random, or they could be modelled for example using differential equations.

2.20

The need for the distribution of summary statistics under network models

For hypothesis testing we would like to know the distribution of the summary statistic under different network models.

Such distributional results, if available at all, are of asymptotic nature. The appropriate asymptotic regime will depend on the parameter values - for example, in Bernoulli random graphs, the number of triangles will be approximately Poisson if p is small, and approximately normal if p is moderately large.

The results have to take dependence between counts into account, and hence we cannot just apply the classical limit theorems but need results for counts in the presence of dependence. As this is not a mathematics course, we shall not derive such distributional results, but we shall mention them briefly in the next chapter.

2.21

3 Summary of distributional results for network statistics

The summary statistics which we focus on are

- ▶ average degree, and/or the degree of a randomly chosen vertex;
- ▶ the global clustering coefficient, and/or the number of triangles;
- ▶ the average shortest path, and/or the shortest path between two randomly chosen vertices.

We shall mainly concentrate on Bernoulli random graphs, denoted by $\mathcal{G}(n, p)$, with n the number of vertices and p the edge probability.

3.1

3.1 The degree of a randomly chosen vertex

The degree of a randomly chosen vertex $\mathcal{G}(n, p)$ is binomial $\text{Bin}(n - 1, p)$;

$$\mathbb{P}(\text{degree}(v) = k) = \binom{n-1}{k} p^k (1-p)^{n-1-k}, \quad k = 0, 1, \dots, n-1,$$

where

$$\binom{n-1}{k} = \frac{(n-1)(n-2)\cdots(n-k)}{k(k-1)\cdots2\cdot1}$$

is the so-called binomial coefficient. For large n this is difficult to compute and hence approximations are useful.

3.2

Recall the Poisson(λ) distribution

$$p_k = e^{-\lambda} \frac{\lambda^k}{k!}, \quad k = 0, 1, \dots$$

One can show that in general, if $P = \text{Binomial}(n, p)$ and if $Q = \text{Poisson}(np)$, then

$$\sup_{A \subset \{0, 1, \dots\}} |P(A) - Q(A)| \leq \min(np^2, p),$$

see for example Chapter 1 in *A.D. Barbour, L. Holst, and S. Janson. Poisson Approximation. Oxford University Press, 1992.*

In particular, if $X_n \sim \text{Binomial}(n, p)$ and $X \sim \text{Poisson}(np)$, then for all $x \in \mathbb{R}$,

$$\mathbb{P}(X \leq x) - p \leq \mathbb{P}(X_n \leq x) \leq \mathbb{P}(X \leq x) + p.$$

The bound holds for any n and p . The approximation is useful if p is small. For example, if $p = \frac{c}{n}$ then the bound goes to 0 as $n \rightarrow \infty$.

3.3

When p is not small, then a normal approximation is more useful. One can show in general that if $X_n \sim \text{Binomial}(n, p)$ then $W_n = \frac{X_n - np}{\sqrt{np(1-p)}}$ has mean 0 and variance 1. If Φ denotes the cumulative distribution function of the standard normal distribution, then one can show that

$$\sup_x |\mathbb{P}(W_n \leq x) - \Phi(x)| \leq 0.4785 \frac{1}{\sqrt{n}} \cdot \frac{p(1-p)^3 + (1-p)p^3}{(p(1-p))^{\frac{3}{2}}},$$

see for example p.45 in *L.H.Y. Chen, L. Goldstein, and Q.-M. Shao. Normal Approximation by Stein's Method. Springer, 2011.*

Again this bound holds for all n and p . When p is not small then the bound will go to 0 as $n \rightarrow \infty$. If $p = \frac{c}{n}$ though then the bound will not be small - the Poisson approximation is much better in that case!

3.4

Dense and sparse

The degree of a randomly chosen vertex in a Bernoulli random graph behaves quite differently depending on whether p is small or not.

The expected degree of a vertex in a Bernoulli random graph is $(n - 1)p$.

In general, a graph with n vertices is called

- ▶ *dense* if the expected degree of a randomly picked vertex tends to infinity as $n \rightarrow \infty$;
- ▶ *sparse* if the expected degree of a randomly picked vertex tends to a finite value as $n \rightarrow \infty$.

3.5

Other models

Also in ER mixture graphs, similar arguments give Poisson and normal approximations, with explicit bounds depending on the edge probabilities. Again a Poisson approximation is good in the sparse regime, and a normal approximation is advised in the dense regime.

Similar results hold for geometric random graphs.

In Barabasi-Albert networks, the expected degree of a randomly chosen vertex obeys a power law; this is how the networks are designed. This power law is usually very different from a normal or a Poisson distribution.

For graphs which are generated under the configuration model, the degree of a randomly picked vertex is given through the degree sequence which underpins the model.

3.6

3.2 The number of triangles

Again we consider a Bernoulli random graphs, denoted by $\mathcal{G}(n, p)$. The probability of a triangle occurring at a triple (a, b, c) of vertices is p^3 . The occurrences of triangles are not independent – if there is a triangle at (a, b, c) , then the probability of a triangle at (a, b, d) with $d \neq a, b, c$ is no longer p^3 but instead p^2 .

This dependence is weak, in the sense that while there are $\binom{n}{3}$ potential triangles in the network, each potential triangle occurrence depends only on $3(n - 3)$ other triangle occurrences.

It is possible to still derive a Poisson and a normal approximation for the number of triangles, again with explicit bounds.

3.5

Poisson approximation

If W_n denotes the number of triangles in $\mathcal{G}(n, p)$, then its expectation is

$$\lambda_n = \mathbb{E}(W_n) = \binom{n}{3}p^3.$$

Let Z denote a Poisson random variable with parameter λ_n .

One can show that

$$\sup_{A \subset \{0, 1, \dots\}} |\mathbb{P}(W_n \in A) - \mathbb{P}(Z \in A)| \leq \binom{n}{3}p^3(3np^3 + 3np^2) \min(1, \lambda_n^{-1}),$$

see for example Chapter 5 in *A.D. Barbour, L. Holst, and S. Janson. Poisson Approximation. Oxford University Press, 1992.*

If $p = \frac{c}{n}$ then the bound is of order cn^{-1} .

If p is not small then the bound will tend to ∞ as $n \rightarrow \infty$.

3.7

3.8

Normal approximation

If W_n denotes the number of triangles in $\mathcal{G}(n, p)$, then its expectation is $\binom{n}{3}p^3$ and one can calculate that its variance is

$$\text{Var}(W_n) = \sigma^2 = \binom{n}{3}p^3[1 - p^3 + 3(n - 3)p^2(1 - p)].$$

Let $T_n = \frac{W_n - \mathbb{E}(W_n)}{\sigma}$. Let Z denote a standard normal random variable. One can show that for any continuous and piecewise continuously differentiable h ,

$$|\mathbb{E}h(T_n) - \mathbb{E}h(Z)| \leq \frac{23n^5p^3}{\sigma^3} ||h'||,$$

see Barbour AD, Karoński M, Ruciński A. *A central limit theorem for decomposable random variables with applications to random graphs*. Journal of Combinatorial Theory, Series B.

