# Lecture 10-11: Network analysis

## Foundations of Data Science:
## Algorithms and Mathematical Foundations

Mihai Cucuringu
mihai.cucuringu@stats.ox.ac.uk

CDT in Mathematics of Random System
University of Oxford

25 September, 2023

A subset of the slides & topics are from the *Probability and Statistics for Network Analysis* course in the Department of Statistics, jointly taught with Gesine Reinert in previous years.

# 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 this lecture does not address graphical models.
- Often one would think of a network as a connected graph, but not always.
- In this lecture we shall use *network* and *graph* interchangeably.
- Here are some of the most well known examples of networks (graphs).

# Marriage network

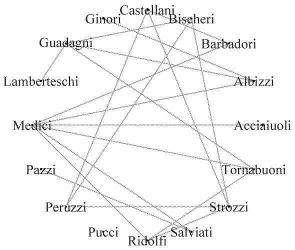



Figure: Marriage relations between Florentine families; two different graphical representations

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 15$^{th}$ century Florence, Italy.

# Zachary's karate club

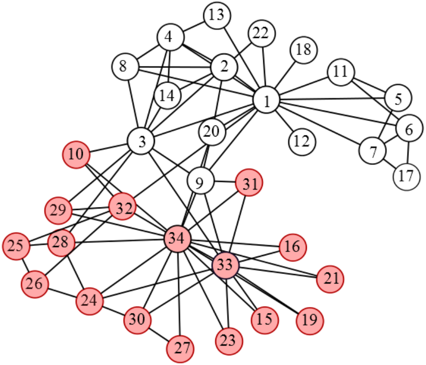



Figure: A friendship network: Zachary's karate club

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.

# Protein interaction networks

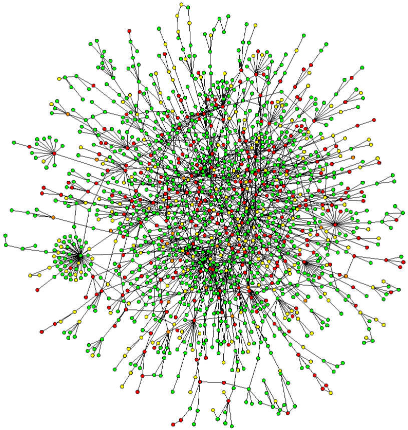


Figure: Yeast protein-protein interactions

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

# Political Blogs

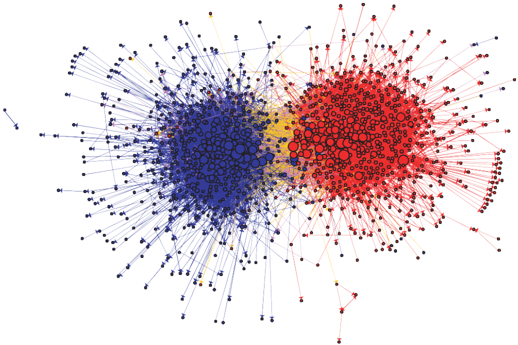
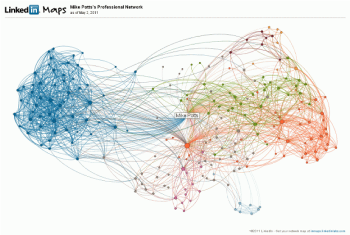


Figure: Political blog data from Adamic and Glance (2005).

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.

# Social networks

- links denote a social interaction
- networks of acquaintances
- collaboration networks
- actor networks
- co-authorship networks
- director networks
- e-mail networks
- phone-call networks (time, duration,location)
- IM networks
- sexual networks



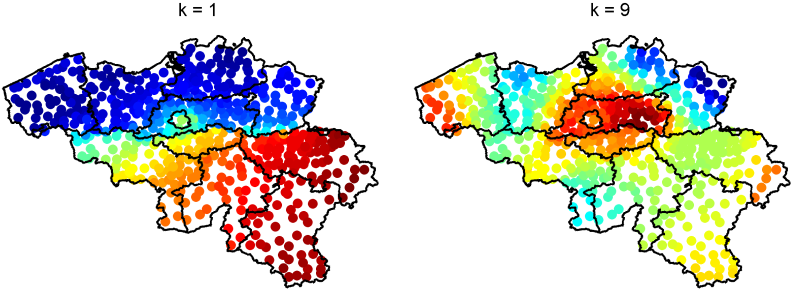

Source: Frieze, Gionis, Tsourakakis, Algorithmic Techniques for Modeling and Mining Large Graphs (KDD 2013)

# Belgium Mobile Phone Network

- Migration statistics within and across 3000+ counties in US
- Eigenvectors of the associated graph Laplacian (we will study later) captures interesting patterns

$$W_{i,j} = \frac{N_{i,j}^2}{P_i \cdot P_j}$$

- $N_{i,j}$: number of people migrating btw county i and county j
- $P_i$: population of county i



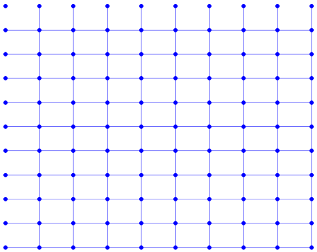

M. Cucuringu, V. Blondel, P. Van Dooren, Extracting spatial information from networks with low-order eigenvectors, Physical Review E 87, 032803 (2013)

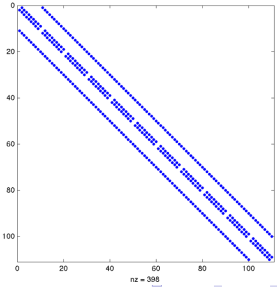
# How to examine a graph

- Take some measurements of it (count # of vertices and edges)
- If not connected, break it into connected components
- Examine the distribution of the degrees of the vertices
- Draw the graph and visualize it

Grid Graph

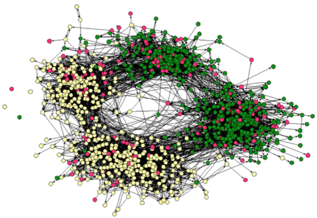
- Very easy to visualize
- Not the case for most real world graphs!
- Impossible to make nice drawings of most graphs

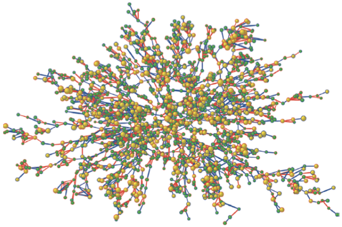




# Analysis of graph data sets in the past

- 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
- visual inspection can reveal a lot of information





# Analysis of graph data sets now

In contrast, starting at around 1997, statistical physicists have turned their attention to large-scale properties of networks.

- unless the network is very small it appears like a hairball, and is difficult to analyse by just looking at it
- more and larger networks appear (byproducts of technological advancement)
  - e.g., internet, web
  - result of our ability to collect more, better-quality, and more complex data
- networks of thousands, millions, or billions of nodes
- need to develop more sophisticated & scalable tools

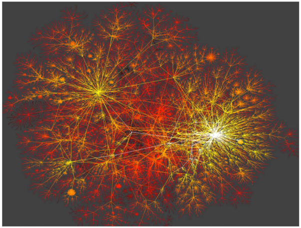




Figure: The Internet graph.

# Network Science

Emerging area with an exponential growth of its literature over the past decade.

The world is full with networks - what do we do with them?

- understand their topology and measure their properties
- study their evolution and dynamics
- create realistic models
- build efficient algorithms that can leverage the network structure
- perform various tasks on the network (e.g., ranking, prediction)

**Question**: how can we leverage structural findings in a network for prediction? (Also allows for a fair comparison of different methods.)

# Further research questions

- 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?

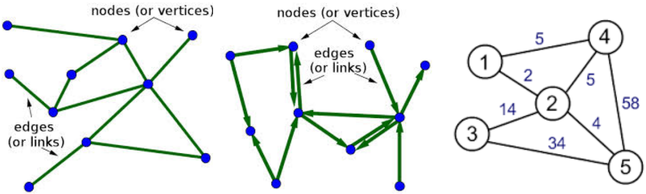
From a statistical viewpoint, questions include

- 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.

# 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).
  - notation $|S|$ denotes the number of elements in the set $S$. Then $|\mathcal{V}|$ is the number of vertices, and $|\mathcal{E}|$ 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* (*digraph*) is a graph where all edges are directed.
- If both endpoints of an edge are the same, the edge is a *loop*
- Simple graphs: without self-loops and multiple edges




(a) Undirected (b) Directed (c) Weighted

- Two vertices are called *adjacent* if they are joined by an edge. A graph can be described by its $|V| \times |V|$ 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

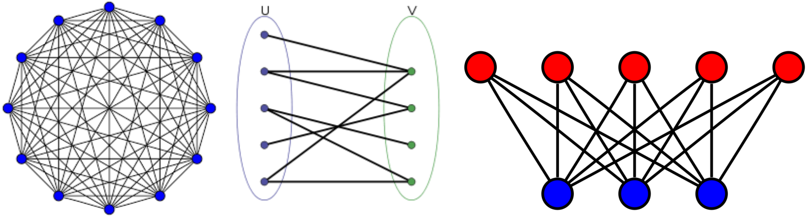
$$a^{(2)}(u, v) = \sum_{w \in \mathcal{V}} a_{u,w} a_{w,v}.$$

- 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.

# Example: Adjacency matrix for Florentine marriages

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

- 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) Complete (b) Bipartite (c) Complete bipartite

- 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 $\mathcal{W}$.

