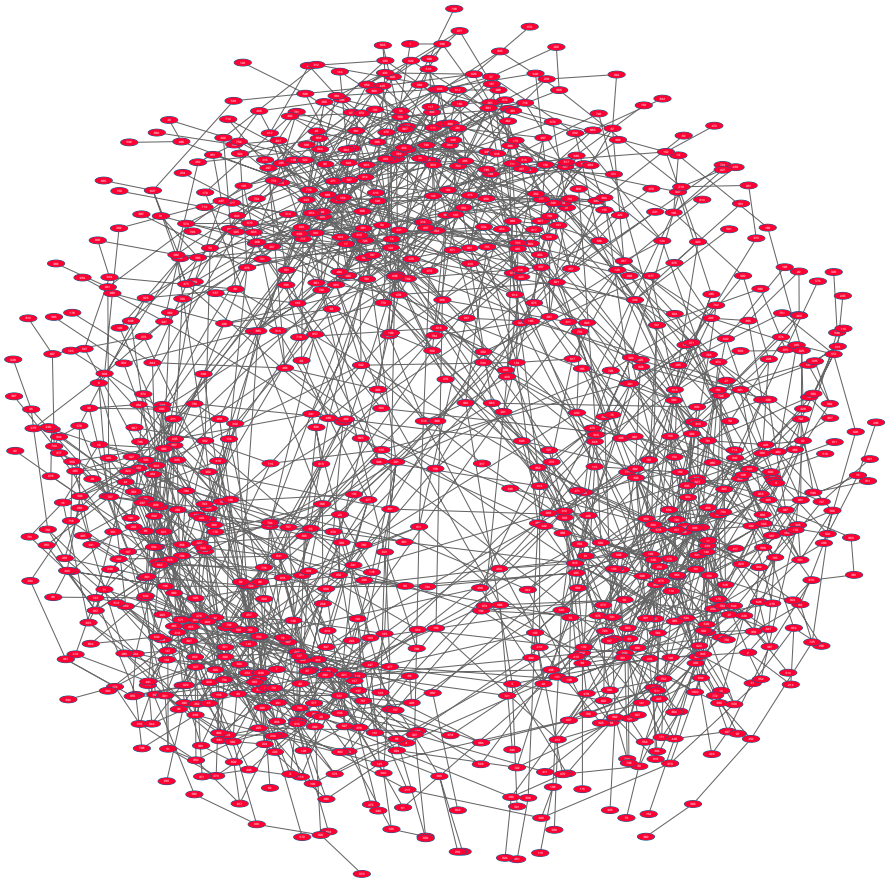
Module 6CCMCS02/7CCMCS02
Theory of Complex Networks

Compact Lecture Notes
Version of 30 September 2016



ACC Coolen
Department of Mathematics
King's College London

1. Introduction

The challenge of large data sets. In recent decades our ability to collect and store vast amounts of quantitative data has increased dramatically. This includes socio-economic data, such as social links between individuals or professional collaboration networks, consumer preferences and commercial interactions, trade dependencies among corporations, credit or insurance obligations between financial institutions. We have access to traffic data on computer and telecommunication systems, satellite networks, the internet, electricity grids, rail, road or air travel connections and distribution networks. We collect and process large amounts of geological and meteorological data, data on sea levels, air and water pollution, volcanic and seismic records, and sizes of polar and glacier ice sheets. Finally, we have seen an explosion in recent years of biomedical data, such as experimental data on biochemical processes and structures at cellular, subcellular and even molecular levels, the topologies of complex composite signalling and information processing systems such as the brain or the immune system, genomic and epigenetic data (gene expression levels, DNA sequences), epidemiological data, and vast numbers of patient records with clinical information.

However, one tends to collect data for a reason. This reason is usually the desire to understand the dynamical behaviour of the complex system that generated the data, to predict with reasonable accuracy its future evolution and its response to perturbations or interventions, or to understand how it was formed. We may want this for commercial gain, to improve and optimise a system's efficiency, to design effective regulatory controls, or (in the case of medicine) to understand and cure diseases. For small and simple systems the translation of observation into qualitative and quantitative understanding of design and function is usually not difficult. In contrast, if we collect data on complex systems with millions of nonlinearly interacting variables, just having a list of the parts and their connections and observations of their collective patterns of behaviour is no longer enough to understand how these systems work.

Networks as data reduction and visualisation tools. A first and useful step in modelling structural data on complex many-variable systems is to visualise these systems as networks or graphs. The idea is to represent each observed system component as a node in the network, and each observed interaction between two components as a link between the two corresponding nodes. Dependent upon one's research domain, the nodes of such networks may represent anything ranging from genes, molecules, or proteins (in biology), via processors or servers (in computer science), to people, airports, power plants or corporations (in social science or economics). The links could refer to biochemical reactions, wired or wireless communication channels, friendships, financial contracts, etc. The price paid for the complexity reduction is the loss of information. Limiting ourselves to a network representation means that we only record *which* parts interact, and disregard *how and when* they interact. However, the rationale is that the topology of the interactions between a

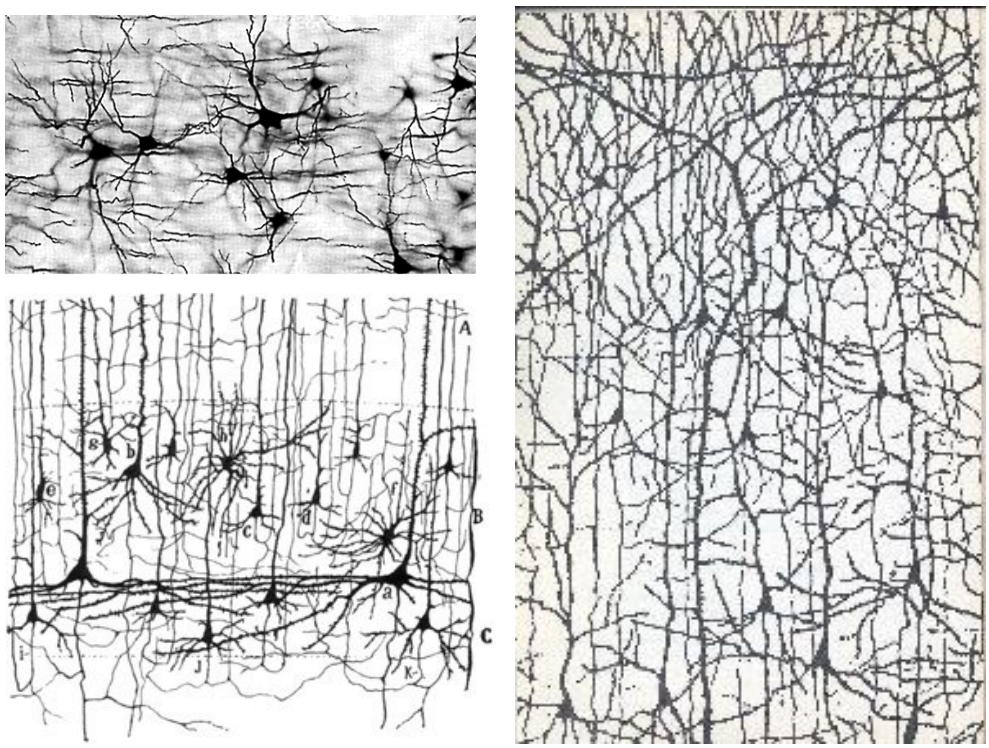
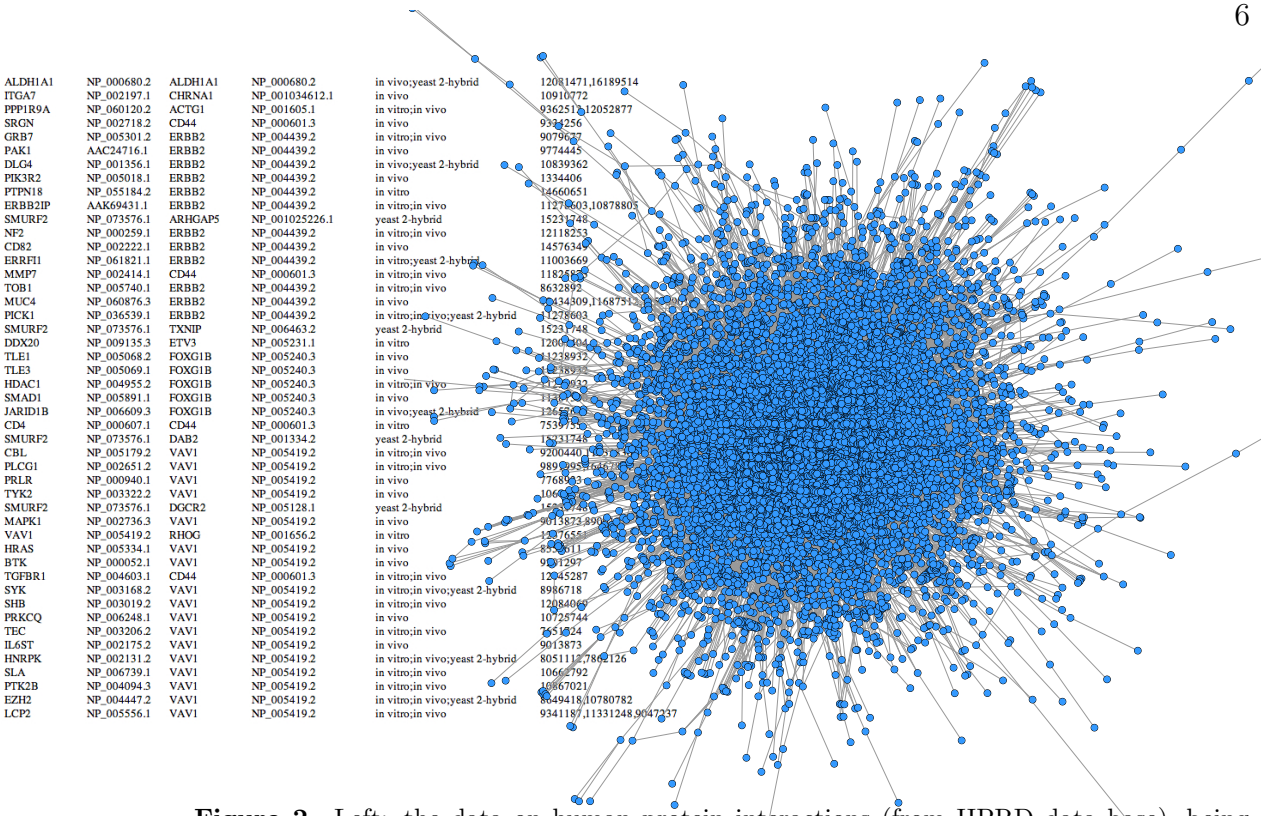


Figure 1. Neural networks, areas of the brain mapped using Golgi's staining technique by Ramon y Cajal around 1900. The nodes are brain cells (neurons), and the links are coated fibres via which the neurone communicate electrical signals. The staining shows only a tiny fraction of the links, in reality a neuron is connected on average to around 10^4 other neurone. The topologies of these networks vary according to the brain area that is being imaged, ranging from rather regular (in areas related to preprocessing of sensory signals) to nearly amorphous (in higher cognitive areas).

system's components should somehow be a fingerprint of its function, and that much can be learned from the topologies of such networks alone.