1989 Oct 1;47(2):125-45.

When p does not depend on n then this expression goes to 0 as $n \rightarrow \infty$. When $p = \frac{c}{n}$ then σ^2 is bounded by a constant and the normal approximation bound does not tend to 0 as $n \rightarrow \infty$.

3.9

Other models

Also in ER mixture graphs, similar arguments give Poisson and normal approximations, with explicit bounds depending on the edge probabilities.

Similar results hold for geometric random graphs.

In Barabasi-Albert networks, if only one edge is attached at a time, then there are no triangles in the network. If more than one edge is attached at the time, then the number of triangles depends on the model specification. One can show that the joint probabilities for where these edges are formed can be specified to achieve any clustering coefficient, see for example the discussion in R. Durrett, *Random Graph Dynamics*. Cambridge University Press, 2007.

For graphs which are generated under the configuration model, the number of triangles depends strongly on the degree sequence which underpins the model.

3.10

3.3 Shortest paths

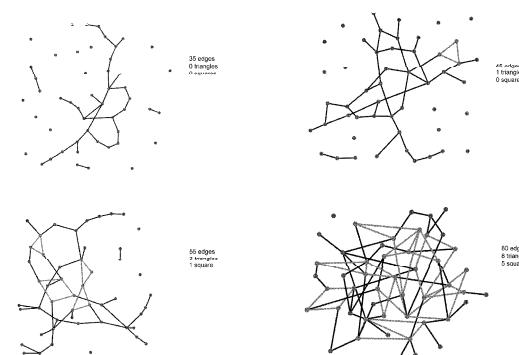
The shortest path between two vertices is infinite when the two vertices are not in the same connected component of the network.

It is possible that a Bernoulli random graph has more than one component. Indeed there is a threshold behaviour which can be observed around $p = \frac{1}{n}$.

Here is a simulation.

3.11

A simulation



3.12

A mathematical formulation

In a Bernoulli $\mathcal{G}(n, p)$ random graph with $p = \frac{c}{n}$, when $c < 1$ then the largest component has size of order at most $\log n$ and when $c > 1$ then the largest component has linear size with probability tending to 1 as n grows to infinity.

To avoid the complication of having a positive probability of infinite shortest paths, we use the Newman-Watts-Strogatz small world model instead.

3.13

Let us consider the following continuous circle mode which is a slight modification of the Newman-Watts-Strogatz model: Let C be a circle of circumference L . On this circle a Poisson $(L\rho/2)$ number of shortcuts added uniformly. Chords between points have length zero. Assume $L\rho > 1$.

Let \mathcal{D} denote shortest distance between two randomly chosen points P and Q . We are interested in the distribution of \mathcal{D} .

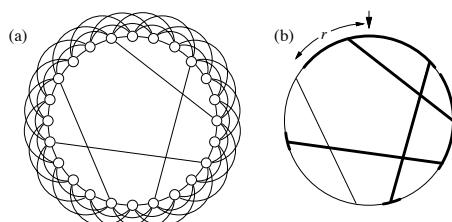
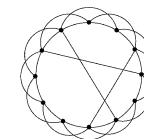


Figure: From Newman, Moore and Watts (1999)

3.15

The Newman-Watts-Strogatz model

Recall the Newman-Watts-Strogatz small world model: Again arrange L vertices of \mathcal{V} on a ring lattice and hard-wire each vertex to its k nearest neighbours on each side on the lattice, where k is small. Now do not rewire edges but instead introduce random shortcuts between vertices which are not hard-wired; the shortcuts are chosen independently, all with the same probability p . What is the distribution of the shortest path between two randomly chosen points?



3.14

Then the following result can be shown:

If T denotes a random variable with distribution given by

$$\mathbb{P}[T > x] = \int_0^\infty \frac{e^{-y} dy}{1 + e^{2x} y}$$

and $D^* = \frac{1}{\rho} \left(\frac{1}{2} \log(L\rho) + T \right)$, then

$$\sup_x |\mathbb{P}[\mathcal{D} \leq x] - \mathbb{P}[D^* \leq x]| = O \left((L\rho)^{-\frac{1}{5}} \log^2(L\rho) \right).$$

Here the O -notation means “of order at most”.

A related result holds for the analogous discrete model. For references see Barbour, A.D. and Reinert, G., 2001. *Small worlds. Random Structures & Algorithms*, 19(1), pp.54-74. Correction: *Random Structures & Algorithms*, 25(1), p.115 (2004) and Barbour, A. and Reinert, G., 2006. *Discrete small world networks. Electronic Journal of Probability*, 11, pp.1234-1283.

3.16

4 Sampling from networks

Often the whole network is too large and we only observe a portion of the network.

Sometimes a large network may be available for study but it is so complex that it may obscure features of interest; then we would like to restrict attention to part of the network.

If we are interested in a rare feature such as the number of cocaine users in a large city, then a simple random sample may not be efficient.

If we want to draw inference from the whole network based on a sample from the network then we need to take the specifics of the sampling procedure into account.

Sampling schemes

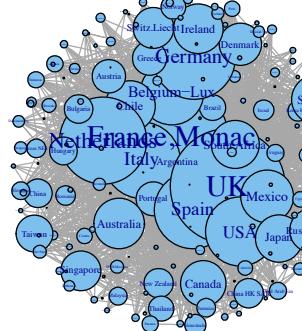
The main sampling schemes for networks which we shall consider are

1. induced subgraph sampling - we take a random sample of vertices and observe all edges between the vertices in the sample.
 2. snowball sampling - in a 1-hop snowball sample we start with one vertex, sample all of its edges and neighbours and, depending on the scheme, the edges between its neighbours too. In a 2-hop snowball sample we start with one vertex and sample all vertices within distance 2 of the vertex (i.e. at most two edges away), and depending on the scheme, also sample the edges between those neighbours.

4.