  Adjacency matrix $A$ of the form:

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

# Graph Theory

- graph theory started in the 18th century, with Leonhard Euler
- the problem of Königsberg bridges (1736)

Figure: Königsberg bridges (map from Euler's time)

- "Graph Theory is the new calculus", Daniel A. Spielman (2007) [2023 Breakthrough Prize in Mathematics for "contributions to computer science & mathematics, including spectral graph theory"]

- Algorithmic graph theory
- Structural graph theory
- Algebraic graph theory
  - **spectral graph theory** - explores connections connections to linear algebra (spectrum of adjacency matrix)
  - using group theory (study symmetric properties of a graph)
  - studying graph invariants (eg, the chromatic polynomial counting the number of its proper vertex colorings)

# 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.

# Degrees and the degree distribution

- The degree $d(v)$ of a vertex $v$ is the number of edges which involve $v$ as an endpoint.
- (For the case of **weighted** graphs, the strength of a node is the sum of the weights of the incident edges)
- 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

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

# Degrees and the degree distribution

- The degree sequence of a given network on a vertex set $\mathcal{V}$ with $n$ elements is the unordered $n$-element set of degrees $\{d(v), v \in \mathcal{V}\}$.
- Example: the degree sequence of a triangle on three vertices is $\{2, 2, 2\}$.
- The degree distribution $(d_0, d_1, d_2, , \ldots)$ 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
  - out-degree as the number of edges that go out from that vertex.

# Example: a scientific collaboration network

- Start with a bipartite graph with authors on one side and papers on the other.
- Edges link papers to their authors.

- Back out a network on just the authors by connecting each pair of authors who have co-authored a paper.
- This naturally leads to a weighted graph
- W(i,j) = the number of papers authors i and j have written together
- Consider the dblp co-authorship graph (SNAP library at Stanford)

# Degrees in a scientific collaboration network

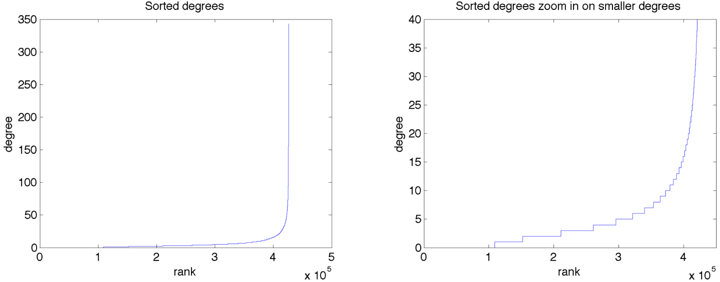



Figure: Sorted degree sequence (zoom-ed in version, on the right)

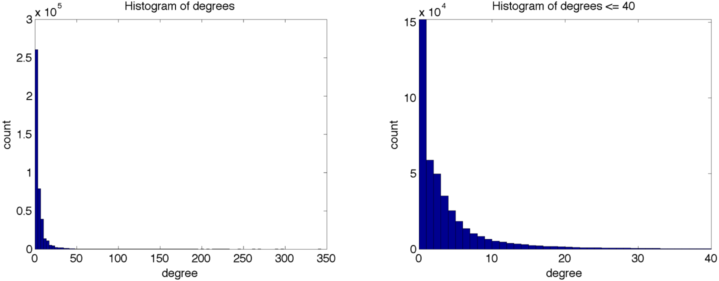



Figure: Histogram of degrees (zoom-ed in version, on the right)

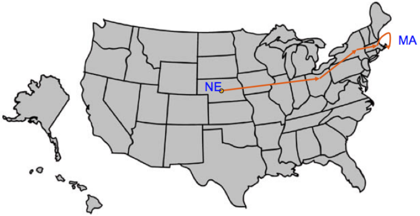
# Milgram's experiment and the small world effect

In 1967, the American sociologist Milgram reported a series of experiments of the following type

- He chose a random person X in Nebraska
- Asked X to deliver a letter to a random person Y (stock broker) in Massachusetts, Lashawn
- Told X the name, address and occupation of Y
- Instructed X to only send letter to people he knows on a first-name basis

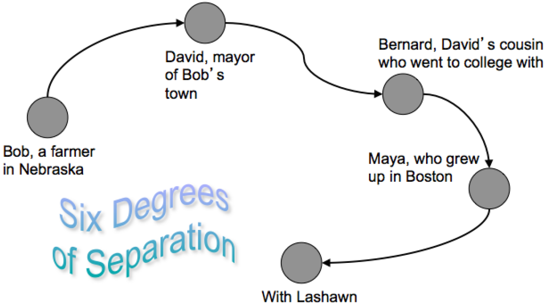
Milgram kept track of how many intermediaries were required until the letters arrived.

- Stanley Milgram (1933-1984) "The man who shocked the world"
- obedience to authority (1963)
- small-World experiment (1967)



# Milgram's experiment and the *small world effect*

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`.

# 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

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

The local clustering coefficient describes how "locally dense" a graph is.

# 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.

# The expected clustering coefficient

- 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.
- Typically, by *clustering coefficient* we mean the global clustering coefficient

# 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, $(v_0, e_1, v_1, e_2, \ldots, v_{n-1}, e_n, v_n)$ such that the endpoints of $e_i$ are $v_{i-1}$ and $v_i$, for $i = 1, \ldots, n$.
- 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.

# Connectivity

- A graph is called *connected* if there is a walk between any pair of vertices in the graph, otherwise it is called *disconnected*.
- The number of connected components is the size of the smallest partition of the nodes into connected subgraphs.
- A graph $\mathcal{G}$ is said to be $k$-connected (or $k$-vertex connected) if there does not exist a set of $k-1$ vertices whose removal disconnects the graph.
- (Notion extends to directed graphs (strong/weak connectivity). Also, analogous statement can be made for $k$-edge connectivity.)
- In a connected graph, the *average shortest path length* is given by

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

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

# Cliques, independent sets, colorings

- A clique in a graph $G$ is a subset $S$ of its nodes such that the subgraph corresponding to it is complete.
- Equivalently, if $S$ is a clique if all pairs of vertices in S share an edge.
- The clique number $w(G)$ is the size of the largest clique of $G$.

---

- An independence set $\mathcal{I}$ of a graph $G$ is a subset $S$ of its nodes such that no two nodes in $S$ share an edge
- Equivalently, $\mathcal{I}$ is a clique in the complement graph $G^c := (V, E^c)$.
- The independence number of $G$ is the clique number of $G^c$.

---

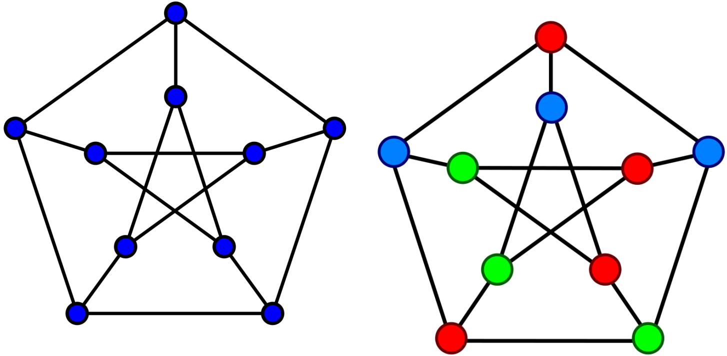
- A vertex coloring of $G$ is a labeling of the graph's vertices with colors such that no two vertices sharing the same edge have the same color.
- The smallest number of colors needed to color a graph $G$ is called its chromatic number, often denoted $\chi(G)$.

---

- What is the relationship between $w(G)$ and $\chi(G)$?

# The Petersen Graph

Often used as an example or counter-example in graph theory.



The Petersen graph has

- a clique number of 2
- an independence number of 4
- chromatic number 3.

# Small graphs and motifs

- In addition to considering 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.

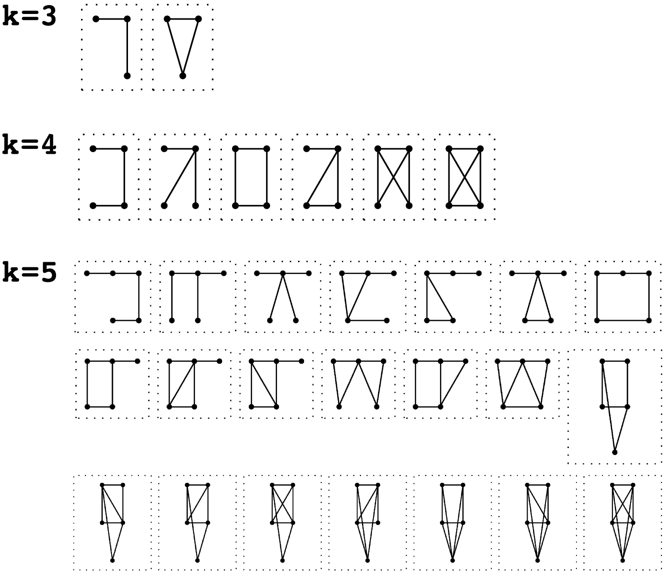



Figure: Some small graphs (motifs)

# Directed Motifs

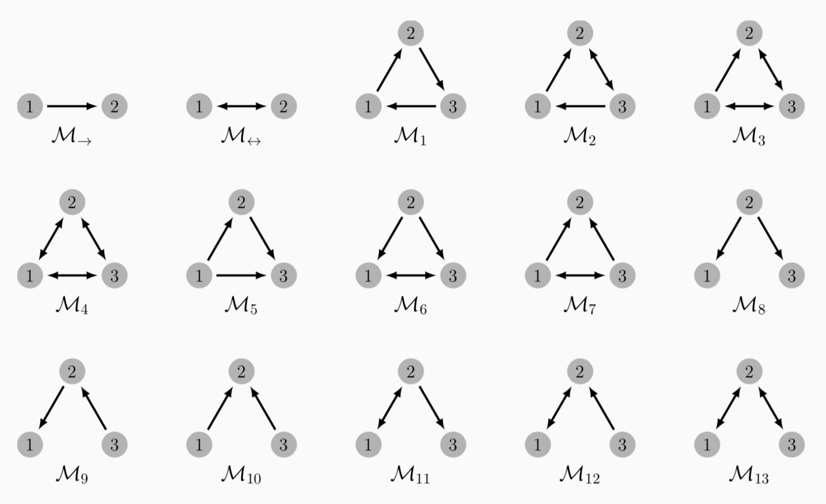



Figure: All directed motifs on at most three vertices.