For example, DNA contains the assembly instructions for large and complicated macromolecules called proteins. These proteins serve as building material for all the parts of the cell, as molecular processing factories, information readers, translators, sensors, transporters and messengers. They interact with each other by forming (temporary) complexes, which are (meta)stable super-molecules formed of multiple proteins that attach to each other selectively. Many experimental groups produce tables of molecular binding partners such as that shown in Fig. 2. Network representations of these data have been very useful to generate intuition on the possible relevance of individual nodes (i.e. proteins), and to suggest functional modules. There are thousands of other examples of complex systems that tend to be modelled as networks, of which a selection is shown in the various figures in this section.



ALDH1A1	NP_000680.2	ALDH1A1	NP_000680.2	in vivo;yeast 2-hybrid	12081471,16189514
ITGA7	NP_002197.1	CHRNA1	NP_001034612.1	in vivo	10910772
PPP1R9A	NP_060120.2	ACTG1	NP_001605.1	in vitro;in vivo	9362515,12052877
SRGN	NP_002718.2	CD44	NP_000601.3	in vivo	9344256
GRB7	NP_005301.2	ERBB2	NP_004439.2	in vitro;in vivo	9079637
PAK1	AAC24716.1	ERBB2	NP_004439.2	in vivo	9774445
DLG4	NP_001356.1	ERBB2	NP_004439.2	in vivo;yeast 2-hybrid	10839362
PIK3R2	NP_005018.1	ERBB2	NP_004439.2	in vivo	1334406
PTPN18	NP_055184.2	ERBB2	NP_004439.2	in vitro	14660651
ERBB2IP	AAK69431.1	ERBB2	NP_004439.2	in vitro;in vivo	11278603,10878805
SMURF2	NP_073576.1	ARHGAP5	NP_001025226.1	yeast 2-hybrid	15231748
NF2	NP_000259.1	ERBB2	NP_004439.2	in vitro;in vivo	12118253
CD82	NP_002222.1	ERBB2	NP_004439.2	in vivo	14576349
ERRFI1	NP_061821.1	ERBB2	NP_004439.2	in vitro;yeast 2-hybrid	11003669
MMP7	NP_002414.1	CD44	NP_000601.3	in vitro;in vivo	11825882
TOB1	NP_005740.1	ERBB2	NP_004439.2	in vitro;in vivo	8632892
MUC4	NP_060876.3	ERBB2	NP_004439.2	in vivo	9434309,11687312
PICK1	NP_036539.1	ERBB2	NP_004439.2	in vitro;in vivo;yeast 2-hybrid	11278603
SMURF2	NP_073576.1	TXNIP	NP_006463.2	yeast 2-hybrid	15231748
DDX20	NP_009135.3	ETV3	NP_005231.1	in vitro	12004304
TLE1	NP_005068.2	FOXG1B	NP_005240.3	in vivo	11238932
TLE3	NP_005069.1	FOXG1B	NP_005240.3	in vivo	11238932
HDAC1	NP_004955.2	FOXG1B	NP_005240.3	in vitro;in vivo	11238932
SMAD1	NP_005891.1	FOXG1B	NP_005240.3	in vivo	11238932
JARID1B	NP_006609.3	FOXG1B	NP_005240.3	in vivo;yeast 2-hybrid	11238932
CD4	NP_000607.1	CD44	NP_000601.3	in vitro	7539753
SMURF2	NP_073576.1	DAB2	NP_001334.2	yeast 2-hybrid	15231748
CBL	NP_005179.2	VAV1	NP_005419.2	in vitro;in vivo	9200440
PLCG1	NP_002651.2	VAV1	NP_005419.2	in vitro;in vivo	9891958
PRLR	NP_000940.1	VAV1	NP_005419.2	in vivo	7768503
TYK2	NP_003322.2	VAV1	NP_005419.2	in vivo	10669
SMURF2	NP_073576.1	DGCR2	NP_005128.1	yeast 2-hybrid	15231748
MAPK1	NP_002736.3	VAV1	NP_005419.2	in vivo	9013873,890
VAV1	NP_005419.2	RHOG	NP_001656.2	in vitro	1376555
HRAS	NP_005334.1	VAV1	NP_005419.2	in vivo	8530311
BTK	NP_000052.1	VAV1	NP_005419.2	in vivo	9311297
TGFBR1	NP_004603.1	CD44	NP_000601.3	in vitro;in vivo	12045287
SYK	NP_003168.2	VAV1	NP_005419.2	in vitro;in vivo;yeast 2-hybrid	8986718
SHB	NP_003019.2	VAV1	NP_005419.2	in vitro;in vivo	12084049
PRKCQ	NP_006248.1	VAV1	NP_005419.2	in vivo	10727444
TEC	NP_003206.2	VAV1	NP_005419.2	in vitro;in vivo	7651824
IL6ST	NP_002175.2	VAV1	NP_005419.2	in vivo	9013873
HNRPK	NP_002131.2	VAV1	NP_005419.2	in vitro;in vivo;yeast 2-hybrid	8051113,7862126
SLA	NP_006739.1	VAV1	NP_005419.2	in vitro;in vivo	10662792
PTK2B	NP_004094.3	VAV1	NP_005419.2	in vitro;in vivo	10867021
EZH2	NP_004447.2	VAV1	NP_005419.2	in vitro;in vivo;yeast 2-hybrid	8649418,10780782
LCP2	NP_005556.1	VAV1	NP_005419.2	in vitro;in vivo	9341187,11331248,9047237

Figure 2. Left: the data on human protein interactions (from HPRD data base), being lists of pairs of protein species that have been observed to interact, together with codes of each species and information on the experiments where interaction was observed. Each line reports on one pair-interaction. This database contains some 70,000 reported interactions (about half of the interactions believed to exist among human proteins). Right: the network representation of the data on the left. Each protein is represented by a node, and each pair-interaction (each line on the left) is represented by a link connecting the two nodes concerned. Since interaction of two proteins is a symmetric property, this graph has nondirected links.

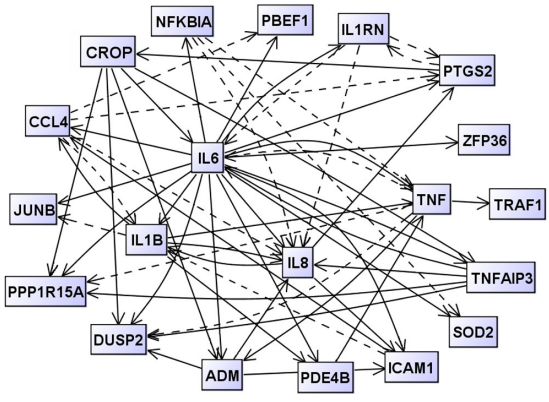


Figure 3. Gene regulation networks: here the nodes represent different genes in humans, and the links (which now are directional) indicate the ability of individual genes (when 'switched on') to affect the extent to which other genes are activated. Here the links are moreover 'weighted' (i.e. they carry a numerical value), indicated by solid arrows (positive value, excitatory effect) or dashed arrows (negative value, inhibitory effect). The figure here shows only a small subset of the genes – the true number of nodes is in the order of 20,000.

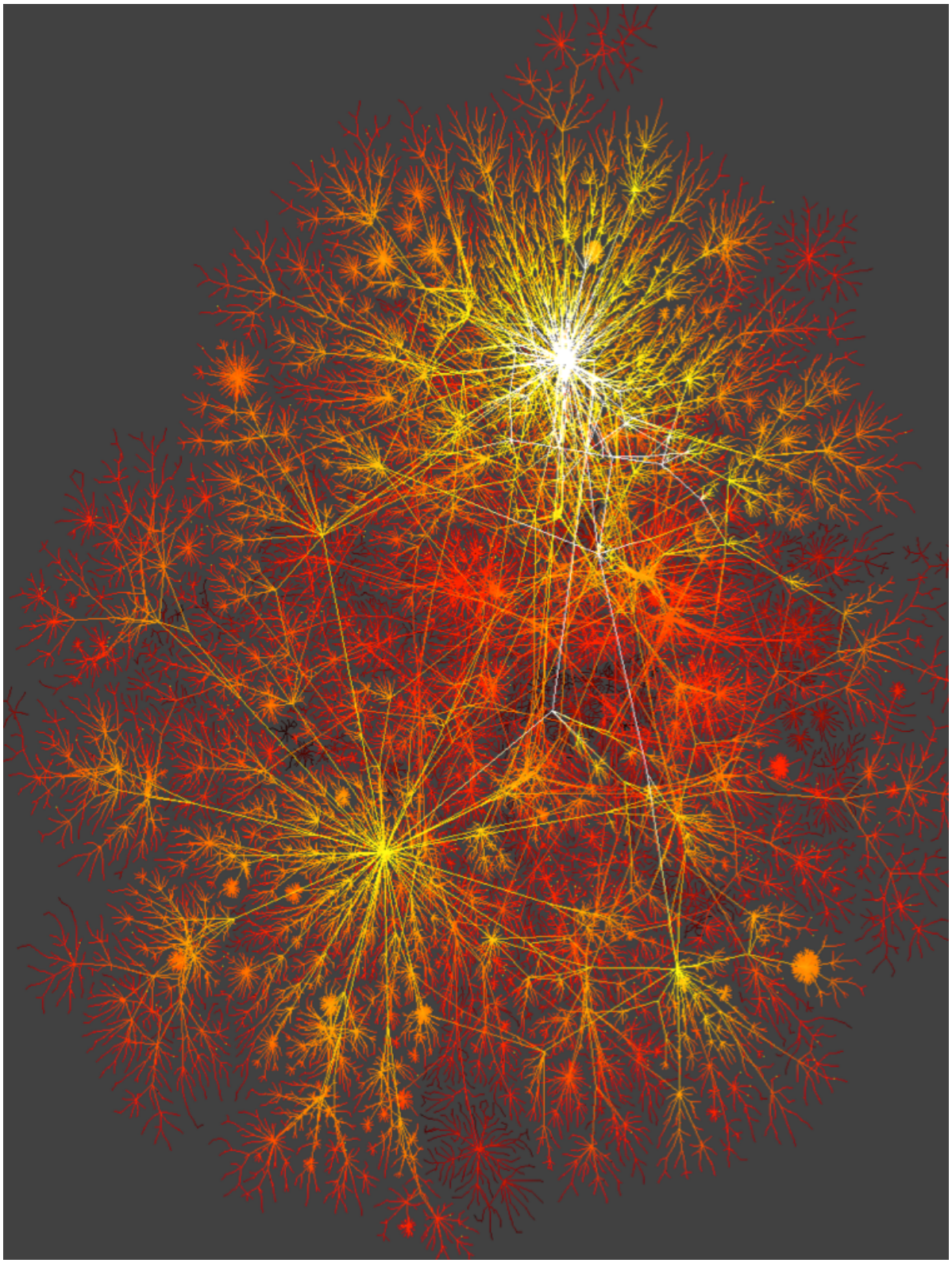


Figure 4. Graphical representation of the internet. Nodes represent the webpages and links represent html jump instructions. In July 2014 it contained some 3.32 billion pages.