4.2

Example: Beverage trade network for the year 200



4.3

Induced subgraph sample

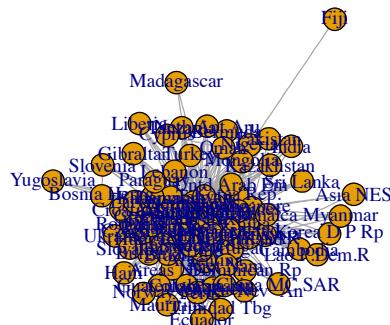
We sample 20 vertices at random - in our realisation the country labels are "China", "Liberia", "St.Kt-Nev-An", "Mauritius", "Oman", "Colombia", "Lebanon", "Madagascar", "Poland", "Viet Nam", "Malaysia", "Ukraine", "Dominican Rp", "Bosnia Herzg", "Gibraltar", "Haiti", "Sri Lanka", "Tanzania", "Untd Arab Em", and "Nigeria".



4.4

Snowball sample

In contrast, the corresponding step-1 snowball sample has 87 vertices and 1030 edges.



4.5

Suppose that we have a population $U = \{1, \dots, N\}$ of units (for example people) and with each unit $i \in U$ is associated a value y_i of interest (for example age). Let

$$\tau = \sum_{i \in U} y_i$$

be the population total, and let

$$\mu = \frac{\tau}{N}$$

be the population average. We also need the population variance.

$$\sigma^2 = \frac{1}{N} \sum_{i \in I^U} (y_i - \mu)^2.$$

Let $S = (i_1, \dots, i_n)$ be a sample of n units from U and observe that we observe y_i for each element in the sample. The goal is to estimate μ and τ .

4.1 Sampling designs

Here we assume that observations are made without measurement error, and that the only source of randomness is the sampling design. This will help to understand the issues which arise from sampling alone, not taking any probabilistic model for the network into account.

4.6

The Horvitz-Thompson estimator

Suppose that each unit $i \in U$ has probability $\pi_i > 0$ to be included in S . Then the *Horvitz-Thompson estimator* of τ is

$$\hat{\tau}_\pi = \sum_{i \in S} \frac{y_i}{\pi_i}.$$

The corresponding estimator of the mean is

$$\hat{\mu}_\pi = \frac{1}{N} \hat{\tau}_\pi.$$

It is possible to show that $E\hat{\tau}_\pi = \tau$; so that $\hat{\tau}_\pi$ is an unbiased estimator for τ . Similarly it can be shown that

$$\text{Var}(\hat{\tau}_\pi) = \sum_{i \in U} \sum_{j \in U} y_i y_j \left(\frac{\pi_{i,j}}{\pi_i \pi_j} - 1 \right).$$

Here $\pi_{i,j}$ is the probability that both i and j are included in S ; when $i = j$ then this is just π_i .

4.7

4.8

Example: Simple random sampling

Simple random sampling is random sampling without replacement. There are $\binom{N}{n}$ possible simple random samples of size n from U , and $\binom{N-1}{n-1}$ contain unit i . Hence

$$\pi_i = \frac{\binom{N-1}{n-1}}{\binom{N}{n}} = \frac{n}{N}.$$

Similarly, for $i \neq j$,

$$\pi_{i,j} = \frac{n(n-1)}{N(N-1)}.$$

The Horvitz-Thompson estimator of the total and the mean are

$$\hat{\mu} = \frac{1}{n} \sum_{i \in S} y_i; \quad \hat{\tau} = N\bar{y}.$$

Their variance can also be calculated; for $n \geq 2$

$$\text{Var}(\hat{\tau}_\pi) = \frac{N^2(N-n)}{n(N-1)} \sigma^2.$$

4.9

4.10

Induced subgraph sampling

In this design, a simple random sample of n vertices is selected from the vertex set \mathcal{V} , without replacement, and edges are observed for all vertex pairs in the sample. Then, for any $i \in \mathcal{V}$,

$$\pi_i = \frac{n}{|\mathcal{V}|}$$

and, for $i \neq j$,

$$\pi_{i,j} = \frac{n(n-1)}{|\mathcal{V}|(|\mathcal{V}|-1)},$$

which is the probability that both u and v , and hence the edge indicator for (u, v) , are observed. Note that these probabilities assume that we know $|\mathcal{V}|$.

4.11

Sampling schemes from networks

The previous section discussed sampling schemes in general. in this section we shall specify them to the situation that the underlying population is represented by a network.

To apply the Horvitz-Thompson theory to networks, there are different choices of populations. As we may be interested in vertex characteristics as well as edge characteristics, for a simple network $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ let us choose

$$U = \mathcal{V} \cup \{(u, v) : u, v \in \mathcal{V}, u \neq v\},$$

the set of vertices together with the set of potential edges. A characteristic of a potential edge could include whether or not the edge is present in \mathcal{E} . So

$$N = |U| = |\mathcal{V}| + |\mathcal{V}|(|\mathcal{V}|-1).$$

Example: the number of edges

Suppose that we would like to estimate $|\mathcal{E}|$, the number of edges in a simple undirected network \mathcal{G} , using a sample of size n from U .

We write the number of edges as half the population total τ of values $y_i, i \in U$, where $y_u = 0$ for $u \in \mathcal{V}$ and $y_{u,v} = \mathbf{1}((u, v) \in \mathcal{E})$, that is, $y_{u,v} = 1$ if $(u, v) \in \mathcal{E}$, and 0 otherwise.

The Horvitz-Thompson estimator for τ is then

$$\hat{\tau}_\pi = \frac{|\mathcal{V}|(|\mathcal{V}|-1)}{n(n-1)} \sum_{u \in S} \sum_{v \in S} \mathbf{1}((u, v) \in \mathcal{E}),$$

which is the average degree in the induced subnetwork, multiplied by $(|\mathcal{V}|(|\mathcal{V}|-1))/(n-1)$.

For the variance we need the probability $\pi_{u,v,s,t}$ that vertices u, v, s and t are included in the sample.

4.12

Snowball sampling

For snowball sampling the probabilities π_u become considerably more involved.

There is one particular case which is still reasonably easy to treat: star sampling. Star sampling is a special case of snowball sampling. A random sample S_0 of n vertices is selected from the vertex set \mathcal{V} . For each vertex in the sample all edges that are incident to it are observed. For simple random sampling of the vertices, if we include only the vertices in S_0 , then for any $i \in \mathcal{V}$, $\pi_i = \frac{n}{|\mathcal{V}|}$ and, for $i \neq j$, $\pi_{i,j} = \frac{n(n-1)}{|\mathcal{V}|(|\mathcal{V}|-1)}$.

If we include in S not only S_0 but also all adjacent vertices of all vertices in S_0 then the vertex inclusion probabilities are getting more complicated.

Variants

Instead of attaching variables to vertices and pairs of vertices, we could attach a value $y_{i,j,k}$ to triples of vertices. Using $y_{i,j,k}$ as the indicator that a triangle is present, this approach can be used to estimate the total number of triangles in the network.

Similarly one could attach values to other collections of vertices.