Benson, A. R., Gleich, D. F., & Leskovec, J. (2016). Higher-order organization of complex networks. Science, 353(6295), 163-166.

# 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, we rely on mathematical models
- Models also allow us to give **mathematical/statistical guarantees** on the performance of algorithms for certain tasks (eg. clustering).

# Bernoulli (Erdös-Renyi) random graphs

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

- The expected number of edges is

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

- Degree distribution
  - each vertex has $n-1$ potential neighbours,
  - each of these $n-1$ edges is present with probability $p$
  - $\Rightarrow$ the degree of a randomly chosen vertex is $Bin(n-1, p)$-distributed

# Bernoulli (Erdös-Renyi) random graphs

- The expected number of triangles in the graph is

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

- Exercise: The expected clustering coefficient is $p$

- In an Erdös-Renyi random graph, your friends are no more likely to be friends themselves than would be a two complete strangers [not realistic].

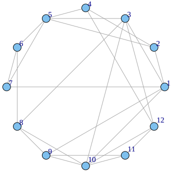
# The Small World phenomenon

- 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!

# 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 $p$ this edge is reconnected to a vertex chosen uniformly at random over the ring, with duplicate edges excludes; otherwise the edge is left in place.
- Repeat “rewiring” by moving clockwise around the ring.
- Next, consider edges that connect vertices to their $2^{nd}$-nearest neighbours, again clockwise, and repeat the rewiring process.
- With $nk$ edges in the network the process stops after $k$ laps.
- The number of edges remains $nk$.

# Small world model

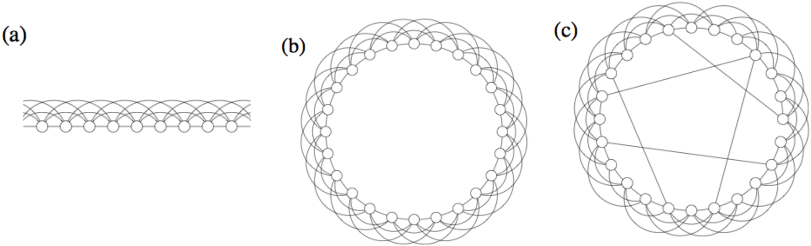


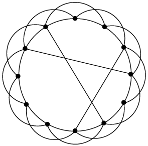

Figure: (a) A one-dimensional lattice with each node connected to its $c$ nearest neighbors, where in this case $c = 6$. (b) The same lattice with periodic boundary conditions, so that the system becomes a ring. (c) The Watts-Strogatz model is created by rewiring a small fraction of the links (in this case five of them) to new sites chosen at random. *Models of the Small World*, A Review, M. E. J. Newman, (2000)

Watts & Strogatz, Collective dynamics of “small-world" networks, Nature 393, 440-442 (1998). Gogole Scholar: 25,307 citations (2015); 38,335 (2019); 43,530 (2020); 47,100 (2021); 49,950 (2022); 52130 (2023)

# The Newman-Moore-Watts model

- Most used version of the Watts-Strogatz model is the *Newman-Moore-Watts model*, also known as the great circle model, *Ball, Mollison and Scalia-Tomba 1997*
  - 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 $2k + Bin(n - 2k - 1, p)$.



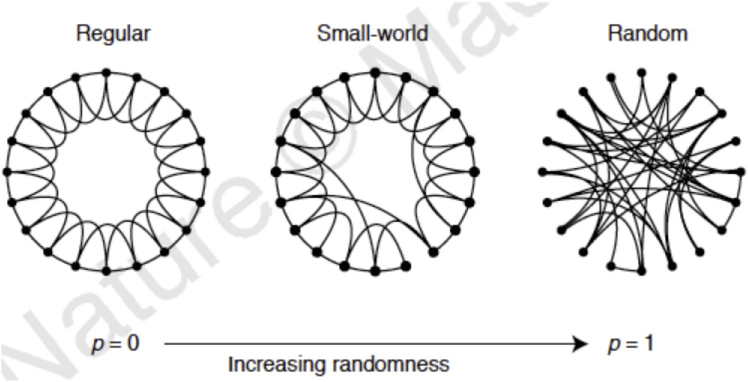



Figure: Original small-world model proposed by Watts & Strogatz, Collective dynamics of 'small-world' networks, Nature 393, 440-442 (1998)

# The Newman-Moore-Watts model

- If there are no shortcuts, then the average distance between two randomly chosen vertices is of the order $n$, the number of vertices.

- But, as soon as there are just a few shortcuts, then the average geodesic distance between two randomly chosen vertices has an expectation of order $\log n$. (if $c = 2k$, and $ncp >> 1$, then the average geodesic distance is of the order $\frac{\log(ncp)}{c^2 p}$)

- 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.

# Power Law and the Barabasi-Albert model

- Barabasi and Albert (1999) noticed that the actor collaboration graph and the World Wide Web had degree distributions

$$d_k \sim Ck^{-\gamma}, \quad \text{for } k \to \infty \tag{1}$$

- *power-law behaviour*; constant $\gamma$ is called the power-law exponent
- subsequently a number of networks have been identified which show this type of behaviour (also called scale-free random graphs).

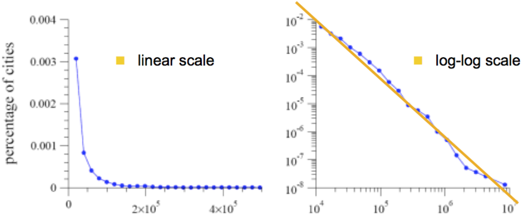



Figure: High skew (asymmetry) in the linear scale. Straight line on a log-log plot.

-Clauset et al. *Power-law distributions in empirical data*. SIAM review 51.4 (2009): 661-703. (Google Scholar citations: 9312 (2021))

- disputed the ubiquity of power-law degree distributions
- presented a statistically principled set of techniques that allow for the validation and quantification of power laws.

# Preferential attachment (the short story)

To explain this behaviour, Barabasi and Albert introduced the *preferential attachment* model for network growth:

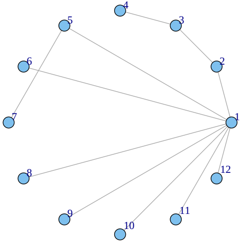
- First considered by Price in 1976 as a model for citation networks (who adapted the work of Herbert Simon, 1978 Nobel Prize in Economics and Turing Award in 1975)
- Suppose that the process starts at time 1 with two vertices linked by $m$ (parallel) edges.
- At every time $t \geq 2$ we add a new vertex with $m$ edges that link the new vertex to vertices already present in the network.
- We assume that the probability $\pi_i$ that the new vertex will be connected to a vertex $i$ depends on the degree $d(i)$ of $i$ so that

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

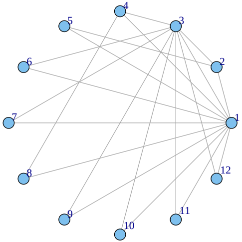
- To be precise, when we add a new vertex we will add edges one at a time, with the second and subsequent edges doing preferential attachment using the updated degrees.

# Preferential attachment

- For $m = 1$




- For $m = 2$



# Preferential Attachment (the long stroy)

- economist Herbert Simon (way ahead of his time in the study of ”complex systems")
- noted the occurrence of power laws in economic data (distribution of people’s personal wealth)
- proposed that people acquire money at a rate that is linearly proportional to their current income
- Simon proved that the ”rich-get-richer" (Yule process) leads to a power-law distribution
- Price adapted Simon’s methods to the network context, and called it *cumulative advantage*

# Preferential Attachment in networks

First considered by Price in the 1970s as a model for citation networks

- each new paper is generated with $c$ citations (mean)
- new papers cite previous papers with probability proportional to their in-degree (citations)
- for papers without any citations:
  - each paper is considered to have a number of default citations = $a$
  - probability of citing a paper with degree $k$, proportional to $k$
- leads to a power law with exponent

$$\alpha = 2 + \frac{a}{c}$$

we are about to prove.