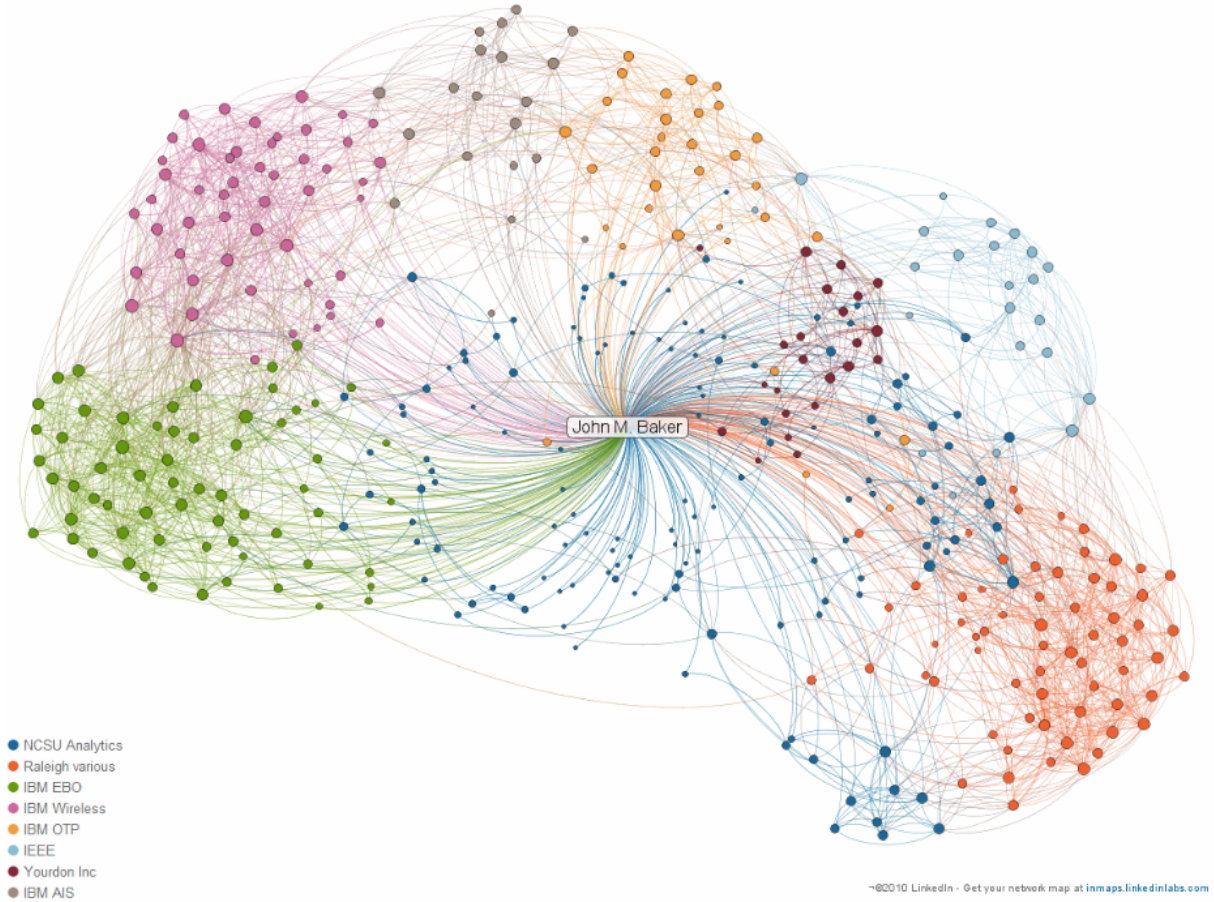


Figure 5. Social network of collaboration partners within an organisation (here: a subset of IBM).

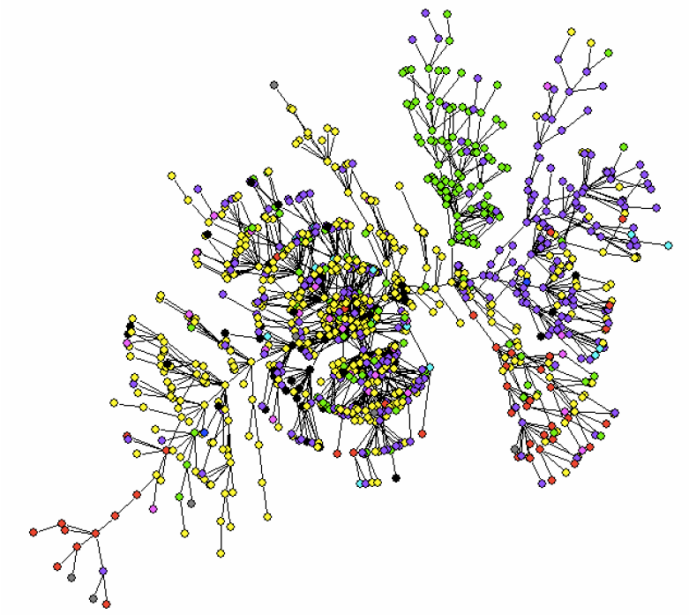


Figure 6. Network representations of interactions observed between share prices. Nodes represent individual companies (with a colour code representing a classification of sectors), and links imply significant observed correlation in the share prices of these companies.

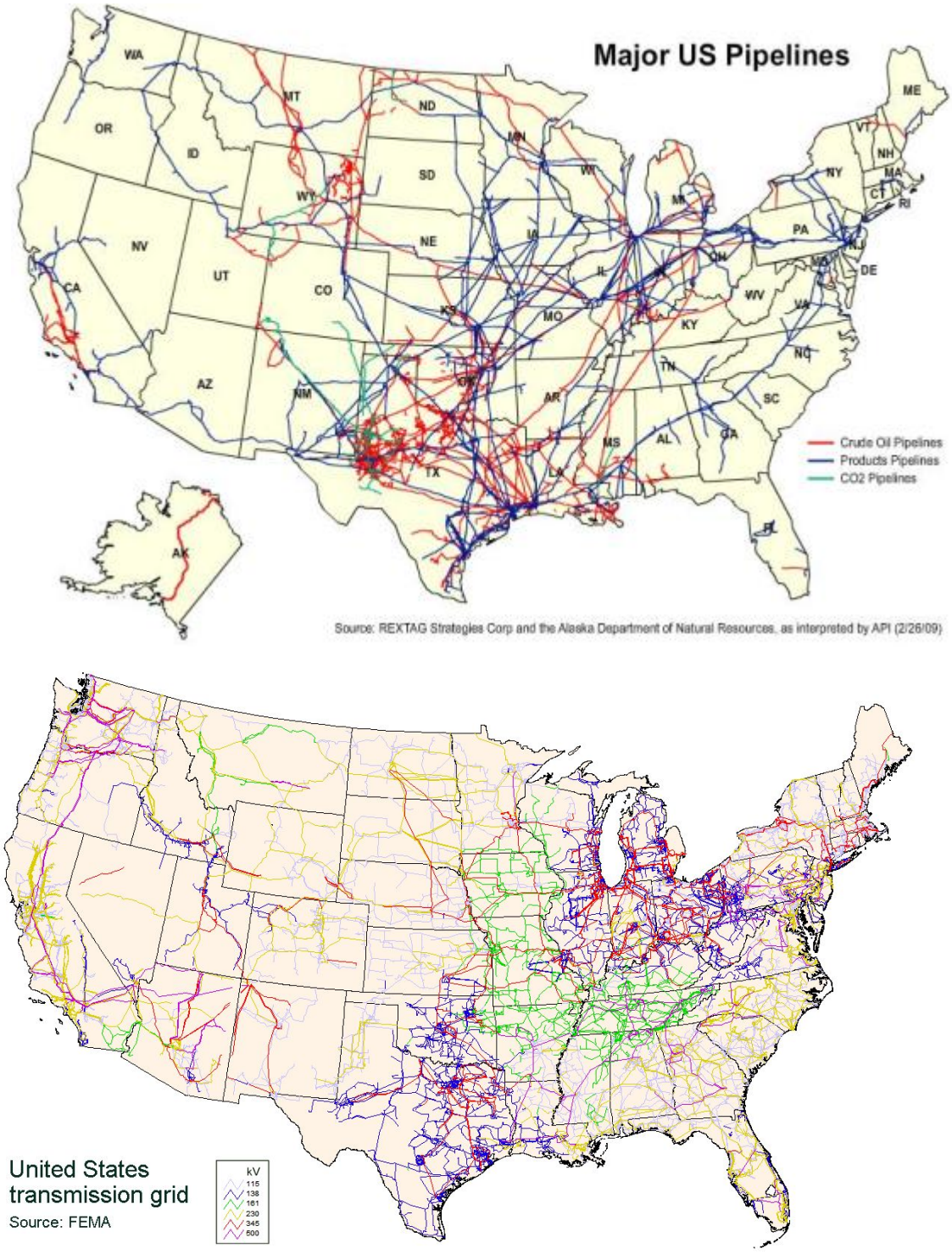


Figure 7. Main arteries of the oil and gas distribution network, and of the national electricity power grid of the USA. Here one would be interested in questions related to the network's vulnerability against targeted attacks, and how to design networks to reduce the damage done by such attacks.