4.13

4.2 Sampling and the randomness in the network: A cautionary tale

In the previous section we assumed that the network was fixed and the interest was in estimating characteristics of the network. When the network $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ itself is random then it is of interest to assess whether a sample from the network would reflect the underlying randomness in the network.

As an illustration, assume that the underlying network is a realisation of a $\mathcal{G}(N, p)$ Bernoulli random graph. Sample n vertices at random, without replacement, and consider the induced subgraph on these n vertices. It is not hard to see that this induced subgraph will look like a sample from a $\mathcal{G}(n, p)$ Bernoulli random graph: an edge between two vertices in the sample is present with probability p , independently of all other edges in the random graph.

This conservation of model structure when sampling is not typical. Even in a Bernoulli random graph, a snowball sample would look different to a Bernoulli random graph - for example the snowball sample of a vertex is by construction connected.

In Stumpf et al. (2005), *Proceedings of the National Academy of Sciences of the United States of America*, 102(12), pp.4221-4224 it is shown that such an induced subgraph sample of a scale-free random graph is no longer scale-free. This lack of conservation results in a conundrum: if the observed network is scale-free but it is known to be a sample from a larger network, then the larger network cannot be scale-free.

Shalizi and Rinaldo (2013), *Annals of Statistics*, 41(2), p.508 show that in exponential random graph models it is usually not the case that a sampled network follows an exponential random graph model also; it is only true under very special model specifications. While these model specifications cover the stochastic blockmodel only when the vertex classes are known.

4.14

4.15

5 Fitting a model: parametric approaches

Parametric just means that we have a finite set of parameters which fully specify the model. Otherwise we call the setting *non-parametric*. For example:

Bernoulli (Erdős-Renyi) random graphs. In the random graph model of Erdős and Renyi (1959), the (finite) vertex set V is given, say $|V| = n$. We denote the set of all potential edges by E ; thus $|E| = \binom{n}{2}$. An edge between two vertices is present with probability p , independently of all other edges. Here p is an unknown parameter.

5.1 Maximum likelihood estimation

In classical (frequentist) statistics we often estimate unknown parameters via the method of maximum likelihood.

The *likelihood* of the parameter given the data is just the probability of seeing the data we see, given the parameter.

Maximum-likelihood estimators have attractive properties; under some regularity conditions they would not only converge to the true parameter as the sample size tends to infinity, but it would also be approximately normally distributed if suitably standardized, and we can approximate the asymptotic variance.

5.1

5.2

Example: Bernoulli random graphs

Our data is the network we see. We describe the data using the adjacency matrix, denote it by \mathbf{x} here because it is the realisation of a random adjacency matrix \mathbf{X} . Recall that for a network on n vertices the adjacency matrix is the square $n \times n$ matrix where each entry is either 0 or 1;

$x_{u,v} = 1$ if and only if there is an edge between u and v .

We denote by $E = \{(u, v) \in \mathcal{V} \times \mathcal{V} : u < v\}$ the set of potential edges, so that $|E| = \binom{n}{2}$. As the number of edges is binomially distributed with parameters $\binom{n}{2}$ and p , the maximum likelihood estimator for p is

$$\hat{p} = \frac{t}{|E|}.$$

5.3

Example: Florentine family data

In the Florentine family data, we observe 16 vertices, 20 edges, an average vertex degree of 2.5, and 3 triangles.

The maximum-likelihood estimate for p is $\hat{p} = \frac{1}{6}$.

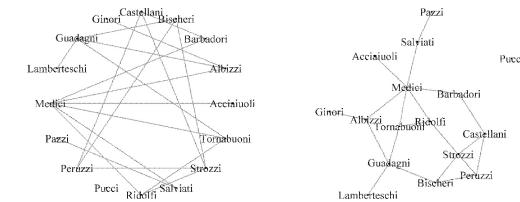


Figure: Marriage relations between Florentine families; two different graphical representations

5.4

Maximum-likelihood estimation also works well in Erdős-Renyi Mixture graphs when the number of types is known, and it works well in Watts-Strogatz small world networks when the number k of nearest neighbours we connect to is known. When the number of types, or the number of nearest neighbours, is unknown, then things become messy.

In Barabasi-Albert models, the parameter would be the power exponent for the vertex degree, as occurring in the probability for an incoming vertex to connect to some vertex i already in the network.

In exponential random graphs, unless the network is very small, maximum-likelihood estimation quickly becomes numerically unfeasible. Even in a simple model like

$$\mathbb{P}(\mathbf{X} = \mathbf{x}) = \frac{1}{\kappa} \exp\{\lambda_1 L(\mathbf{x}) + \lambda_2 S_2(\mathbf{x}) + \lambda_3 S_3(\mathbf{x}) + \lambda_4 T(\mathbf{x})\}$$

the calculation of the normalising constant κ becomes numerically impossible very quickly.

5.5

5.6

5.2 Markov Chain Monte Carlo estimation

When the likelihood is not available, Markov chain Monte Carlo methods can often be set up.

A Markov chain is a stochastic process where the state at time n only depends on the state at time $n - 1$, plus some independent randomness; a random walk is an example.

A Markov chain is *irreducible* if any set of states can be reached from any other state in a finite number of moves.

The Markov chain is *reversible* if you cannot tell whether it is running forwards in time or backwards in time.

A distribution is *stationary* for the Markov chain if, when you start in the stationary distribution, one step after you cannot tell whether you made any step or not; the distribution of the chain looks just the same.

There are mathematical definitions for these concepts, but we only need the main result here:

If a Markov chain on a finite state space is irreducible and reversible, then it will have a unique stationary distribution, and no matter in which state you start the chain, it will eventually converge to this stationary distribution.

5.7

5.8

We make use of this fact by looking at our target distribution, such as the distribution for \mathbf{X} in an exponential random graph model, as the stationary distribution of a Markov chain.

This Markov chain lives on graphs, and moves are adding or deleting edges, as well as adding types or reducing types.

Finding suitable Markov chains is an active area of research.

The `ergm` package has MCMC implemented for parameter estimation. We need to be aware that there is no guarantee that the Markov chain has reached its stationary distribution. Also, if the stationary distribution is not unique, then the results can be misleading. Unfortunately in exponential random graph models it is known that unless the model is close to a Bernoulli random graph, often the likelihood is very flat and convergence takes a very long time.

5.9

5.3 Assessing the model fit

Suppose that we have estimated our parameters in our model of interest. We can now use this model to see whether it does actually fit the data.

To that purpose we study the (asymptotic) distributions of our summary statistics *vertex degree*, *clustering coefficient*, and *shortest path length*. Then we see whether our observed values are plausible under the estimated model. Note that the model should reflect the research question; typically no model will capture all aspects of the observed network.