# The *attachment kernel*

- most important modeling choice in constructing an attachment mechanism of network formation
- gives a rule for how new objects in the network connect to existing objects
- e.g., if nodes are added one at a time to a network and that each new node forms an edge with a single existing node *i* with prob.

$$q_i = \frac{a_i}{\sum_i a_i} \tag{2}$$

- $a_i$ is called the attachment kernel, usually depends only structural properties of node *i* (degree, local clustering coefficient, etc.)
- one can consider attachment mechanisms in which
  - new nodes form edges to more than one existing node at a time
  - rewiring between existing nodes occurs
  - edges or nodes are removed
  - network structures other than single nodes are added in each time step

# Unweighted, undirected network

- the usual preferential attachment mechanism is called *linear preferential attachment*
- has an attachment kernel that is a linear function of degree

$$a_i = k_i$$

  where $k_i$ is the degree of node $i$

- probability that a new edge attaches to node $i$ is given by

$$q_i = \frac{k_i}{\sum_i k_i} = \frac{k_i}{\overline{k} n} \tag{3}$$

- $\overline{k} = \sum_i k_i / n$ is the mean degree of the network on $n$ nodes.
- can consider separately in-degre and out-degree

# The Price model (1976)

- Price considered citation networks (directed networks)
- track in-degrees and out-degrees separately
- new papers cite papers that already exist
  - each node represents a paper
  - each directed edge represents a citation
- nodes can never be removed
- denote by $c$ the mean number of papers cited by a new paper
- the mean out-degree of the network is thus $c$
- a new paper cites an existing paper $i$ with probability

$$q_i = \frac{a + k_i}{\sum_i (a + k_i)} \tag{4}$$

- $a > 0$ is a "bonus" applied to each paper
- attachment kernel $a + k_i$

# Estimating $p_k$

Denote by

- $k_i \stackrel{\text{def}}{=}$ the in-degree of each node (number of citations)
- probability that a new edge attaches to node $i$ is given by

$$q_i = \frac{a+k_i}{\sum_i (a+k_i)} = \frac{a+k_i}{na+n\overline{k}} = \frac{a+k_i}{n(a+c)} \tag{5}$$

where $\overline{k} = \frac{1}{n}\sum_i k_i$

- and keeping in mind that the sum of all in-degrees equals to the sum of all out-degrees
- each new paper cites $c$ papers on average
- expected number of new citations to node $i$, when a new node comes in, is given by

$$c\frac{a+k_i}{n(a+c)} \tag{6}$$

# Estimating $p_k$

- $p_k(n) \stackrel{\text{def}}{=}$ fraction of nodes with in-degree $k$ for a network of size $n$
- there are $np_k(n)$ nodes with in-degree $k$
- expected number of new citations to all nodes with in-degree $k$ is given by (using (6))

$$np_k(n) \quad c\frac{a+k}{n(a+c)} = \frac{c(a+k)}{a+c}p_k(n) \tag{7}$$

- to study the dynamics of the equation, one approach is to write a *master equation*[1] for the evolution of the in-degree
- add a single node to a network with $n$ nodes:
  - number of nodes with in-degree $k$ increases by 1 for every node with previous degree $k-1$ that receives a new citation
  - expected number of such new nodes of in-degree k is

$$\frac{c(a+k-1)}{a+c}p_{k-1}(n)$$

[1] Master equations are equations of motion/differential equations used to describe the evolution of probabilities

# Adding a node...

Adding a single node to a network with $n$ nodes also implies:

- one node of in-degree $k$ is lost every time that such a node receives a new citation
- we have already computed the expected number of such nodes receiving citations in (7):

$$\frac{c(a+k)}{a+c} p_k(n)$$

Altogether, the **expected number of nodes with in-degree** $k$ **after one new node is added** is

$$(n+1)p_k(n+1) = \underbrace{np_k(n)}_{\text{\# previous nodes of in-degree k}} + \underbrace{\frac{c(a+k-1)}{a+c} p_{k-1}(n)}_{\text{number of new nodes of in-degree } k} - \underbrace{\frac{c(a+k)}{a+c} p_k(n)}_{\text{number of nodes of in-degree k that are lost}}, \quad k \geq 1 \tag{8}$$

# Estimating $p_k$

$\mathbb{E}$[number of nodes with in-degree $k$ after one new node is added] =

$$(n+1)p_k(n+1) = np_k(n) + \frac{c(a+k-1)}{a+c}p_{k-1}(n) - \frac{c(a+k)}{a+c}p_k(n), k \geq 1$$

- note that a separate equation is needed for $k = 0$ (omitted)
- taking $n \to \infty$, and letting $p_k$ denote $p_k(n \to \infty)$

$$p_k = \frac{c}{a+c}\left[(a+k-1)p_{k-1} - (a+k)p_k\right], \quad k \geq 1$$

$$p_0 = 1 - \frac{ac}{a+c}p_0, \quad k = 0$$

- With additional work, in the limit $n \to \infty$, arrive at

$$p_k \sim k^{-\beta}, \text{ for } k >> a$$

where the exponent is

$$\beta = 2 + \frac{a}{c} > 2$$

Note that $\beta = 3$ when $a = c$.

# Barabási-Albert Model (extension of Price)

- undirected networks (as opposed to Price's directed network)
- the number of new connections for each new node is exactly $c$ (thus now an integer), as opposed to Price where the number of connections was required only to take an average value of $c$ (but might vary from step to step)
- the probability can be shown to become

$$p_k = \frac{2c(c+1)}{k(k+1)(k+2)}, \quad \text{for } k \geq c \tag{9}$$

which, in the limit $k \to \infty$, recovers the power-law tail

$$p_k \sim k^{-3} \tag{10}$$

- Barabási, Albert-László and Albert, Réka. *Emergence of scaling in random networks*, Science 286.5439 (1999)

• Note 1: Not a whole lot of extra work for: 22,048 citations (2015); (36,484) (2020); (40,180) (2021); (42,393) (2022); (44,600) (2023)

• Note 2: *"A general theory of bibliometric and other cumulative advantage processes"*, Price, Journal of the American Society for Information, 1976 (1,083 citations) (2015); (1,721) (2022); (2,530) (2023);

# The Stochastic Block Model (SBM)

- also *Erdös-Renyi mixture model*, *latent block models* (*Holland, Laskey and Leinhardt (1983), Nowicky and Snijders (2001)*)
- assumes that vertices are of different types, say, there are $L$ different types (also $L$ clusters)
- 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

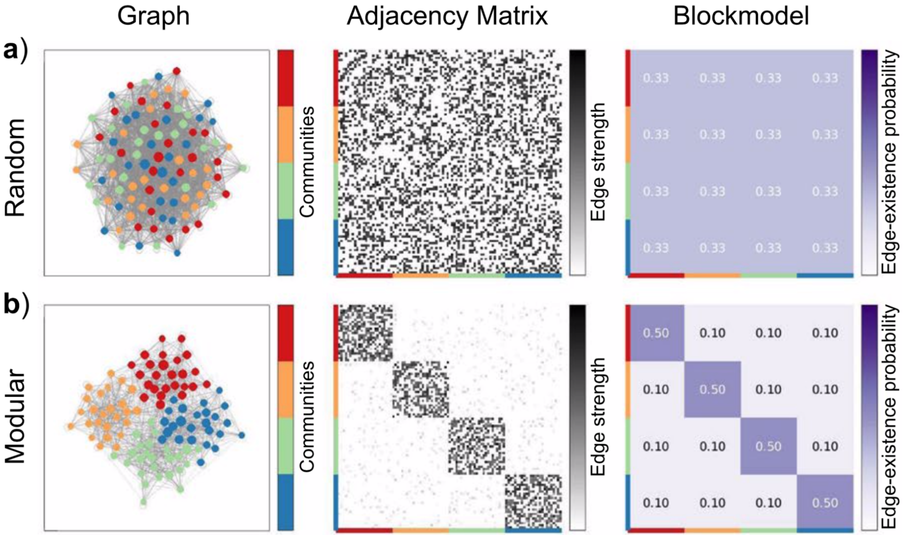
$$p_{i,j} = \mathbb{P}((u, v) \in \mathcal{E} | u \text{ is of type } i, v \text{ is of type } j) \tag{11}$$

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

$$\mathbb{E}(d(v)) = \sum_\ell \alpha_\ell \left( (|\mathcal{V}|\alpha_\ell - 1) p_{\ell,\ell} + \sum_{k \neq \ell} |\mathcal{V}| \alpha_k p_{k,\ell} \right) \tag{12}$$

- $|\mathcal{V}| \cdot \alpha_\ell$ is the expected size of cluster $\ell$
- Often the type allocation itself is not known.

# The Stochastic Block Model (SBM)




*Weighted Stochastic Block Models of the Human Connectome across the Life Span*, Joshua Faskowitz, Xiaoran Yan, Xi-Nian Zuo, Olaf Sporns, Scientific Reports, Volume 8, Article number: 12997 (2018)

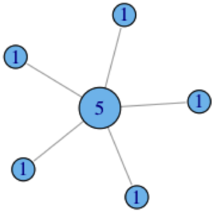
# Network Centrality Measures

- Mostly tools from Social Network Analysis
- Attempt to quantify the importance of nodes, 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
- Various notions/concepts of centrality (vary by context and purpose)
- Simplest measure: Degree Centrality. Limitations?
- Centralities are often not robust to small perturbations either of their definition or of network structure
- Be cautious about interpreting the results of centrality calculations

# Network Centrality Measures

- Probably hundreds of different types of centralities, though many of them are very similar to each other
- We will only discuss a few well-known examples
  - Degree Centrality (simplest type of centrality)
  - Closeness Centrality
  - Betweenness Centrality
  - Eigenvector Centrality
  - Katz Centrality
  - Page Rank algorithm
  - Bonachich Power Centrality

# Simplest centrality: Degree Centrality




- **Social network**: individuals who have connections to many others might have more influence, more prestige than those who have fewer connections.
- **Citation network**: the number of citations a paper receives from other papers, which is simply its in-degree in the citation network, is a measure of whether the paper has been influential or not.

# Closeness Centrality

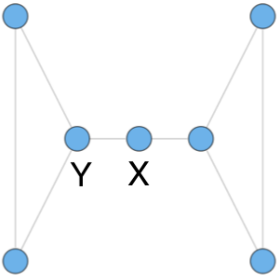
- Consider an undirected network
- Measures the mean distance from a vertex to other vertices (several definitions of it exist in the literature)
- A version of closeness centrality, which one might call an exponentially weighted closeness centrality (appropriate for both connected and disconnected networks) is

$$C_c(i) = \sum_{j \in G_i} 2^{-L_{ij}} \tag{13}$$

- $L_{ij}$ is the geodesic distance (i.e., the length of the shortest path) between vertices $i$ and $j$
- $G_i$ is the connected network component reachable from vertex $i$ (but excluding vertex $i$ itself)
- Path lengths in a weighted network need to be computed with a distance matrix rather than an adjacency matrix

# Betweenness Centrality

- The betweenness centrality of an object in a network measures the extent to which it lies on short paths
- A higher betweenness indicates that it lies on more short paths and hence should somehow be important for traversing between different parts of a network
- How many pairs of individuals would have to go through you in order to reach one another in the minimum number of hops? Who has higher betweenness, X or Y?