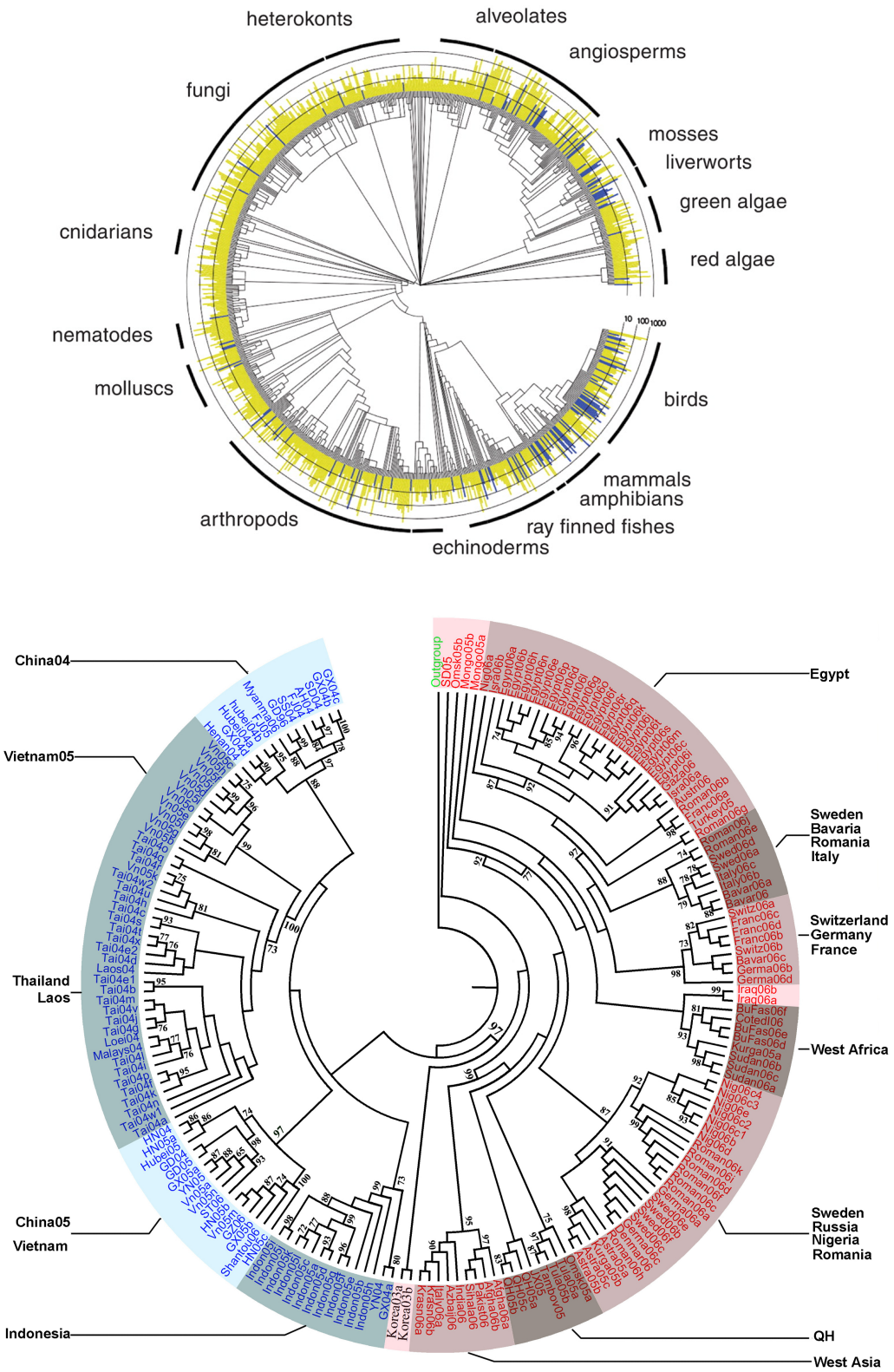


Figure 8. Phylogenetic trees, constructed from genome similarity. Here the nodes represent species of organisms, and links represent the most plausible evolutionary ancestor relationships. Top: general phylogenetic tree showing the main organism families and their genome lengths. Bottom: focus on different strains of human influenza.

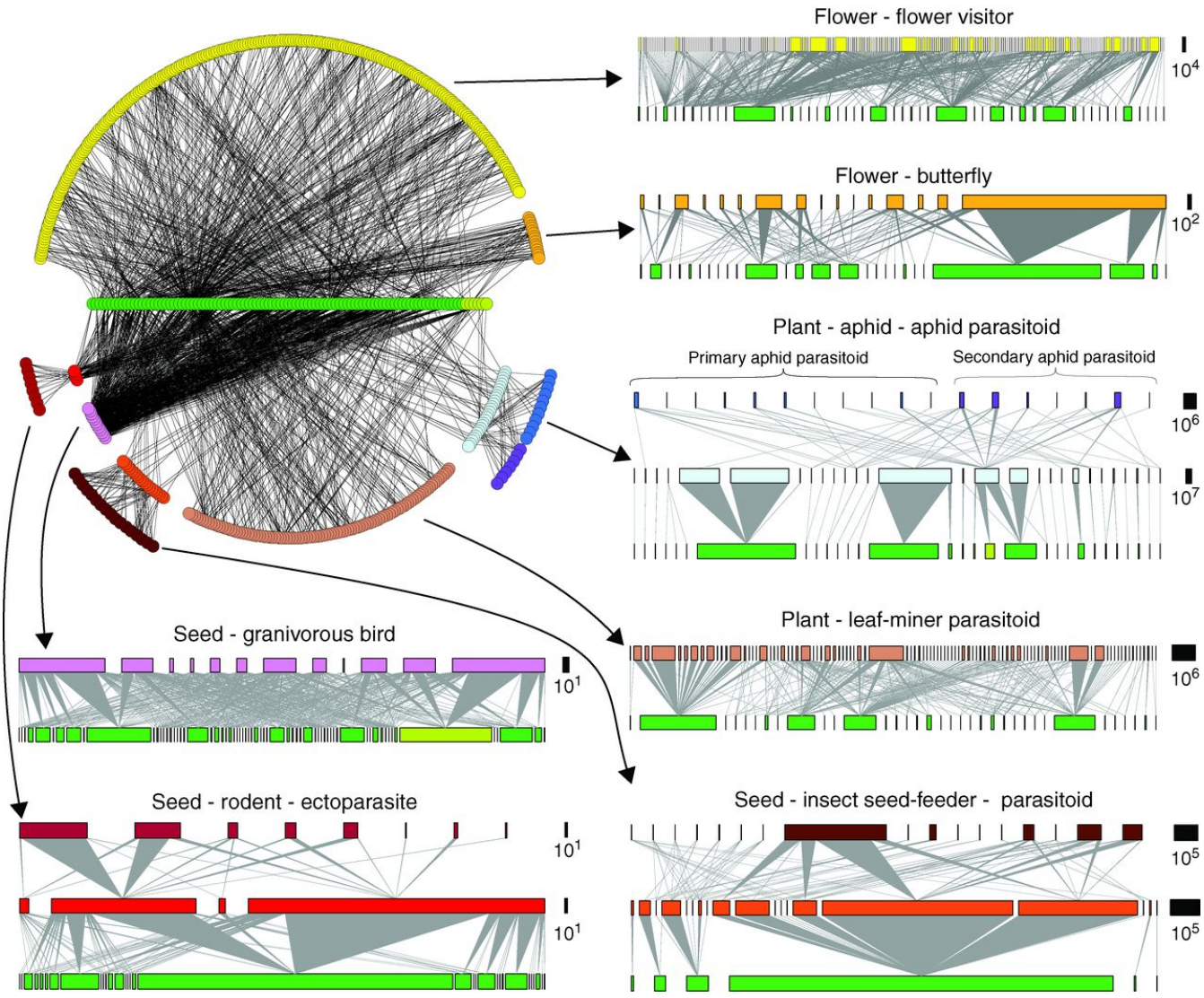
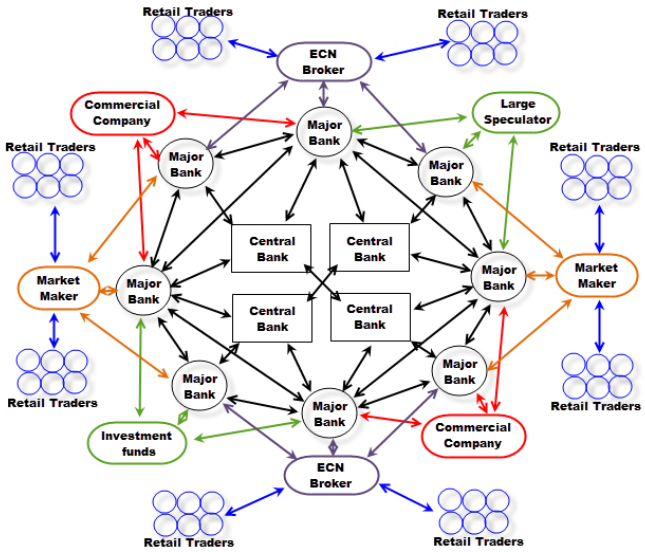


Figure 9. Ecological networks, mapping species and their mutual dependencies in a given area (predator/prey or parasitic relationships, supply of food or other resources, reproduction, etc). The nodes represent the different species of organisms, and links indicate a mutual dependency.



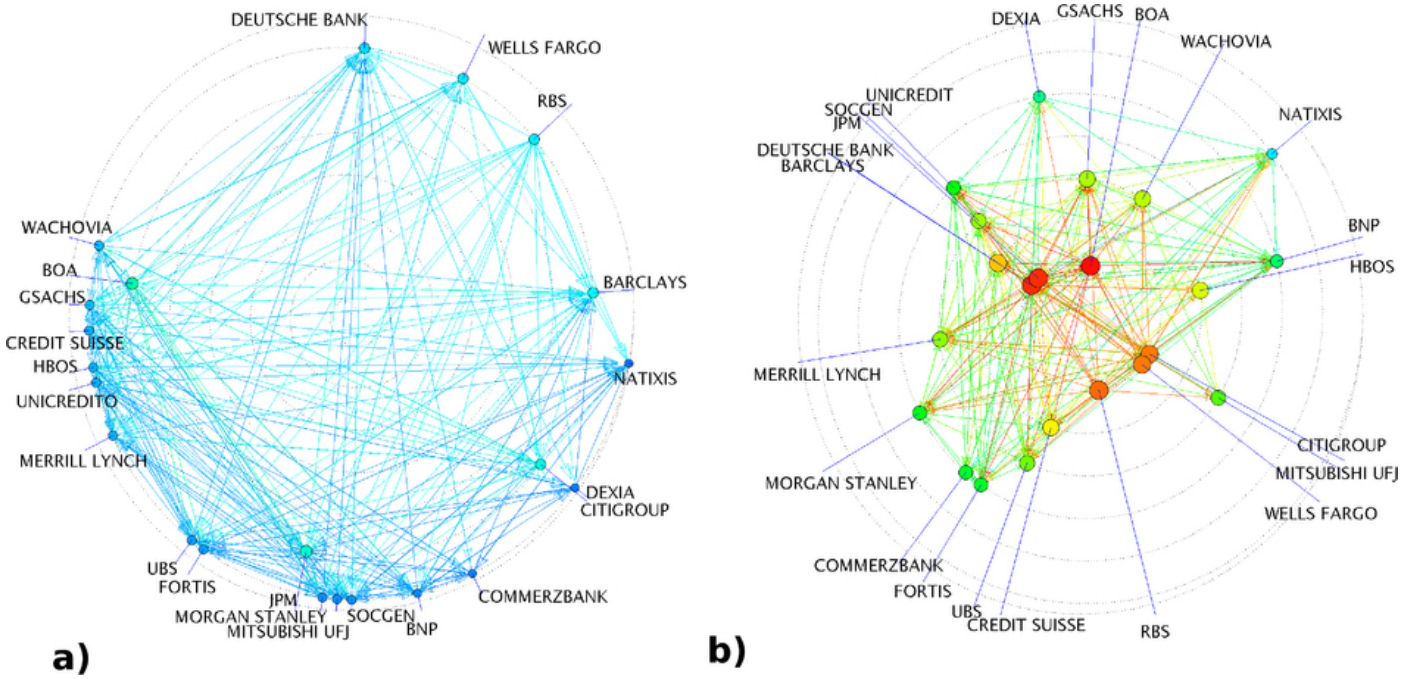


Figure 10. Networks representing economic and financial relationships between the main players in financial markets. It is now increasingly (and painfully) being realised by regulators that the complex interconnected nature of the international financial system means that new mathematical approaches are needed to understand, predict, and prevent future financial crises. Top: the type of players required in models. Bottom: example of big players and their dependencies and relations in the international banking network.