Often, secretly we would like to find that the observed value of the statistics are not plausible under the null model! Because then we can reject, say, the simple random graph model, and conclude that something more complicated is going on.

5.10

Example: Florentine family data

In the Florentine family data, we observe 16 vertices, 20 edges, an average vertex degree of 2.5, and 3 triangles. We want to assess the null hypothesis that the data come from a Bernoulli random graph.

Let us assume that the null hypothesis is true. Then we estimate

$$\hat{p} = \frac{20}{\binom{16}{2}} = \frac{20 \times 2}{16 \times 15} = \frac{1}{6}.$$

In a Bernoulli random graph T , the number of triangles, is approximately Poisson distributed with parameter $\binom{n}{3}p^3$; here $T \approx Po(2.592593)$.

5.11

The probability under the null hypothesis that $T \geq 3$ would then be approximately

$$\mathbb{P}(Z \geq 3) \approx 0.48,$$

so there is not enough evidence to reject the null hypothesis.

How good is the Poisson approximation in this case? The bound on the total variation distance which we derived using the Stein-Chen method gives a value of about 1.55, making the approximate p -value rather unreliable.

5.12

Quantile-quantile plots

We could also assess the fit visually. A much used plot in Statistics a *quantile-quantile plot*.

The *quantiles* of a distribution are its "percent points"; for example the 0.5 quantile is the 50 % point, i.e. the median. Mathematically, the (*sample*) *quantiles* q_α , are defined for $0 \leq \alpha \leq 1$ so that a proportion of at least α of the data are less or equal to q_α and a proportion of at least $1 - \alpha$ is greater or equal to q_α . There are many (at least 8) definitions of q_α if αn is not an integer.

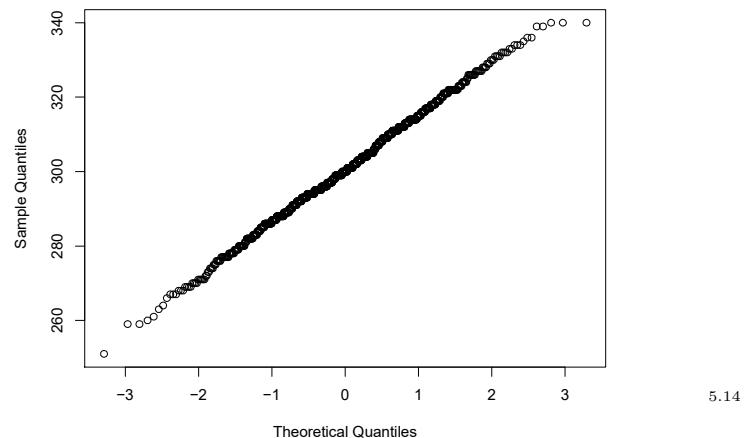
We plot the quantiles of our observed (empirical) distribution against the quantiles of our hypothesised (null) distribution; if the two distributions agree, then the plot should result in a roughly diagonal line.

5.13

Example: degrees in a dense Bernoulli random graph

Simulate a Bernoulli random graph on 1000 vertices with edge probability $p = 0.3$. The normal q-q plot for the degrees looks satisfactory.

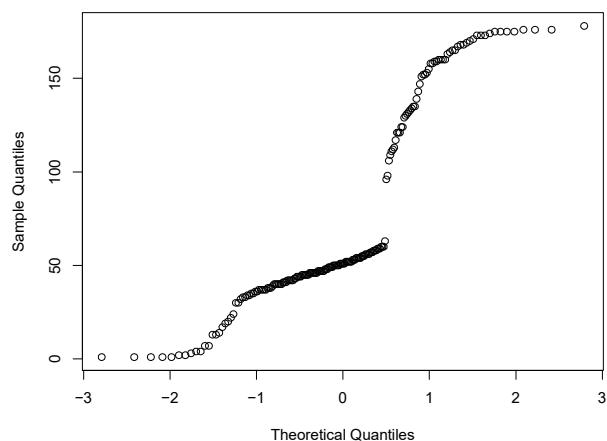
Quantile-quantile plot for degrees of $G(1000, 0.3)$



Example: degrees in a trade network

For the world trade network of the year 2000, the normal q-q plot for the degrees does not look satisfactory.

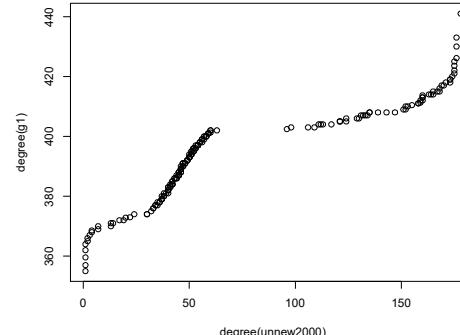
Quantile-quantile plot for degrees of trade in 2000



5.15

We can also use a quantile-quantile plot for two sets of simulated data, or for one set of simulated data and one set of observed data. The interpretation is always the same: if the data come from the same distribution, then we should see a diagonal line; otherwise not. Here we compare the degrees of the trade network in the year 2000 to the degrees of a Bernoulli random graph with same density (0.393985) - not a good fit.

Trade network degree quantiles vs degree quantiles of $G(1000, 0.3)$



5.16

6 Nonparametric methods

What if we do not have a suitable test statistic for which we know the distribution? We need some handle on the distribution, so here we assume that we can simulate random samples from our null distribution. There are a number of methods available.

6.1

1. Observe the actual value t^* for T_0 , calculated from the data
2. Simulate a random sample of size n from F_0
3. Order the set $\{t^*, t_{01}, \dots, t_{0n}\}$
4. Reject H_0 if the rank of t^* in this set (in decreasing order) is $\leq m$.

The basis of this test is that, under H_0 , the random variable T^* has the same distribution as the remainder of the set and so, by symmetry,

$$\mathbb{P}(t^* \text{ is among the largest } m \text{ values}) = \frac{m}{n+1}.$$

The procedure is exact however small n might be. However, increasing n increases the power of the test. A reasonable rule is to choose n such that $m \geq 5$. Note that we will need more simulations to test at smaller values of α .

In discrete data, we will typically observe ties. We can break ties randomly, then the above procedure will still be valid.

6.1 Monte-Carlo tests

The Monte Carlo test, attributed to Dwass (1957) and Barnard (1963), is an exact procedure of virtually universal application and correspondingly widely used.

Suppose that we would like to base our test on the statistic T_0 . We only need to be able to simulate a random sample T_{01}, T_{02}, \dots from the distribution, call it F_0 , determined by the null hypothesis. We assume that F_0 is continuous, and, without loss of generality, that we reject the null hypothesis H_0 for large values of T_0 . Then, provided that $\alpha = \frac{m}{n+1}$ is rational, we can proceed as follows.