# Betweenness Centrality of a Vertex

- The geodesic betweenness $B_n(i)$ of a **vertex** in a weighted, undirected network is

$$B_n(i) = \sum_{s,t \in G} \frac{\Psi_{s,t}(i)}{\Psi_{s,t}}$$

  where vertices $s, t, i$ are all different from each other
- $\Psi_{s,t}$ denotes the number of shortest paths (geodesics) between vertices $s$ and $t$
- $\Psi_{s,t}(i)$ denotes the number of shortest paths (geodesics) between vertices $s$ and $t$ **that pass through vertex** $i$.
- The geodesic betweenness $B_n$ of a network is the mean of $B_n(i)$ over all vertices $i$

# Different centralities

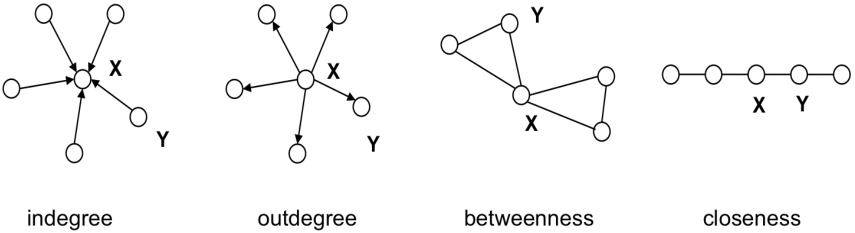



Figure: In each of the following networks, X has higher centrality than Y according to a particular measure

# Eigenvector Centrality

- Eigenvector centrality posits that a vertex is important because it is connected to other vertices that are important
- This notion is inherently recursive and leads naturally to an eigenvalue problem
- Closely related to an old notion of centrality due to Katz and PageRank centrality
- PageRank: famous because of its role in the Google search engine

Recall the Perron-Frobenius theorem

- a real square matrix with positive entries has a unique eigenvalue of largest magnitude and that eigenvalue is real ($\lambda_1 > \lambda_2$), with corresponding eigenvector with strictly positive components
- similar claim holds true for nonnegative matrices, under certain assumptions (irreducible/strongly connected).

# Eigenvector Centrality

- The eigenvector centrality $C_e(i)$ of vertex $i$ in an **undirected**, connected network is proportional to the sum of the eigenvector centralities of the neighboring vertices

$$C_e(i) = \frac{1}{\lambda} \sum_{j \in \mathcal{N}(i)} C_e(j)$$

where $\lambda$ is the largest eigenvalue of $A$

$$C_e(i) = \frac{1}{\lambda} \sum_{j} A_{ij} C_e(j)$$

in both weighted and unweighted networks

- $C_e(i)$ is the i-th component of the leading eigenvector of $A$

$$AC_e = \lambda C_e$$

- Perron-Frobenius: each entry of this eigenvector is positive
- If $G$ is not connected, the $A$ becomes block diagonal (one for each component), and one computes eigenvector centralities separately for each component.

# Issues with Eigenvector Centrality

$$C_e(i) = \frac{1}{\lambda} \sum_{j \in \mathcal{N}(i)} C_e(j)$$

- Say $u$ is connected to the rest of the network, but only outgoing edges and no incoming ones (directed graph setting)
- Then $u$ has centrality zero because there are no terms in the sum; which is acceptable, however...
- Assume $v$ has only one incoming edge ($u \rightarrow v$), then $v$ also ends up with centrality zero
- It holds true that only vertices that are in a strongly connected component of two or more vertices, or the out-component of such a component, can have non-zero eigenvector centrality
- Issue: acyclic networks (citation networks) have no strongly connected components of more than one vertex, thus all vertices will have centrality zero

# Katz centrality (1953)

- A variant of eigenvector centrality which allows each vertex a small amount of centrality *for free* regardless of its position in the network or the centrality of its neighbors

$$x_i = \alpha \sum_{j=1}^{n} A_{ij} x_j + \beta \tag{14}$$

- $\alpha, \beta > 0$
- $x = \alpha A x + \beta \mathbf{1}$
- (with $\beta = 1$) the Katz centrality $C_k(i)$ of vertex $i$, is given by the i-th component of the eigenvector

$$C_k = (I - \alpha A)^{-1} \mathbf{1}$$

- $I$ is the identity matrix, $\mathbf{1}$ denotes the all-ones vector
- $\alpha \in (0, \lambda)$ is a free parameter governing the balance between the eigenvector term and the constant term in (14)

# Page-Rank

- Issue with Katz centrality: if a vertex with high Katz centrality points to many others then those others also get high centrality
- A high-centrality vertex (like *Yahoo*) pointing to one million others gives *all* one million of them high centrality!
- Fix: share *Yahoo*'s contribution to the centrality of those one million pages to which it is pointing
- Centrality of vertex $u$ is obtained from that of its neighbors proportional to their centrality divided by their out-degree

$$x_i = \alpha \sum_{j=1}^{n} A_{ij} \frac{x_j}{k_j^{out}} + \beta \quad (15)$$

where $A_{ij} = 1$ if webpage $j$ links to webpage $i$ ($j \mapsto i$) and $A_{ij} = 0$ otherwise.
- Note: vertices with no out-going edges ($k_j^{out} = 0$) should contribute zero to the centrality of others, so set $k_j^{out} = 1$
- $D$ diagonal matrix with $D_{ii} = \max\{k_i^{out}, 1\}$

$$x = \alpha A D^{-1} x + \beta \mathbf{1} \quad (16)$$

# Page-Rank

- $x = \alpha A D^{-1} x + \beta \mathbf{1}$
- $D$ diagonal matrix with $D_{ii} = \max\{k_i^{out}, 1\}$

$$x = \beta(\boldsymbol{I} - \alpha A D^{-1})^{-1}\mathbf{1} = \beta D(D - \alpha A)^{-1}\mathbf{1}$$

- Set $\beta = 1$, just a re-scaling and arrive at the Page-Rank centrality

$$C_{PR} = D(D - \alpha A)^{-1}\mathbf{1}$$

- $\alpha$ governs the balance between the eigenvector term and the constant term
- In practice, Google uses $\alpha = 0.85$, no rigorous theory behind this choice
- Can customize the additive constant for each vertex

$$x_i = \alpha \sum_{j=1}^{n} A_{ij} \frac{x_j}{k_j^{out}} + \beta_i \qquad (17)$$

$$C_{PR} = D(D - \alpha A)^{-1}\boldsymbol{\beta}$$

# Almost-Page-Rank

- Without any additive constant

$$x_i = \alpha \sum_{j=1}^{n} A_{ij} \frac{x_j}{k_j^{out}} \tag{18}$$

- If $A$ is undirected, note that $x_i = k_i$ does the job (recovers degree centrality)

- For the undirected case, it yields a new network centrality though not very popular in practice.

# Almost-Page-Rank

Wishful thinking: would like the following *influece*-based weight system

- (a) webpages that link to i, and have high PageRank scores themselves, should be given more weight
- (b) webpages that link to i, but link to a lot of other webpages in general, should be given less weight

$$x_i = \sum_{j \mapsto i} \frac{x_j}{k_j^{out}} = \sum_{j=1}^{n} A_{ij} \frac{x_j}{k_j^{out}} \tag{19}$$

- where $A_{ij} = 1$ if webpage $j$ links to webpage $i$ ($j \mapsto i$) and $A_{ij} = 0$ otherwise.
- $k_j^{out}$ number of webpages $j$ links to

Indeed, for $j \mapsto i$, this is in line with our wishful thinking above, as the weight contribution $\frac{x_j}{k_j^{out}}$

- increases with $x_j$ as in (a)
- decreases with $k_j^{out}$ as in (b)

# Almost-Page-Rank written in matrix notation

$$x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}; \quad A = \begin{pmatrix} A_{11} & A_{13} \dots A_{1n} \\ A_{21} & A_{23} \dots A_{2n} \\ \vdots & \\ A_{n1} & A_{n3} \dots A_{nn} \end{pmatrix}; \quad M = \begin{pmatrix} k_1^{out} & 0 \dots\dots 0 \\ 0 & k_2^{out} \dots\dots 0 \\ \vdots & \vdots \qquad \vdots \\ 0 & 0 \dots\dots k_n^{out} \end{pmatrix} \tag{20}$$

- The previous system of linear equations can be simply written as

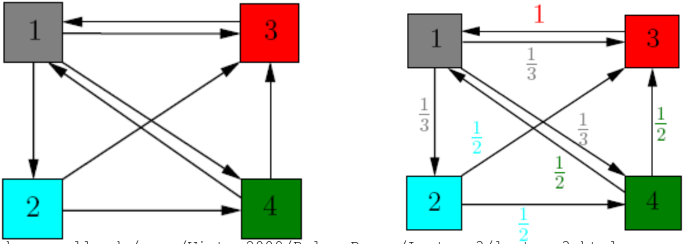
$$x = AM^{-1}x \tag{21}$$

- Letting $W = AM^{-1}$, this amounts to $x = Wx$, meaning that $x$ is an eigenvector of the matrix $W$ with eigenvalue 1.
- Good problem to consider - we know very well how to compute eigenvalues and eigenvectors of $W$, and there are very fast methods even for the case when $W$ is very large and sparse.

Questions to consider:

- how do we even know that $W$ has an eigenvalue of 1, so that such a vector $x$ even exists?
- even if it exists, is it unique? Is the problem well-defined?