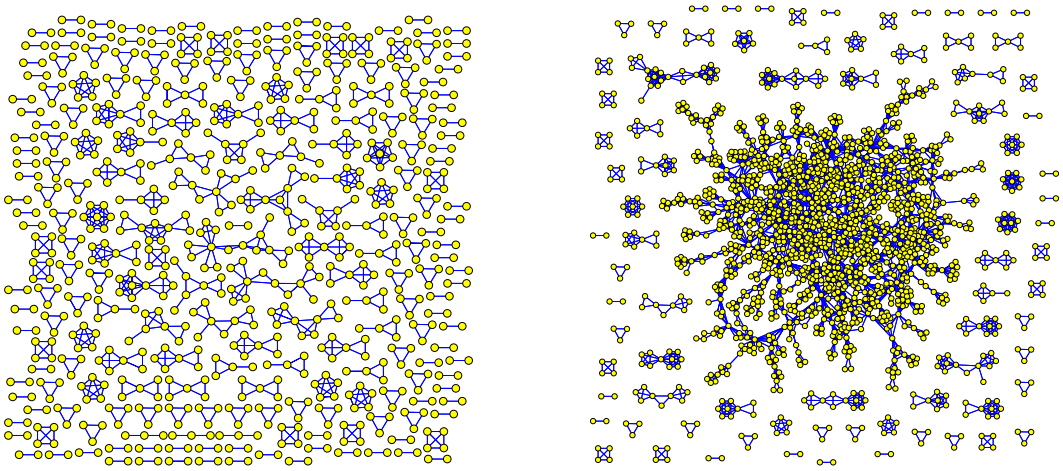


Figure 11. Immune networks. Nodes represent different 'T-clones' (families of immune cells that coordinate the adaptive immune response to specific invaders). Links indicate that the T-clones interact with a common B-clone (the B-clones actually trigger the destruction of the invaders). In humans there are typically around 10^8 T-clones (i.e. nodes).

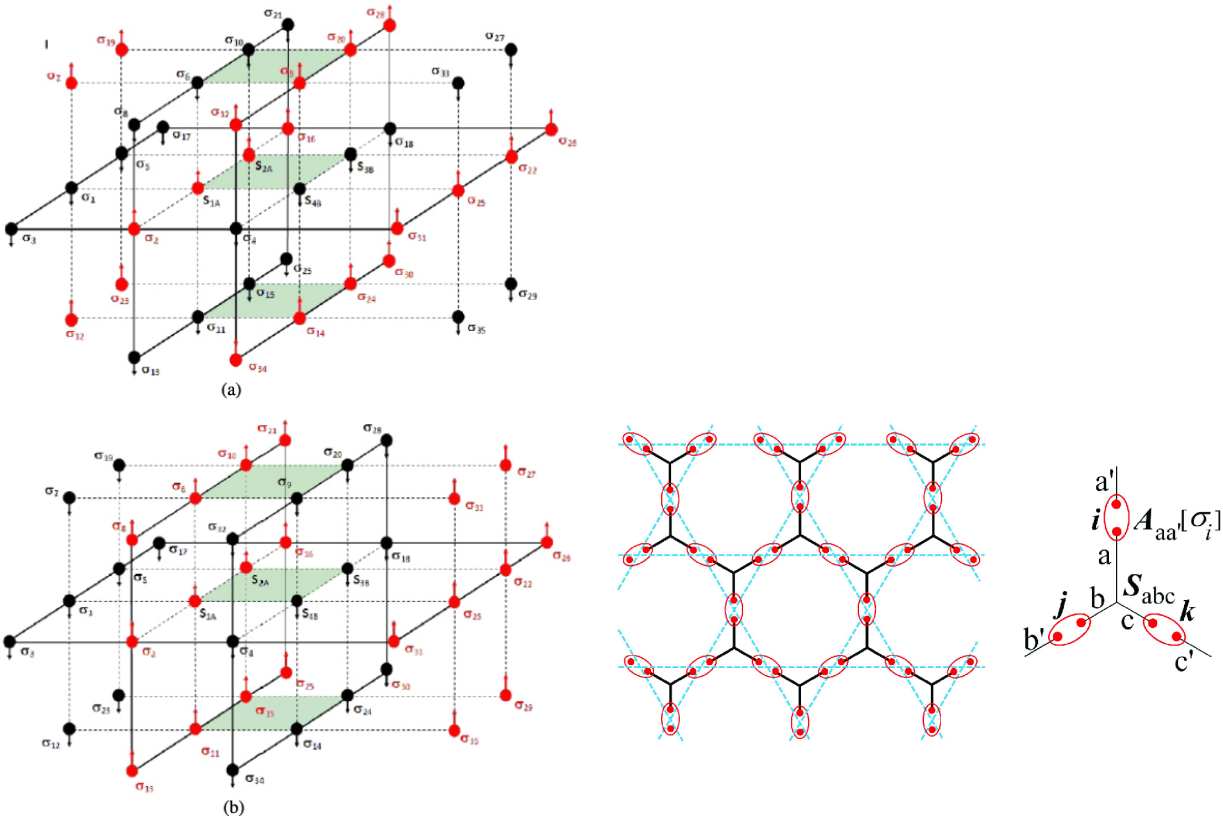


Figure 12. Simplified models of magnetic systems in statistical physics. Here the nodes (of which there are of the order of 10^{24}) represent sites on regular (2-dim or 3-dim) lattices, occupied by atoms with intrinsic magnetic fields. The links indicate which pairs of magnetic atoms are believed to be able to interact and exchange energy.

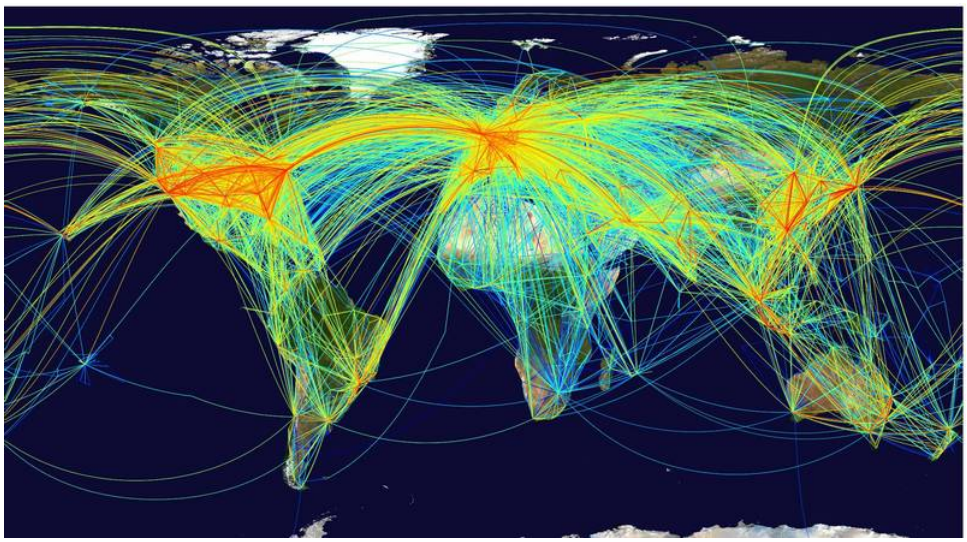


Figure 13. Mobility networks. These are very important in the modelling and prevention of the spread of epidemics. Nodes are the main global population centres, and links represent pairs of population centres with the most intensive traffic of people between them.

Note that network *images* in themselves are subjective. Different software tools will use different protocols for deciding on the most appealing placements of nodes, and hence generally will produce different visual representations of the same network (see e.g. Fig. 14). It is perfectly fine to use graphical representations as intuitive aids, but we will need precise mathematical descriptions when it comes to extracting information and testing hypotheses.

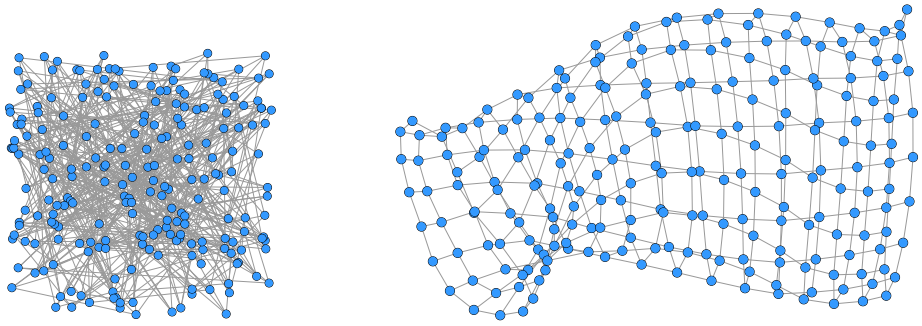


Figure 14. The two graphs shown here may look different, but are in fact topologically identical. They differ only in the choices made for the placements of the nodes in the image plane, i.e. in cosmetics.

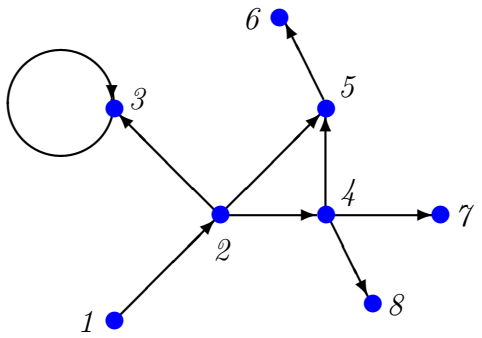
This course will describe some of the mathematical and computational tools that have been developed over the years to characterise and quantify network and graph topologies, quantify the complexity of large networks, identify modules in networks, and to understand the resilience of processes on networks against node or link removal. We also study how to define and generate synthetic networks with controlled topological features.

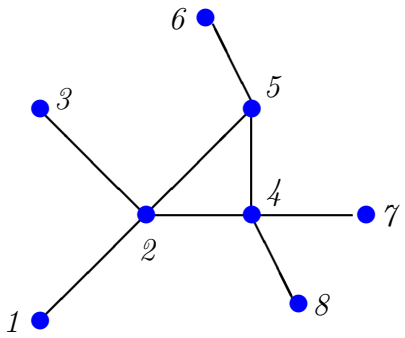
2. Definitions and notation

2.1. Networks or graphs

Some authors use 'networks' to denote the physical objects in the real world, and 'graphs' for their mathematical description. Here we will not make this distinction.