6.2

Example: Florentine family data

In the Florentine family data, we observe 16 vertices, 20 edges, an average vertex degree of 2.5, and 3 triangles. We want to assess the null hypothesis that the data come from a Bernoulli random graph on 16 vertices with $p = \frac{1}{6}$.

We generate 99 Bernoulli random graphs on 16 vertices, with $p = \frac{1}{6}$. In my simulation 33 of them have at least 3 triangles. Hence the Monte Carlo p-value is 0.34, and there is no reason to reject the null hypothesis.

6.3

6.4

For random graphs, Monte Carlo tests often use shuffling edges with the number of edges fixed, or fixing the vertex degree distribution, or fixing some other summary.

Suppose we want to see whether our observed clustering coefficient is "unusual" for the type of network we would like to consider. Then we may draw many networks uniformly at random from all networks having the same vertex degree sequence, say. We count how often a clustering coefficient at least as extreme as ours occurs, and we use that to test the hypothesis.

In practice these types of test are the most used tests in network analysis. They are called *conditional uniform graph tests*.

Some caveats:

In Bernoulli random graphs, the number of edges asymptotically determines the number of triangles when the number of edges is moderately large. Thus conditioning on the number of edges (or the vertex degrees, which determine the number of edges) gives degenerate results.

More generally, we have seen that vertex degrees and clustering coefficient (and other subgraph counts) are not independent, nor are they independent of the shortest path length. By fixing one summary we may not know exactly what we are testing against.

6.5

"Drawing uniformly at random"

"Drawing uniformly at random" from complex networks is not as easy as it sounds. Algorithms may not explore the whole data set. Even in Bernoulli random graphs, when the expected number of edges is moderate, so that a normal approximation would hold for the number of edges, then, asymptotically, the number of 2-stars and the number of triangles is already completely determined by the number of edges!

"Drawing uniformly at random", conditional on some summaries being fixed, is related to sampling from exponential random graphs. We have seen already that in exponential random graphs there may be more than one stationary distribution for the Markov chain Monte Carlo algorithm; this algorithm is similar to the one used for drawing at random, and so we may have to expect similar phenomena.

6.7

6.6

6.8

6.2 Fitting scale-free networks

Barabasi and Albert introduced networks such that the distribution of the degree of a randomly picked vertex V is of the type

$$\mathbb{P}(d(V) = k) \sim Ck^{-\gamma}$$

for $k \rightarrow \infty$. Such behaviour is called *power-law behaviour*, the constant γ is called the *power-law exponent*. The networks are also called *scale-free*, because of the following property.

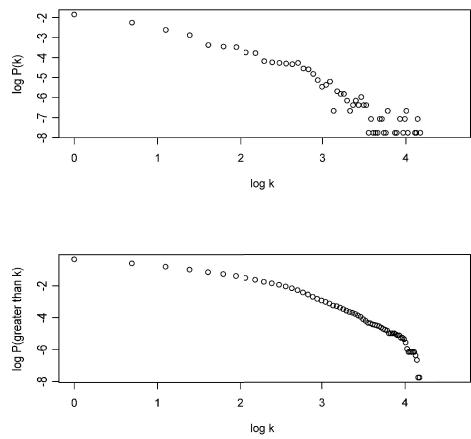


Figure: Yeast data log-log plots

6.9

If $\alpha > 0$ is a constant, then

$$\mathbb{P}(d(V) = ak) \sim C(ak)^{-\gamma} \sim C'k^{-\gamma},$$

where C' is just a new constant. That is, scaling the argument in the distribution changes the constant of proportionality as a function of the scale change, but preserves the shape of the distribution itself. If we take logarithms on both sides:

$$\begin{aligned}\log \mathbb{P}(d(V) = k) &\sim \log C - \gamma \log k \\ \log \mathbb{P}(d(V) = ak) &\sim \log C - \gamma \log \alpha - \gamma \log k;\end{aligned}$$

scaling the argument results in a linear shift of the log probabilities only. This equation also leads to the suggestion to plot the $\log \text{relfreq}(d(v) = \alpha k)$ of the empirical relative degree frequencies against $\log k$. Such a plot is called a *log-log plot*. If the model is correct, then we should see a straight line; the slope would be our estimate of γ .

6.10

These plots have a lot of noise in the tails. As an alternative, Newman (2005) suggests to plot the log of the empirical cumulative distribution function instead, or, equivalently, our estimate for

$$\log \mathbb{P}(d(V) \geq k).$$

If the model is correct, then one can calculate that

$$\log \mathbb{P}(d(V) \geq k) \sim C'' - (\gamma - 1) \log k.$$

Thus a log-log plot should again give a straight line, but with a shallower slope. The tails are somewhat less noisy in this plot.

In both cases, the slope is estimated by least-squares regression: for our observations, $y(k)$ (which could be log probabilities or log cumulative probabilities, for example) plotted against $x(k)$, which is $\log k$ in our case, we find the line $a + bx(k)$ which minimises

$$\sum (y(k) - a - bx(k))^2.$$

6.11

6.12

7 Inferring vertices and edges in networks

In the statistical analysis of networks we are often interested in inferring "local" properties, such as the existence of an edge from the position of the potential edge in the network.

7.1

Some sets

If $\mathbf{X} = (X_{i,j})_{i,j=1,\dots,n}$ is the (symmetric) adjacency matrix of our random graph (with \mathbf{x} its realisation) then we can view it as an ordered set $\{X_{i,j}, i, j = 1, \dots, n\}$. We define three related sets in an intuitive way:

$$\begin{aligned}\mathbf{X}_{i,j}^+ &= \{X_{k,l}, k, l = 1, \dots, n, \{k, l\} \neq \{i, j\}; X_{i,j} = X_{j,i} = 1\} \\ \mathbf{X}_{i,j}^- &= \{X_{k,l}, k, l = 1, \dots, n, \{k, l\} \neq \{i, j\}; X_{i,j} = X_{j,i} = 0\} \\ \mathbf{X}_{i,j}^c &= \{X_{k,l}, k, l = 1, \dots, n, \{k, l\} \neq \{i, j\}\}.\end{aligned}$$

Thus $\mathbf{X}_{i,j}^+$ is the adjacency matrix but with the (i, j) -entry and the (j, i) -entry replaced by the value 1.

Similarly $\mathbf{X}_{i,j}^-$ is the adjacency matrix but with the (i, j) -entry and the (j, i) -entry replaced by the value 0.

Finally $\mathbf{X}_{i,j}^c$ is the set of edge indicators except the indicators for edge (i, j) and (j, i) .

7.3

Exponential random graphs

Exponential random graph models are particularly well fitted to answer such questions. Recall:

Exponential random graph (p^*) models model the whole adjacency matrix \mathbf{X} of a graph simultaneously, making it easy to incorporate dependence.