# -Almost-Page-Rank




http://pi.math.cornell.edu/~mec/Winter2009/RalucaRemus/Lecture3/lecture3.html

Need to solve $x = Wx$ where

$$W = \begin{pmatrix} 0 & 0 & 1 & \frac{1}{2} \\ \frac{1}{3} & 0 & 0 & 0 \\ \frac{1}{3} & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{3} & \frac{1}{2} & 0 & 0 \end{pmatrix} \qquad (22)$$

Matrix $W$ is an example of a column-stochastic matrix:

- square matrix with all entries non-negative
- entries in each column sum to 1.

Stochastic matrices arise in the study of Markov chains (suitable for modelling problems in economics and operations research).

# Markov Chains

- random process on graph with states $1, 2, \ldots, n$, where a surfer moves between states (each move is a step of the process)
- $x^{(0)}$ is the $n$-dim vector with the starting probabilities; after one step $x^{(1)} = W^T x^{(0)}$ has the probabilities of being in each state
- the Almost-PageRank problem is that of Markov Chain, where the states are the webpages, and the transition probability matrix is given by $W^T$. Recall that $W^T_{ij} = W_{ji} = \frac{A_{ji}}{k_i^{out}}$
- the Markov Chain can be described as

$$\text{Prob}(\text{move } i \mapsto j) = \begin{cases} \frac{1}{k_i^{out}} & \text{if } i \mapsto j \\ 0 & \text{otherwise} \end{cases} \tag{23}$$

- random surfer on the web by clicking links uniformly at random
- a stationary distribution (left-eig) of a Markov chain is a probability vector $x$ (i.e., its entries are $\geq 0$ and sum to 1) with $x = Wx$.
- if MC is strongly connected (any state can be reached from any other state), the stationary distribution $x$ exists and is unique
- stationary dist. construed as the proportion of visits the MC pays to each state after a very long time (*Ergodic Theorem*).

# Spectra of a column-stochastic matrix

## Theorem

*A column-stochastic matrix $W$ has an eigenvalue $\lambda = 1$, which is also its largest eigenvalue.*

## Proof.

(a) Let $W$ be an $n \times n$ column-stochastic matrix.

- first note that $W$ and $W^T$ have the same eigenvalues (their eigenvector will usually be different)
- denote $\mathbf{1} = [1, 1, \ldots 1]^T$ to be the all-ones vector of length $n$
- since $W$ is column-stochastic, $W^T\mathbf{1} = \mathbf{1}$ (since all columns of $W$ sum up to 1)
- hence $\mathbf{1}$ is an eigenvector of $W^T$ (but not of $W$) with corresponding eigenvalue $\lambda = 1$
- therefore $\lambda = 1$ is also an eigenvalue of $W$.

□

To prove the second part, we need to make a detour first.

# Gershgorin circle theorem

- Let $A \in \mathbb{C}^{n \times n}$. For $i = 1, \ldots, n$, let

$$r_i = \sum_{j \neq i} |A_{ij}| \tag{24}$$

be the sum of the absolute values of the non-diagonal entries in row $i$

- Let $D_i$ denote the closed disk in the complex place centered at $A_{ii}$ with radius $r_i$

$$D_i := D(A_{ii}, r = r_i) = \{z \in \mathbb{C} : |z - A_{ii}| \leq r_i\} \tag{25}$$

- Such a disk is referred to as a *Gershgorin disc*.

## Theorem

*Every eigenvalue of A lies within at least one of the Gershgorin discs* $D_i, i = 1, \ldots, n$.

## Theorem (Stronger version)

*All the eigenvalues of A lie in the union of the disks* $D_i$ *for* $i = 1, \ldots, n$. *If some set of k overlapping disks is disjoint from all the other disks, then exactly k eigenvalues lie in the union of these k disks.*

# Spectra of a column-stochastic matrix (cont)

## Proof.

(b) Let us now prove that $\lambda = 1$ is also the largest eigenvalue of $W$

- application of the Gershgorin circle theorem to $W^T$
- consider the $k^{th}$ row of $W^T$
- denote the diagonal elements as $w_{k,k}$, with the radius given by

$$r_i = \sum_{i \neq k} |w_{k,i}| = \sum_{i \neq k} w_{k,i} \tag{26}$$

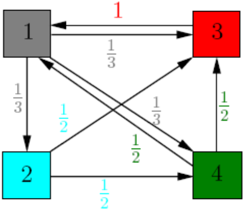
  since $w_{k,i} \geq 0, \forall k, i$ ($W$ has all entries non-negative).
- this is a circle centered at $w_{k,k} \in [0, 1]$, with radius

$$r_i = \sum_{i \neq k} w_{k,i} = 1 - w_{k,k} \leq 1 \tag{27}$$

- this circle has 1 on its perimeter, and this holds true for all Gershgorin circles of the matrix $W$
- since all eigenvalues of $W$ lie in the union of the Gershgorin circles, all eigenvalues $\lambda_i$ must satisfy

$$|\lambda_i| \leq 1 \tag{28}$$

# -Almost-Page-Rank




$$W = \begin{pmatrix} 0 & 0 & 1 & \frac{1}{2} \\ \frac{1}{3} & 0 & 0 & 0 \\ \frac{1}{3} & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{3} & \frac{1}{2} & 0 & 0 \end{pmatrix} \quad (29)$$

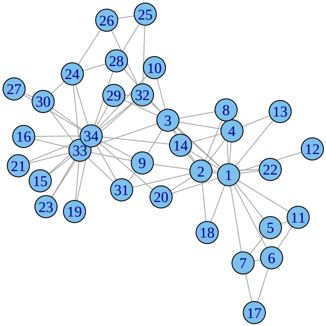
When solving $Wx = \lambda x$, we obtain as eigenvector of $W$ associated to eigenvalue $\lambda = 1$ the vector $x = [x_1, x_2, x_3, x_4]^T = [\frac{12}{31}, \frac{4}{31}, \frac{9}{31}, \frac{6}{31}]^T$

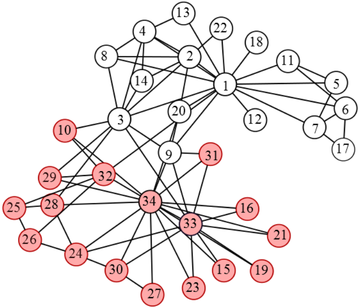
- somewhat surprisingly, page 3 is no longer the most important one, but page 1 is
- the apparently important page 3 (which has three webpages linking to it) has only one outgoing link, which gets all its “voting power”, and that link points to page 1.

• Important: we only need to compute the eigenvector associated with the eigenvalue 1 (also the largest one), which can be computed fast with standard *power iteration*, also very scalable for very large graphs (even with billions on nodes!).

# Zachary's Karate club network

Arguably the most famous network data set: the Zachary's Karate club network, 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 (one of the few cases when a proxy for the *"ground truth"* is known).

# Modularity in unweighted & undirected networks

- for a given a partition of a network into $k$ non-overlapping groups of nodes, we would like to know what is the expected number of connections within a group
- if nodes $i$ and $j$ have degrees $k_i$ and $k_j$, the probability that in a random network they are neighbours, ie. $(i,j) \in E$, is given by

$$\mathbb{P}[(i,j) \in E] = \frac{k_i k_j}{2m} \tag{30}$$

  where $m$ is the total number of edges in $G$
- the expected number of edges within a group is

$$\frac{1}{2} \sum_{i,j} \frac{k_i k_j}{2m} \delta(x_i, x_j) \tag{31}$$

  where $x_i$ is the community to which node $i$ is assigned

$$\delta(x_i, x_j) = \begin{cases} 1 & \text{if } g_i = g_j \\ 0 & \text{if } g_i \neq g_j \end{cases} \tag{32}$$

# Modularity in unweighted & undirected networks

- if there are $k$ groups of nodes in the network, then the numbder of edges within groups is given by

$$\frac{1}{2}\sum_{i,j} A_{ij}\delta(x_i, x_j) \tag{33}$$

- consider the difference between the actual versus the expected number of edges that connect vertices of similar types/groups

$$Q = \underbrace{\frac{1}{2}\sum_{i,j} A_{ij}\delta(x_i, x_j)}_{\text{what we observe in the network}} - \underbrace{\mathbb{E}\left[\frac{1}{2}\sum_{i,j} A_{ij}\delta(x_i, x_j)\right]}_{\text{what we expect at random}} \tag{34}$$

$$= \frac{1}{2}\sum_{i,j} A_{ij}\delta(x_i, x_j) - \frac{1}{2}\sum_{i,j} \frac{k_i k_j}{2m}\delta(x_i, x_j) \tag{35}$$

$$= \frac{1}{2}\sum_{i,j}\left(A_{ij} - \frac{k_i k_j}{2m}\right)\delta(x_i, x_j) \tag{36}$$

- in practice, one considers the fraction of edges, not the number

$$\frac{1}{2m}\sum_{i,j}\left(A_{ij} - \frac{k_i k_j}{2m}\right)\delta(x_i, x_j) \tag{37}$$

# Modularity in unweighted & undirected networks

$$Q = \frac{1}{2m} \sum_{i,j} \left( A_{ij} - \frac{k_i k_j}{2m} \right) \delta(x_i, x_j) \quad (38)$$

$$= \frac{1}{2m} \sum_{i,j} B_{ij} \delta(x_i, x_j) \quad (39)$$

- where the modularity matrix $B$ is defined as

$$B_{ij} = A_{ij} - \frac{k_i k_j}{2m} \quad (40)$$

- can be extended to
  - weighted networks, with the strengths replacing the degrees
  - to directed networks, $\frac{k_i k_j}{m}$ replacing $\frac{k_i k_j}{2m}$
- the modularity $Q$ can be seen as the quality of the partition
- we aim to maximize $Q$ in order to find communities
- community-detection methods optimize for partitions with high modularity
- very rich literature on modularity optimization (including spectral).