- Definition: an N-node graph $G(V, E)$ is defined by a set of vertices (or nodes) $V = \{1, \ldots, N\}$, and a set of edges (or links) $E \subseteq \{(i, j) |\ i, j \in V\}$
- Definition: a simple graph is a graph without self-links, i.e. $\forall (i, j) \in E :\ i \neq j$
- Definition: a nondirected graph is a graph with symmetric links only, i.e. if $(i, j) \in E$ then also $(j, i) \in E$
- Definition: a directed graph is one that contains non-symmetric links, i.e. $\exists (i, j) \in E$ such that $(j, i) \notin E$





$V = \{1, 2, 3, 4, 5, 6, 7, 8\}$

$E = \{(2, 1), (3, 2), (4, 2), (5, 2), (3, 3), (5, 4), (7, 4), (8, 4), (6, 5)\}$

$V = \{1, 2, 3, 4, 5, 6, 7, 8\}$

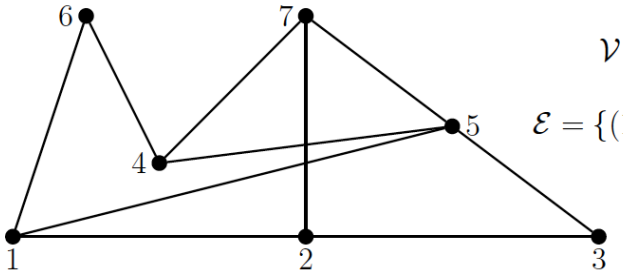
$E = \{(2, 1), (1, 2), (3, 2), (2, 3), (4, 2), (2, 4), (5, 2), (2, 5), (5, 4), (4, 5), (7, 4), (4, 7), (8, 4), (4, 8), (6, 5), (5, 6)\}$

Figure 15. Left: example of a directed graph. It is not simple, since it has a self-link $(3, 3)$ (note that in principle we could leave out arrows when drawing self-links, since there are reciprocal by definition). Right: example of a simple non-directed graph.

- Conventions: in drawing graphs we use the following conventions, see e.g. Fig 15

  (i) a node i is represented by a small filled circle
  (ii) in directed graphs a link (i, j) is drawn as an arrow from node j to node i
  (iii) in simple non-directed graphs a link (i, j), which in such graphs is always accompanied by a link (j, i), is drawn as a line segment between nodes i and j
  (iv) a self-interaction (i, i) is drawn by a small circle that starts at i and ends at i

In situations where in the context of the problem at hand it is clear that we have only non-directed graphs, we leave out the explicit mentioning of both (i, j) and (j, i) and simply give



$$\mathcal{V} = \{1, 2, 3, 4, 5, 6, 7\}$$

$$\mathcal{E} = \{(1,2), (1,6), (1,5), (2,7), (2,3), (3,5), (4,6), (4,5), (4,7), (5,7)\}$$

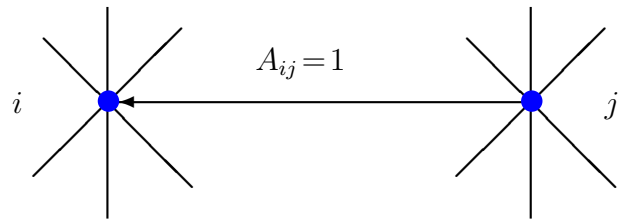
In these lectures we will limit ourselves to non-weighted graphs, i.e. we will *not* consider graphs in which links carry a numerical value to represent their sign or strength. We consider links to be binary objects: they are either present or absent.

2.2. The adjacency matrix of a network

Next we switch to a less cumbersome representation of graphs than sets of links, which will also make subsequent calculations more easy. In an N-node graph there are $N \times N$ potential links, the presence of each of which can be coded by a binary number, which we arrange as the entries of a matrix:

- Definition: the adjacency matrix $\boldsymbol{A} \in \{0, 1\}^{N \times N}$ of an N-node graph $G(V, E)$ is defined by the following entries:

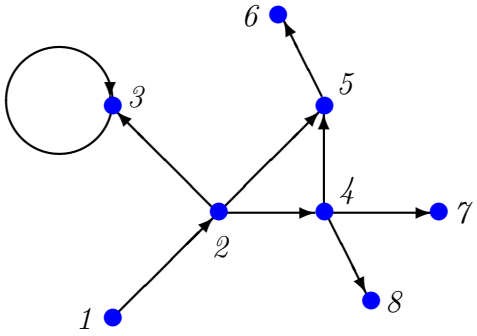
$$\forall (i,j) \in \{1, \ldots, N\}^2 : \quad \begin{array}{ll} A_{ij} = 1 & \text{if } (i,j) \in E, \quad i.e. \quad \text{if there is a link } j \to i \\ A_{ij} = 0 & \text{if } (i,j) \notin E, \quad i.e. \quad \text{if there is no link } j \to i \end{array} \tag{1}$$



- Consequence: a simple N-node graph has an $N \times N$ adjacency matrix with zero diagonal elements, i.e. $A_{ii} = 0 \; \forall i \in \{1, \ldots, N\}$.
- Consequence: a nondirected N-node graph has a symmetric $N \times N$ adjacency matrix, i.e. $A_{ij} = A_{ji} \; \forall (i,j) \in \{1, \ldots, N\}^2$.
- Consequence: a directed N-node graph has a nonsymmetric $N \times N$ adjacency matrix, i.e. $\exists (i,j) \in \{1, \ldots, N\}^2$ such that $A_{ij} \neq A_{ji}$.

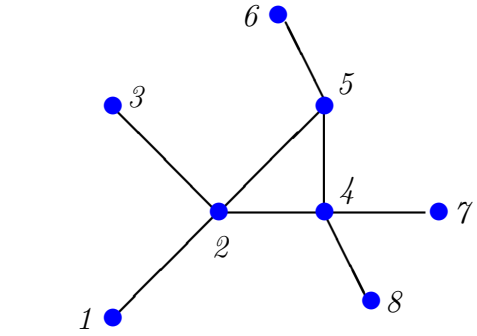
There is a one-to-one correspondence between the N^2 binary entries of $\boldsymbol{A}$ and the specification of which links are present in an N-node graph, so each N-node graph corresponds to a unique adjacency matrix and vice versa. We could therefore equivalently also have defined our graphs and their types (e.g. directed, non directed, simple) on the basis of the adjacency matrices and their properties, instead of starting from the edge and vertex sets.

We can verify that the two network examples of Fig 15 correspond to the following two adjacency matrices:



corresponds to $\boldsymbol{A} =$

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



corresponds to $\boldsymbol{A} =$

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

We observe indeed that the second adjacency matrix is symmetric (as it corresponds to a non-directed graph), in contrast to the first.

2.3. Paths in networks

Consider products of a graph's adjacency matrix entries of the following form, with $i, j \in \{1, \ldots, N\}$, and with $i_\ell \in \{1, \ldots, N\}$ for all ℓ:

$$\prod_{\ell=1}^{k-1} A_{i_\ell i_{\ell+1}} = \overbrace{A_{i_1 i_2} A_{i_2 i_3} A_{i_3 i_4} \ldots A_{i_{k-2} i_{k-1}} A_{i_{k-1} i_k}}^{k-1 \ \ factors} \tag{2}$$

Since each individual factor is either 0 or 1 we must conclude that

$$\prod_{\ell=1}^{k-1} A_{i_\ell i_{\ell+1}} = 1 \;\; if \;\; A_{i_\ell i_{\ell+1}} = 1 \;\; \forall \ell \in \{1, \ldots, k-1\} \tag{3}$$

$$\prod_{\ell=1}^{k-1} A_{i_\ell i_{\ell+1}} = 0 \;\; otherwise \tag{4}$$

But this implies

$$\prod_{\ell=1}^{k-1} A_{i_\ell i_{\ell+1}} = 1 \;\; if \; the \; graph \; contains \; the \; path \; of \; connected \; links \; i_k \to i_{k-1} \to \ldots \to i_2 \to i_1 \tag{5}$$

$$\prod_{\ell=1}^{k-1} A_{i_\ell i_{\ell+1}} = 0 \;\; if \; it \; does \; not \tag{6}$$

For instance, in the first (directed) graph of Fig 15 we have

$A_{54}A_{42} = 1:$ *the graph contains the path* $2 \to 4 \to 5$

$A_{45}A_{52} = 0:$ *the graph does not contain the path* $2 \to 5 \to 4$

$A_{65}A_{54}A_{42}A_{21} = 1:$ *the graph contains the path* $1 \to 2 \to 4 \to 5 \to 6$

$A_{76}A_{65}A_{54}A_{42} = 0:$ *the graph does not contain the path* $2 \to 4 \to 5 \to 6 \to 7$

and so on.

- Definition: a closed path (or cycle) is a path that starts and ends at the same node, so

$$\prod_{\ell=1}^{k-1} A_{i_\ell i_{\ell+1}} = 1 \;\; with \; i_1 = i_k \;\; if \; the \; graph \; contains \; the \; cycle \; i_1 \to i_{k-1} \to \ldots \to i_2 \to i_1 \tag{7}$$

$$\prod_{\ell=1}^{k-1} A_{i_\ell i_{\ell+1}} = 0 \;\; with \; i_1 = i_k \;\; if \; it \; does \; not \tag{8}$$

- Definition: a simple cycle in a graph is one that contains no repeated vertices or edges, i.e. $\prod_{\ell=1}^{k-1} A_{i_\ell i_{\ell+1}} = 1$ *with* $i_1 = i_k$ and all node labels in $\{i_1, \ldots, i_{k-1}\}$ are nonidentical.