The general form of the model is

$$\mathbb{P}(\mathbf{X} = \mathbf{x}) = \frac{1}{\kappa} \exp \left\{ \theta^T \mathbf{z}(\mathbf{x}) \right\},$$

where θ is a vector of model parameters and $\mathbf{z}(\mathbf{x})$ is a vector of network statistics; $\kappa = \kappa(\theta)$ is a normalising quantity so that the probabilities sum to 1.

7.2

The odds ratio is

$$\begin{aligned}\frac{\mathbb{P}(X_{i,j} = 1; \mathbf{X}_{i,j}^c = \mathbf{x}_{i,j}^c)}{\mathbb{P}(X_{i,j} = 0; \mathbf{X}_{i,j}^c = \mathbf{x}_{i,j}^c)} &= \frac{\exp \left\{ \theta^T \mathbf{z}(\mathbf{x}_{i,j}^+) \right\}}{\exp \left\{ \theta^T \mathbf{z}(\mathbf{x}_{i,j}^-) \right\}} \\ &= \exp \left\{ \theta^T \mathbf{z}(\mathbf{x}_{i,j}^+) - \theta^T \mathbf{z}(\mathbf{x}_{i,j}^-) \right\}.\end{aligned}$$

Taking logarithms,

$$\log \left(\frac{\mathbb{P}(X_{i,j} = 1; \mathbf{X}_{i,j}^c = \mathbf{x}_{i,j}^c)}{\mathbb{P}(X_{i,j} = 0; \mathbf{X}_{i,j}^c = \mathbf{x}_{i,j}^c)} \right) = \theta^T (\mathbf{z}(\mathbf{x}_{i,j}^+) - \mathbf{z}(\mathbf{x}_{i,j}^-)).$$

Thus the odds ratio of an edge being present rather than absent, given the rest of the network, has an easy form which does not involve the normalising constant κ .

7.4

Florentine families revisited

In his Florentine family data set Padgett also estimated wealth from tax records from 1427, and recorded them in 1000 Florins. We can fit an exponential random graph model to the data using MCMC estimation. For the model that the logit depends on the number of edges, the number of 2-stars, the number of triangles, and the absolute wealth difference, we obtain (using the `sna` and the `ergm` packages in R)

	Estimate	Std. error	p-value
edges	-0.809	0.689	0.24
2-stars	-0.226	0.354	0.52
triangles	0.3999	0.820	0.62
absolute wealth difference	0.020	0.009	0.03

The difference in wealth is significant as explanatory variable for the odds of a marriage relation. An increase of wealth difference by 1 unit increases the odds of a marriage by a factor of about $e^{0.02} \approx 1.05$.

7.5

Stochastic blockmodel with known groups

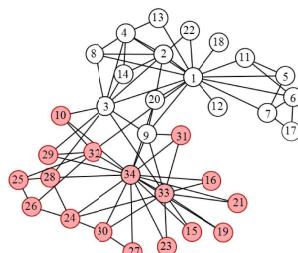
Suppose that we know the group memberships in a stochastic blockmodel and we would like to infer whether or not there is an edge between two nodes. Then we can

- (1) fit a stochastic blockmodel to the data with the particular edge variable excluded;
- (2) assess the odds of an edge to be present from the edge probability estimates.

7.6

Example: Zachary's karate club

Zachary's Karate club network is a friendship network of 34 members of a karate club at a US university in the 1970s. The club split into two different factions as a result of an internal dispute, and the members of each faction are known.



7.7

In Zachary's paper, the adjacency matrix of relationships in the club is ambiguous – while the network is undirected, there is an edge reported from 23 to 34, but not from 34 to 23.

In the graphical representation of the club in that paper, the edge between 23 and 34 is absent. Yet, in the standard network databases the edge is assumed to be present.

How likely is it that the edge was indeed present in the data?

7.8

To answer this question, we assume that there is no observation of presence or absence of an edge between 23 and 34 available. This new network still has 34 vertices but only 77 edges.

There are 17 vertices in group 2, and 31 edges between vertices in Group 2 in the reduced data set. This gives a maximum likelihood estimate for the edge probability in Group 2 as

$$\hat{p} = \frac{31}{\binom{17}{2} - 1} = \frac{31}{135} = 0.2298$$

(with standard error $\sqrt{\hat{p}(1 - \hat{p})/135} = 0.0362$), and the estimated odds for an edge between Vertex 23 and Vertex 34 are

$$0.2298/(1 - 0.2298) \approx 0.2981.$$

Hence the evidence points against an edge between Vertex 23 and Vertex 34.

8 Modules and communities

Finding modules, or communities, which make up the network, has been of interest not only for social networks.

Statistically speaking, we would like to apply a *clustering method* to find out more about the structure of the network.

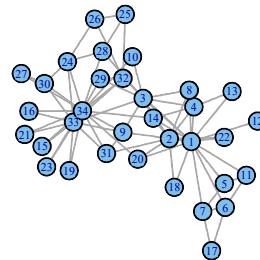
There is an abundance of clustering methods available.

7.9

8.1

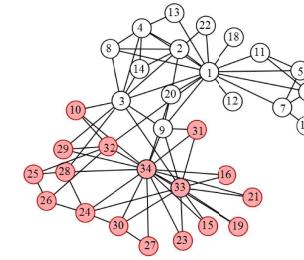
The karate club example

Recall the Zachary's Karate club network; a friendship network of 34 members of a karate club at a US university in the 1970s.



The karate club example

For Zachary's Karate club friendship network: the club split into two different factions as a result of an internal dispute, and the members of each faction are known.



8.2

8.3

8.1 Modularity

As a guidance to how many communities a network should be split into, often *modularity* is used.

Assume that the vertex set is partitioned into g groups, say, and for vertex v , c_v is the unique group that v is assigned to. Then the modularity is defined as

$$Q = \frac{1}{2m} \sum_{u,v} \left(a_{u,v} - \frac{d(u)d(v)}{2m} \right) \mathbf{1}(c_u = c_v),$$

where m is the number of edges, $A = (a_{u,v})$ is the adjacency matrix, and $d(v)$ is the degree of vertex v .

We note that $\frac{d(u)d(v)}{2m}$ is roughly the probability that $a_{u,v} = 1$ in the configuration model. In this sense the modularity compares the observed network to a configuration model.

The Newman-Girvan algorithm

A much used algorithmic approach is the algorithm by Newman and Girvan, see for example *Newman and Girvan (2004)*.

Recall that the betweenness of an edge is defined to be the number of shortest paths between vertex pairs that run along the edge in question, summed over all vertex pairs.