# The Newman-Girvan algorithm

- a popular algorithm to maximize $Q$ by Newman & Girvan

  Girvan, Michelle, and Mark EJ Newman. "Community structure in social and biological networks". Proceedings of the National Academy of Sciences (PNAS) (2002). Google Scholar: 16,201 (2021); 18,700 (2023)

- recall: the betweenness centrality of an edge is the number of shortest paths between vertex pairs that run along the edge in question, summed over all vertex pairs
- the Girvan & Newman algorithm involves simply
  - calculating the betweenness of all edges in the network,
  - removing the one with highest betweenness
  - 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

Many other algorithmic approaches: spectral, Louvain modularity algorithm, ...

# Milestones in Community Structure Detection

| 2002 | 2003 | 2004 | 2006 | 2007-2017 |
|---|---|---|---|---|

**Girvan, Newman** (2002)
Community structure in social and biological networks

**Newman** (2003)
Fast algorithm for detecting community structure in networks

**Clauset, Newman, Moore** (2004)
Finding community structure in very large networks

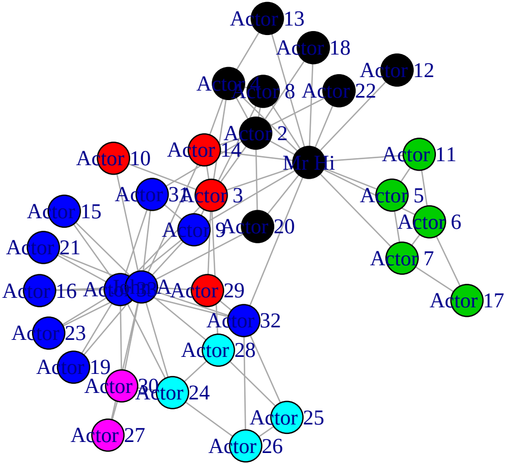
**Newman** (2006)
Modularity and Community Structure Detection

**Over 5000 papers** (2007-2017)
On the topic Community Structure Detection

Source: Chandole, Kabre, Aggarwal

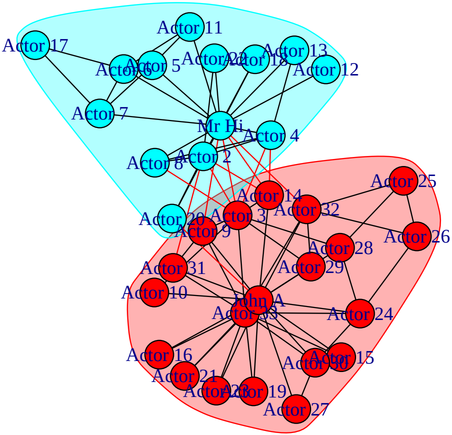
# The karate club example

The Girvan-Newman algorithm on the karate club example:



# The karate club example

The Girvan-Newman algorithm on the karate club example forced to produce exactly 2 groups.

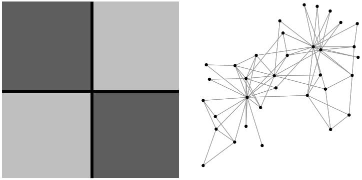


# Further topics within networks

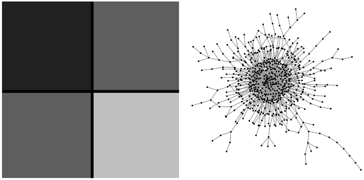
- many networks have a hierarchical structure (food webs, GICS decomposition: sectors, groups, etc). How do we take this into account when, for example, denoising the empirical correlation matrix?
- networks may also be dynamic; each given network data is a snapshot in time
- edges in networks may be of different types (as in knowledge graphs), leading to multilayer/multiplex networks
- certain networks have structures beyond the usual clustering, eg core-periphery
- anomaly detection (fraud detection in financial transaction networks)
- network change-point detection (NCPD) (identification of market regimes)

# A range of structures in networks

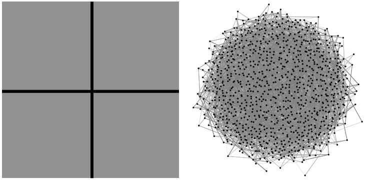
Idealized block models of network adjacency matrices; darker blocks correspond to denser connections among its component nodes.



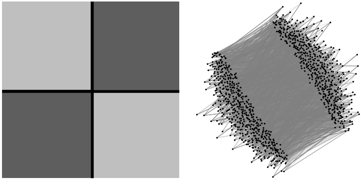
(a) Low-dimensional structure



(b) Core-periphery structure



(c) Expander or complete graph

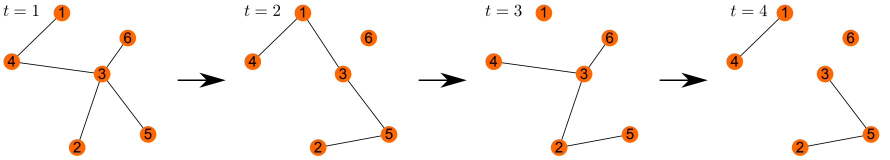


(d) Bipartite structure

Jeub, L. G., Balachandran, P., Porter, M. A., Mucha, P. J., & Mahoney, M. W. (2015). Think locally, act locally: Detection of small, medium-sized, and large communities in large networks. Physical Review E, 91(1), 012821

# Temporal (or time-dependent) networks

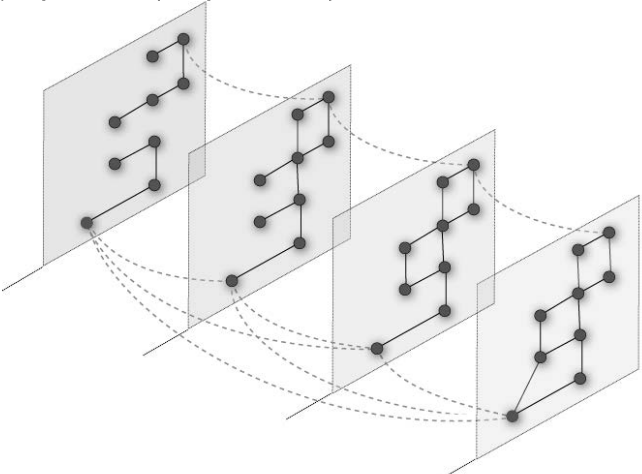
- networks for which either the nodes or the edges (or both) change over time
- voting networks (U.S. Senate roll call vote similarities) encoding the pairwise agreement/similarity between their voting, specified independently for each 2-year Congress
- social networks where new connections are formed over time, and also new users join the network.
- daily correlation matrices, co-occurrence network, co-jumps networks




Credit: Heather Harrington

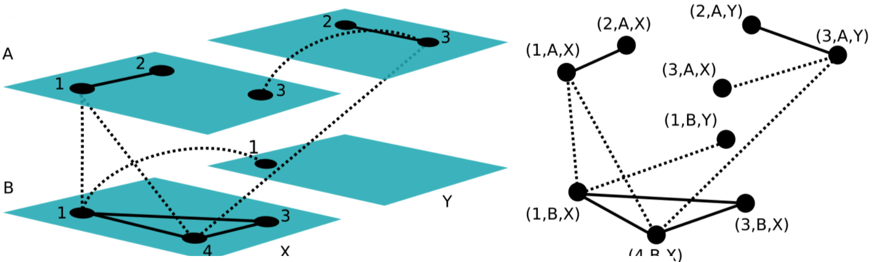
# Temporal (or time-dependent) networks

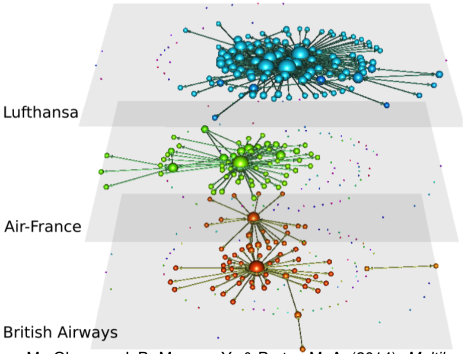
- network for which either the nodes or the edges (or both) change over time
- *interslice connections* (dashed lines) are encoded by $T_{jrs}$, specifying the coupling of node $j$ to itself between slices $r$ and $s$



Mucha, P. J., Richardson, T., Macon, K., Porter, M. A., & Onnela, J. P. *Community structure in time-dependent, multiscale, and multiplex networks*. Science, 328(5980), 876-878 (2010)

# Multilayer networks

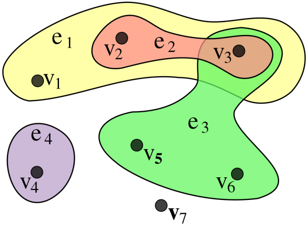



Kivelä, M., Arenas, A., Barthelemy, M., Gleeson, J. P., Moreno, Y., & Porter, M. A. (2014). *Multilayer networks.* Journal of complex networks, 2(3), 203-271.