We can also ask whether there is *any* path of a specified length from a given initial node j to a specified target node i. Now we do not care what exactly are the intermediate nodes visited in between i and j. We are obviously interested in the presence or absence of paths of length 2 or more, since having a path of length 1 means simply that $A_{ij} = 1$ (which can be read off directly from the adjacency matrix):

$$\sum_{i_1=1}^{N} \ldots \sum_{i_k=1}^{N} A_{ii_1} \Big(\prod_{\ell=1}^{k-1} A_{i_\ell i_{\ell+1}} \Big) A_{i_k j} > 0 \;\; \Leftrightarrow \;\; there \; exists \; at \; least \; one \; path \; of \; length \; k+1 \; from \; node \; j \; to \; node \; i \tag{9}$$

$$\sum_{i_1=1}^{N} \ldots \sum_{i_k=1}^{N} A_{ii_1} \Big(\prod_{\ell=1}^{k-1} A_{i_\ell i_{\ell+1}} \Big) A_{i_k j} = 0 \;\; \Leftrightarrow \;\; there \; exists \; no \; path \; of \; lenght \; k+1 \; from \; node \; j \; to \; node \; i \tag{10}$$

But since the summations over the indices $\{i_1, \ldots, i_k\}$ in the latter formulae are equivalent to doing matrix multiplications, we can simplify these formulae. We remember the definitions of matrix multiplication and powers of matrices, e.g.

$$(\boldsymbol{AB})_{ij} = \sum_{r=1}^{N} A_{ir} B_{rj}, \qquad (\boldsymbol{A}^0)_{ij} = \delta_{ij}, \qquad (\boldsymbol{A}^{k+1})_{ij} = \sum_{r=1}^{N} (\boldsymbol{A}^k)_{ir} A_{rj}$$

(with the Kronecker δ-symbol, defined as $\delta_{ii} = 1$ for all i and $\delta_{ij} = 0$ for all index pairs $i \neq j$), and we conclude from these that the following is true

$$(\boldsymbol{A}^{k+1})_{ij} > 0 \quad \Leftrightarrow \quad \textit{there exists at least one path of lenght } k+1 \textit{ from node } j \textit{ to node } i \tag{11}$$

$$(\boldsymbol{A}^{k+1})_{ij} = 0 \quad \Leftrightarrow \quad \textit{there exists no path of lenght } k+1 \textit{ from node } j \textit{ to node } i \tag{12}$$

This shows already the benefit of working with adjacency matrices as opposed to the sets (V, E) of nodes and links; for large N it would become painful to trace lines in images or match entries in sets of index pairs, but instead we can simply do matrix multiplication.

Finally, the last step will not come as a surprise. If we don't care about path lengths but only ask about connectivity, we may write

$$\sum_{k\geq 0} (\boldsymbol{A}^{k+1})_{ij} > 0 \quad \Leftrightarrow \quad \textit{there exists at least one path from } j \textit{ to } i \tag{13}$$

$$\sum_{k\geq 0} (\boldsymbol{A}^{k+1})_{ij} = 0 \quad \Leftrightarrow \quad \textit{there exists no path from } j \textit{ to } i$$

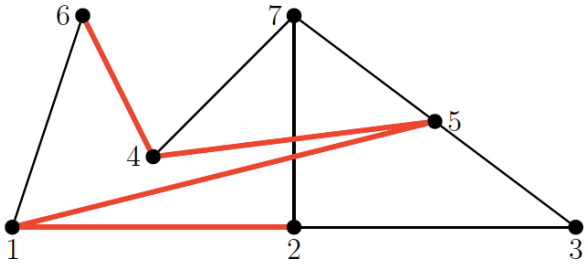
Two final definitions in relation to paths in graphs with special properties:

- Definition: an Eulerian path in a graph is one in which each link (or edge) is traversed exactly once.
- Definition: a Hamiltonian path in a graph is one in which each node (or vertex) is visited exactly once.

2.4. Graphs within graphs

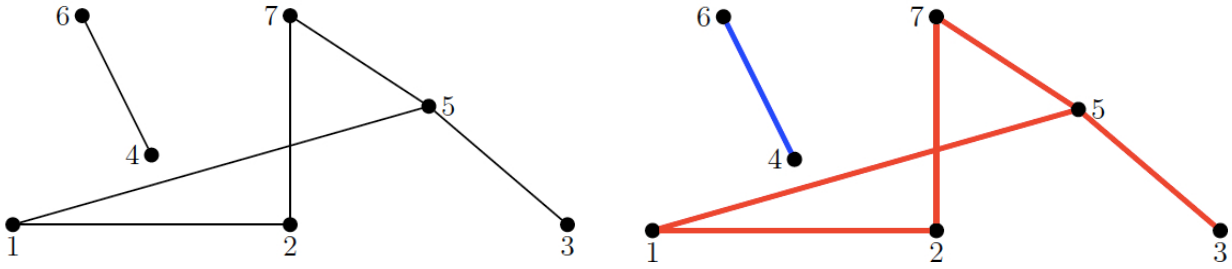
- Definition: a subgraph $G'(V', E')$ of a graph $G(V, E)$ is a graph such that $V' \subseteq V$ and $E' \subseteq E$.

For instance, the red graph below has $V' = \{1, 2, 4, 5, 6\}$ and $E' = \{(1, 2), (2, 1), (1, 5), (5, 1), (4, 5), (5, 4), (4, 6), (6, 4)\}$, and is clearly a subgraph of the one at the top of page 15.

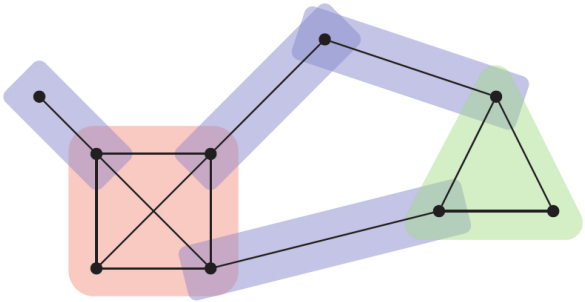


- Definition: the connected components of a graph G are the largest sub-graphs of G such that for each subgraph there exists a path between all vertices within the subgraph.

For instance, the graph in black on the left has the connected components shown in blue and red on the right:



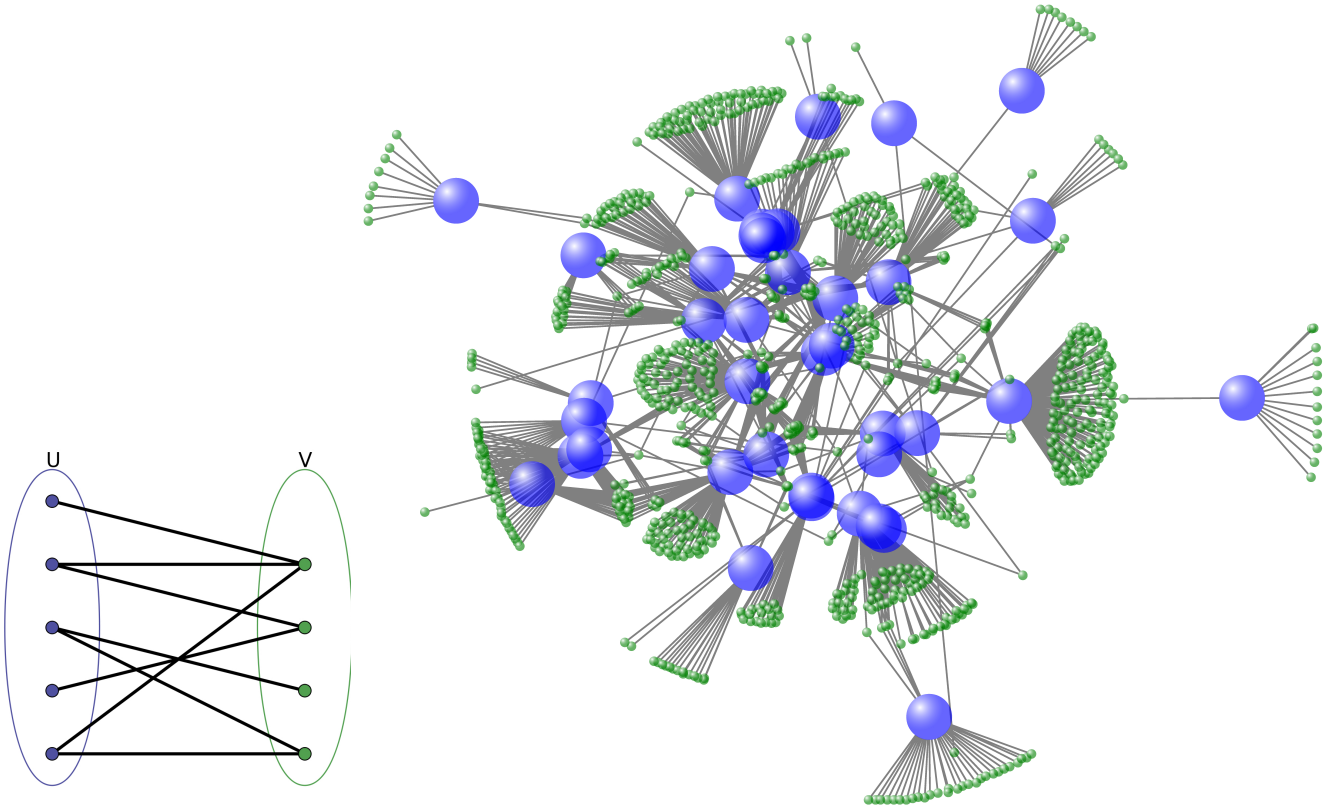
- Definition: a clique in a graph G is a maximal subset $V' \subseteq V$ of vertices in the graph such that every member of the subset has an edge connecting to every other. Here 'maximal' means that it is impossible to add any further edge to V' such that the new edge connects to all previous edges in V'.



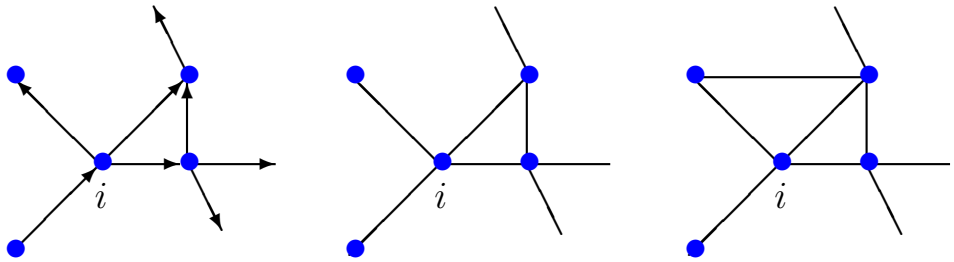
In this graph we all cliques of size 2, 3 and 4 shown respectively in purple, green and pink. Note: the immune networks in Figure 11 consist strictly of connected cliques.

- Definition: a bipartite graph G is one in which the vertices can be divided into two nonempty disjoint subsets, i.e. $V = V_1 \cup V_2$ with $|V_1|, |V_2| > 0$ and $V_1 \cap V_2 = \emptyset$, such that all edges $(i, j) \in E$ have either $i \in V_1$ and $j \in V_2$ or $j \in V_1$ and $i \in V_2$.

The nodes in bipartite graphs can be divided into two qualitatively different groups, and there are no links between indices in the same group. Examples are networks that represent sexual relations in heterosexual groups, or graphs that represent relations between diseases (node group 1) and clinical features (node group 2), networks mapping researchers and the journals in which they publish, networks of resource generators and resource consumers, etc:



Generalisations to tripartite graphs (three disjunct vertex subsets, with links only between nodes from different subsets) and to higher order partitions are obvious.