The algorithm of Girvan and Newman then involves simply calculating the betweenness of all edges in the network and removing the one with highest betweenness, and repeating this process until no edges remain.

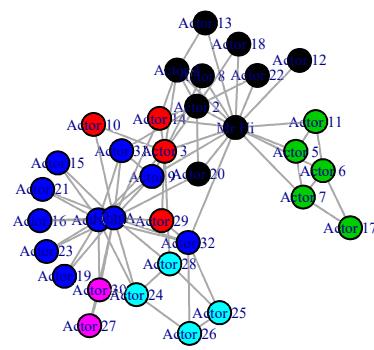
If two or more edges tie for highest betweenness then one can either choose one at random to remove, or simultaneously remove all of them.

8.4

8.5

The karate club example

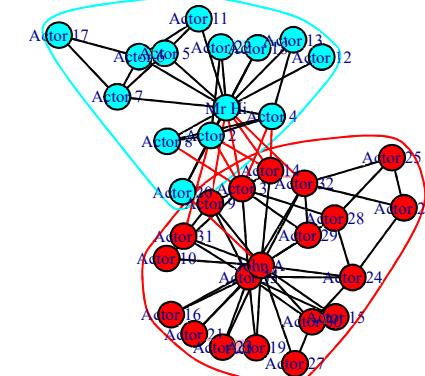
Newman-Girvan on the karate club example:



8.6

The karate club example

Newman-Girvan on the karate club example forced to have exactly 2 groups:



8.7

The karate club example



Credit: Aaron Clauset

8.8

The degree-corrected stochastic blockmodel

The degree-corrected stochastic blockmodel by *Karrer and Newman (2011)* assumes that each vertex u has a “degree parameter” θ_u associated with it so that for every class k ,

$$\sum_u \theta_u \mathbf{1}(c_u = k) = 1$$

and

$$\mathbb{P}(u \sim v) = \theta_u \theta_v p_{c_i, c_j}.$$

Here (c_1, \dots, c_n) is the class assignment vector for the vertices.

Karrer and Newman actually use a Poisson version of this model, the number of edges between two vertices u and v follows a Poisson distribution with parameter $\mathbb{P}(u \sim v)$.

8.10

8.2 The stochastic blockmodel

Instead of modularity, the stochastic blockmodel can be viewed as reflecting community structure, with vertices of the same type interpreted as communities.

Recall: The stochastic blockmodel assumes that vertices are of L different types. Then edges are constructed independently, such that two vertices u, v of types i and j are connected with probability

$$p_{i,j} = \mathbb{P}(u \sim v | u \text{ is of type } i, v \text{ is of type } j).$$

Here we write $u \sim v$ to indicate that there is an undirected edge between u and v .

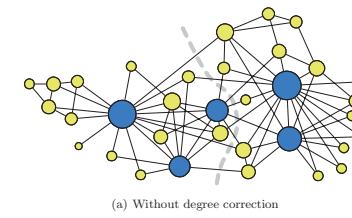
In particular one can fit a stochastic blockmodel to a network and use the output for community detection.

Unfortunately this approach usually does not work well in real networks, as it tends to cluster vertices according to their degrees. In contrast, the degree-corrected stochastic blockmodel tends to perform much better.

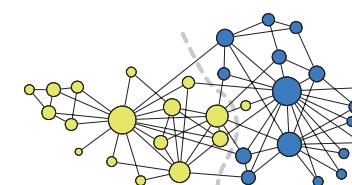
8.9

The karate club example

The stochastic blockmodel, with and without degree correction (from Karrer and Newman (2011)). The size of a vertex is proportional to its degree.



(a) Without degree correction



(b) With degree-correction

8.11

9 Some further topics

Many networks are directed.

Edges may be weighted and the interpretation of weights may not be straightforward. For example in mobile phone networks, short phone calls indicate a different type of interaction than long phone calls.

Some networks are *signed*, where the positive/negative edge weights may encode similarity/dissimilarity information.

Communities may be overlapping.

Networks may be *dynamic*; our data are snapshots in time.

Edges in networks may be of different types. More generally, a set of related networks may be available on essentially the same set of vertices. *Multi-layer network analysis* is being developed for this situation.

We look at two of these issues in a bit more depth.

9.1

9.1 Longitudinal network data

Longitudinal studies may provide a collection of networks.

An example is treated in *L. Mercken et al., Smoking-based selection and influence in gender-segregated friendship networks: a social network analysis of adolescent smoking; Addiction 105 (2010), 1280–1289.*

Four measurements in nine junior high schools in Finland were taken; 1163 adolescents (average age 13.6 years), 605 males and 558 females.

Self-administered questionnaires were distributed initially, and similar questionnaires during follow up (12 month, 24 month, 30 month).

9.2

Findings

The questions concerned friendship ties, smoking behaviour, parental and sibling smoking behaviour, school achievement, and alcohol consumption.

For each wave, and for each school, a female and a male network were constructed; cross-gender friendships were excluded.

For each network, a so-called *dynamic actor-based model* was estimated; this models how individuals add or delete friendship ties from one time step to the next.

9.3

There is some evidence that friendship selection is influenced by smoking.

Only for females is there evidence that they are influenced to smoke by their peer group.

Both males and females were influenced by parental smoking behaviour.

A prevention campaign which targets resistance to peer pressure may be more successful in adolescent girls than boys.

9.4

9.2 Patterns in temporal networks

Time series analysis of networks is a very active area of research. Patterns over time can be used to explain observed outcomes.

As an example, *Estrades, Carmen, Manuel Flores, and Guillermo Lezama. The role of export restrictions in agricultural trade. IATRC Commissioned Paper 20, April 2017* consider trade networks and conflict for the period 1962-2000. Conflict between countries is measured by the initiation of militarised interstate disputes.

The aim is to model the binary outcome which is 1 for the first year a militarised dispute takes place between a pair of countries. Subsequent years are discarded from the analysis.

They model edges as independent, using endogenous network features as well as exogenous features such as trade measures and measures which describe the willingness and ability of countries to engage in militarised conflict. They ignore the potential dependence between edges and other network features.

9.5

Findings

They find that the effects of trade vary by commodity. The variation is largely explained by the strategic importance of commodities and the ease with which they can be substituted.

Trade of energy, metals and electronics creates vulnerability between countries, which is associated with increased conflict.

Trade of chemicals and arms is associated with less conflict.

These are associations; there is no claim of a causal relationship!

Summary

These lectures provided a glimpse into the world of network analysis.

Many procedures are heuristic; yet understanding the variability of outcomes is essential to draw statistical conclusions.

There is a Networks Club which meets Tuesdays 12-1 during term time, in the Mathematical Institute.

New research questions may require new statistics!

9.7

9.8