# Hypergraphs

- certain applications require moving beyond pairwise connections
- eg., co-authorship networks: multiple authors write a joint paper
- in hypergraphs, the edges (denoted as hyperedges) are allowed to connect more than two nodes
- *k*-regular hypergraph: every node has degree *k*, i.e., it is contained exactly in *k* hyperedges;
- often represented as *tensors*
- in a financial context, it can be used to capture
  - co-jump behaviour: events in which a subset of stocks co-jump within the same short time interval
  - co-occurrence events: multiple companies mentioned in the same news article

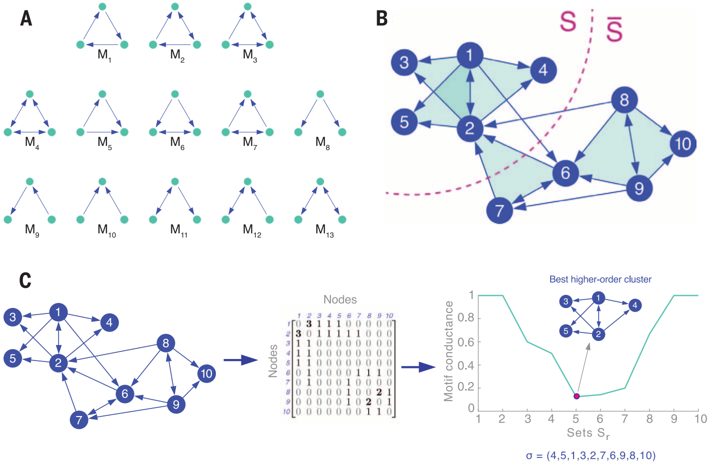



Credit: Wikipedia

Note there is a reading group organized this term in the Statistics Department, by Gesine Reinert where some of our joint students and postdocs will be involved.

# Higher-order organization of complex networks

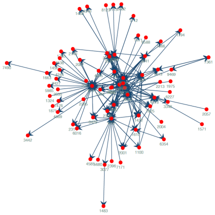
- explores higher-order organization of complex networks at the level of small network subgraphs/motifs
- networks exhibit rich higher-order organizational structures exposed by clustering on higher-order connectivity patterns



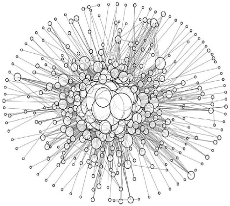

Credit: Benson, Austin R., David F. Gleich, and Jure Leskovec. *Higher-order organization of complex networks.* Science 353.6295 (2016): 163-166.

# Core-Periphery Networks

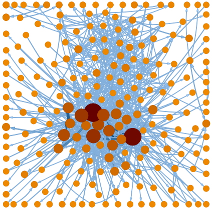
Chaojun Wang, *Core-Periphery Trading Networks*, (2016)



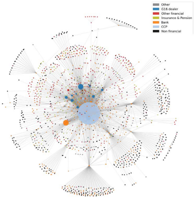
Li and Schüerhoff - muni bonds



Hollifield, Neklyudov, Spatt - ABS



Bech and Atalay - Fed funds



ESRB - Interest rate swaps

ESRB - Credit default swaps

ESRB - FX forwards

Figure: Core-periphery trading networks in OTC markets.

# Core-Periphery structure in (undirected) networks

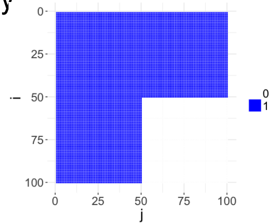
Structure that consists of densely connected core vertices $C$ and sparsely connected peripheral vertices $P$ (where $V = C \cup P$)

- core vertices in $C$ tend to be well-connected both among themselves and to peripheral vertices in $P$
- peripheral vertices are sparsely connected to other vertices.

A stochastic block model for core-periphery

$$A = \begin{array}{|c|c|} \hline p_{cc} & p_{cp} \\ \hline p_{cp} & p_{pp} \\ \hline \end{array} \qquad (41)$$

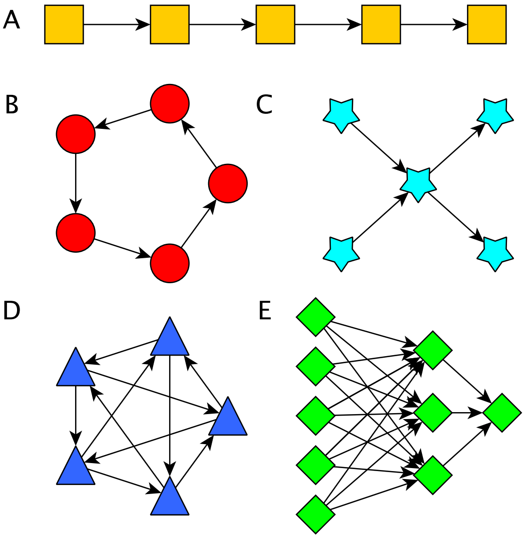
$$\widetilde{A} = \begin{array}{|c|c|} \hline 1 & 1 \\ \hline 1 & 0 \\ \hline \end{array} \qquad (42)$$




- $p_{cc} = p_{cp} = 1$ and $p_{pp} = 0$: idealized block model, leading to an adjacency matrix of rank 2
- $n_c$ & $n_p$ the number of core and peripheral vertices; $n_c + n_p = n$
- extension to directed networks: A. Elliott, A. Chiu, M. Bazzi, G. Reinert, M. Cucuringu, *Core-periphery structure in directed networks*, Proc. of the Royal Society A 476, no. 2241

# Anomaly detection - identification of heavy structures

The weights of the edges within such structures are considerably larger than the average weight of the ambient graph.



# Network Change-point detection (NCPD)

- dynamic networks that are temporal sequences of graph snapshots
- goal: detect abrupt changes in their structure
- cross-correlation networks of daily stock returns (computed over a one-month interval) from the SP 500 index in a period of ~ 20 years (Feb. 2000 - Dec. 2020).
- covariate information available: volatility, volume, etc
- move beyond the traditional uni/multi-variate time series for change-point detection (often based on cumsum statistics)

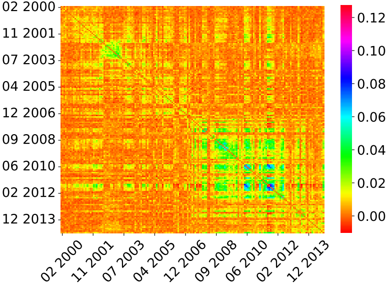



Figure: Matrix of Adjusted Rand Index values between the partitions obtained for each pair of graph snapshots in the correlation network of S&P 500 stock returns. The first two digits denote the month, followed by the year.


Figure: Change-point detection statistics, as obtained by various methods.

## Graph Neural Networks

Intuition behind GNN embedding approaches: at each iteration (search depth, layer) the nodes aggregate information from their local neighbors; as this process iterates, nodes gather information from further reaches of the graphs

- at each layer $k$, a representation of each node is available $h^k$
- each node $v$ AGGREGATEs the representations/embedding of the nodes in its immediate neighborhood $\{h_u^{k-1}, \forall u \in \mathcal{N}(v)\}$ into a single vector $h_{\mathcal{N}(v)}^{k-1}$ (this depends on node embeddings from previous iteration $k-1$)
- next, $h_{\mathcal{N}(v)}^{k-1}$ is CONCATenated with $h_v^{k-1}$, and fed through a fully connected layer with nonlinear activation function $\sigma$, whose output is $\mapsto h_v^k$

$$h_v^k \leftarrow \sigma\left(W^k \cdot \mathrm{CONCAT}(h_v^{k-1}, h_{\mathcal{N}(v)}^{k-1})\right) \quad (43)$$

- in order to learn useful, predictive representations in a fully unsupervised setting, one can apply a graph-based loss function to the final output embeddings $z_u$, and tune the weight matrices $W^k$ and parameters of the AGGREGATE function via stochastic gradient descent.

Remarks

- graph-based loss fcn. promotes nearby nodes to have similar representations, while enforcing that the representations of far-away nodes are highly distinct
- when the representations are used for a specific downstream task, the unsupervised loss above is replaced/augmented by a task-specific objective

# Relevant literature for applications of networks in finance


1. *"Sentiment Correlation in Financial News Networks and Associated Market Movements"*, Wan, Yang, Marinov, Calliess, Zohren, Dong, Scientific Reports 11, 3062 (2021)
2. *"Temporal Graph Networks for Deep Learning on Dynamic Graphs"*, Rossi, Chamberlain, Frasca, Eynard, Monti, Bronstein, `https://arxiv.org/abs/2006.10637`
3. *"Topological structures in the equities market network"*, Gregory Leibon, Scott Pauls, Daniel Rockmore, and Robert Savell, PNAS 2008, 105 (52) 20589-20594
4. *"Modeling the Stock Relation with Graph Networks for Overnight Stock Movement Prediction"*, IJCAI-20, `https://www.ijcai.org/Proceedings/2020/0626.pdf`
5. *"Modeling the Momentum Spillover Effect for Stock Prediction via Attribute-Driven Graph Attention Networks"*, AAAI Conference on Artificial Intelligence, 35(1), 55-62
6. *"Knowledge Graph-based Event Embedding Framework for Financial Quantitative Investments"*, Cheng, Yang, Wang, Zhang, Zhang, SIGIR 2020
7. *"Analysis of Equity Markets: A Graph Theory Approach analysis of equity markets a graph theory approach"*, `https://evoq-eval.siam.org/Portals/0/Publications/SIURO/Volume%2010/Analysis_Equity_Markets_A_Graph_Theory_Approach.pdf?ver=2018-02-28-145946-083`
8. *"Stock Network Stability After Crashes Based on Entropy Method"*, Front. Phys., 12 June 2020 `https://doi.org/10.3389/fphy.2020.00163`
9. *"Stock market network's topological stability: Evidence from planar maximally filtered graph and minimal spanning tree"*, International Journal of Modern Physics B (2015)
10. *"Correlation based networks of equity returns sampled at different time horizons"* `https://arxiv.org/pdf/physics/0605251.pdf`
11. *"Stability Analysis of Company Co-Mention Network and Market Graph Over Time Using Graph Similarity Measures"*, J. Open Innov. Technol. Mark. Complex. 2019