$k_i^{\text{in}}(\boldsymbol{A}) = 1$
$k_i^{\text{out}}(\boldsymbol{A}) = 3$

$k_i(\boldsymbol{A}) = 4$
$C_i(\boldsymbol{A}) = 1/6$
$T_i(\boldsymbol{A}) = 1$

$k_i(\boldsymbol{A}) = 4$
$C_i(\boldsymbol{A}) = 2/6 = 1/3$
$T_i(\boldsymbol{A}) = 2$

Figure 16. Left: in- and out degrees $k_i^{\text{in}}(\boldsymbol{A})$ and $k_i^{\text{out}}(\boldsymbol{A})$, i.e. the number of arrows flowing into and out of the given node, in directed graphs. Middle and right: degrees $k_i(\boldsymbol{A})$, triangle counters $T_i(\boldsymbol{A})$, and clustering coefficients $C_i(\boldsymbol{A})$ in non-directed graphs. $C_i(\boldsymbol{A})$ gives the fraction of distinct neighbour pairs of i that are themselves connected. In the absence of link directionality, there is no distinction in nondirected graphs between left- and right-degrees.

3. Microscopic structural characteristics of graphs

3.1. Node-specific quantities

Node degrees. To characterize graph topologies more intuitively, we first inspect simple quantities that inform us about their structure in the vicinity of individual nodes. The first of these are the so-called in- and out-degrees $k_i^{\text{out}}(\boldsymbol{A}) \in \mathbb{N}$ and $k_i^{\text{out}}(\boldsymbol{A}) \in \mathbb{N}$ of each node i in graph $\boldsymbol{A}$. They count, respectively, the number of arrows flowing into and out of node i:

- Definition: the in-degree of node i in an N-node graph with adjacency matrix $\boldsymbol{A}$ is defined as $k_i^{\text{in}}(\boldsymbol{A}) = \sum_{j=1}^{N} A_{ij}$
- Definition: the out-degree of node i in an N-node graph with adjacency matrix $\boldsymbol{A}$ is defined as $k_i^{\text{out}}(\boldsymbol{A}) = \sum_{j=1}^{N} A_{ji}$

We denote the pair of in- and out-degrees for a node i as $\vec{k}_i(\boldsymbol{A}) = (k_i^{\text{in}}(\boldsymbol{A}), k_i^{\text{out}}(\boldsymbol{A})) \in \mathbb{N}^2$. In nondirected graphs we find that always $k_i^{\text{in}}(\boldsymbol{A}) = k_i^{\text{out}}(\boldsymbol{A})$ (see exercises). Here we can drop the superscripts and simply refer to 'the degree' of a node:

- Definition: the degree of node i in a non-directed N-node graph with adjacency matrix $\boldsymbol{A}$ is defined as $k_i(\boldsymbol{A}) = \sum_j A_{ij}$.
- Definition: the degree *sequence* of a non-directed N-node graph with adjacency matrix $\boldsymbol{A}$ is defined as the vector $(k_1(\boldsymbol{A}), k_2(\boldsymbol{A}), \ldots, k_N(\boldsymbol{A})) \in \mathbb{N}^N$.

Clustering coefficients and closed path counters. There are many ways to characterise a graph's local structure beyond counting the neighbours of a node. For simple nondirected

graphs, the clustering coefficient gives the fraction of node pairs linked to i that are themselves connected:

- Definition: the clustering coefficient $C_i(\boldsymbol{A})$ of node i with degree ≥ 2 in a non-directed N-node graph with adjacency matrix $\boldsymbol{A}$ is defined as

$$C_i(\boldsymbol{A}) = \frac{\text{number of } \textit{connected} \text{ node pairs among neighbours of } i}{\text{number of node pairs among neighbours of } i}$$
$$= \frac{\sum_{j,k=1}^{N}(1-\delta_{jk})A_{ij}A_{jk}A_{ik}}{\sum_{j,k=1}^{N}(1-\delta_{jk})A_{ij}A_{ik}} \in [0,1] \tag{14}$$

  (for nodes i with degree 0 or 1 we simply define $C_i(\boldsymbol{A}) = 0$).

We have already seen that products of entries of the adjacency matrix of a graph can be used to identify paths. We can use this to count the numbers of closed paths of a given length:

- Claim: the number $L_\ell(\boldsymbol{A})$ of closed paths of length $\ell > 0$ in an N-node graph with adjacency matrix $\boldsymbol{A}$ (directed or non-directed) is given by

$$L_\ell(\boldsymbol{A}) = \sum_{i_1=1}^{N} \cdots \sum_{i_\ell=1}^{N} \Big(\prod_{k=1}^{\ell-1} A_{i_k,i_{k+1}} \Big) A_{i_\ell,i_1} = \sum_{i=1}^{N} (\boldsymbol{A}^\ell)_{ii} \tag{15}$$

This follows directly from our earlier identities on paths. Note that the sum of the diagonal entries of a matrix is called its *trace*, $\mathrm{Tr}(\boldsymbol{B}) = \sum_i B_{ii}$, so we have $L_\ell(\boldsymbol{A}) = \mathrm{Tr}(\boldsymbol{A}^\ell)$.

- Definition: the number of triangles $T_i(\boldsymbol{A})$ involving node i in a non-directed simple N-node graph with adjacency matrix $\boldsymbol{A}$ is defined as $T_i(\boldsymbol{A}) = \frac{1}{2}\sum_{j,k=1}^{N} A_{ij}A_{jk}A_{ki} \in \mathbb{N}$.

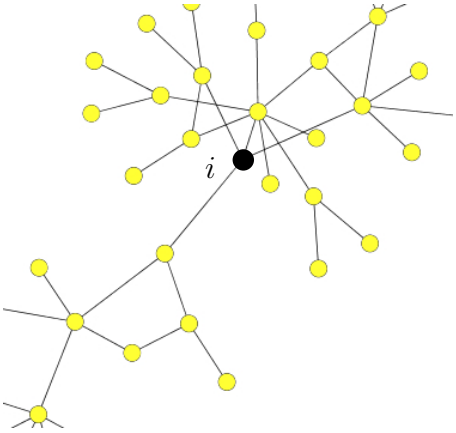
$T_i(\boldsymbol{A})$, which can also be written as $T_i(\boldsymbol{A}) = \frac{1}{2}(\boldsymbol{A}^3)_{ii}$, counts the number of distinct nondirected loops of length three, in which node i participates. The factor $\frac{1}{2}$ in $T_i(\boldsymbol{A})$ corrects for overcounting: any triangle starting and ending in node i can be drawn with two possible orientations. Note that in simple non-directed graphs one has $C_i(\boldsymbol{A}) = 2T_i(\boldsymbol{A})/k_i(\boldsymbol{A})[k_i(\boldsymbol{A}) - 1]$ (see exercises). In Fig. 16 we illustrate the various node characteristics with some simple examples.

3.2. Generalised degrees

The concept of degrees can be generalised in obvious ways. For instance, the generalised degrees of order $\ell \geq 1$ count the number of distinct paths of a given length ℓ that either flow out of, or into a node i:

$$k_i^{(\ell)\text{in}}(\boldsymbol{A}) = \sum_{j_1=1}^{N} \cdots \sum_{j_\ell=1}^{N} A_{ij_1}A_{j_1j_2}\ldots A_{j_{\ell-1}j_\ell} \tag{16}$$

$$k_i^{(\ell)\text{out}}(\boldsymbol{A}) = \sum_{j_1=1}^{N} \cdots \sum_{j_\ell=1}^{N} A_{j_\ell j_{\ell-1}}\ldots A_{j_2j_1}A_{j_1i} \tag{17}$$



$k_i^{(1)}(\boldsymbol{A}) = 4$

$k_i^{(2)}(\boldsymbol{A}) = 20$

Figure 17. Degrees and generalised degrees in non-directed graphs. At the minimal level one specifies for each node i (black vertex) only the degree $k_i(\boldsymbol{A}) = k_i^{(1)}(\boldsymbol{A}) = \sum_j A_{ij}$ (the number of its neighbours). Next one provides for each node the pair $(k_i^{(1)}(\boldsymbol{A}), k_i^{(2)}(\boldsymbol{A}))$, in which $k_i^{(2)}(\boldsymbol{A}) = \sum_{j\ell} A_{ij}A_{j\ell}$ is the number of length-two paths ending in i. Since the graph at hand here is non-directed, we can drop the superscripts 'in' and 'out'.

For $\ell = 1$ these formulae reduce to the previous expressions for ordinary in- and out-degrees. We can similarly introduce quantities that generalise the triangle counters $T_i(\boldsymbol{A})$ by counting closed paths in the graph $\boldsymbol{A}$ of arbitrary lengths $\ell \geq 3$, that pass through node i.

3.3. Quantities related to pairs of nodes

The distance between two nodes. We define the distance $d_{ij}(\boldsymbol{A})$ between nodes i and j in a non-directed graph with adjacency matrix $\boldsymbol{A}$ as the length of the *shortest* path from node i to node j. It can be expressed in many ways, e.g. upon using our earlier formulae involving paths:

- Definition: the distance $d_{ij}(\boldsymbol{A})$ between nodes i and j is defined as follows

$$if\ there\ is\ no\ path\ from\ j\ to\ i: \quad d_{ij}(\boldsymbol{A}) = \infty \tag{18}$$

$$if\ there\ is\ a\ path\ from\ j\ to\ i: \quad d_{ij}(\boldsymbol{A}) = \ smallest\ \ell \geq 0\ such\ that\ (\boldsymbol{A}^\ell)_{ij} > 0 \tag{19}$$

Note 1: this distance is not the same as the distance between nodes in an *image* of the graph, it is only based on how many links need to be crossed when walking from i to j.

Note 2: a path along which the shortest distance between two nodes is realised (there could be more than one) is also called a *geodesic*.

An alternative way to obtain the distances between nodes in the graphs, without checking one by one all possible routes from i to j, is provided by the following identity. Here the inverse $\boldsymbol{C}^{-1}$ of an $N \times N$ matrix $\boldsymbol{C}$ (if it exists) is the unique matrix with the property $\boldsymbol{C}\boldsymbol{C}^{-1} = \boldsymbol{C}^{-1}\boldsymbol{C} = \mathbb{1}$, in which $\mathbb{1}$ denotes the $N \times N$ identity matrix.

- Claim:
$$d_{ij} = \lim_{\gamma \downarrow 0} \frac{\log[(\mathbb{I} - \gamma \boldsymbol{A})^{-1}]_{ij}}{\log \gamma} \tag{20}$$

  Proof:

  We first note that $(\mathbb{I} - \gamma \boldsymbol{A})^{-1} = \sum_{\ell \geq 0} \gamma^\ell \boldsymbol{A}^\ell$ (see exercises). This series is always convergent for sufficiently small γ. It then follows that

$$\begin{aligned}(\mathbb{I} - \gamma \boldsymbol{A})^{-1}]_{ij} &= \sum_{\ell \geq 0} \gamma^\ell (\boldsymbol{A}^\ell)_{ij} = \sum_{\ell \geq d_{ij}} \gamma^\ell (\boldsymbol{A}^\ell)_{ij} = \gamma^{d_{ij}} (\boldsymbol{A}^{d_{ij}})_{ij} + \mathcal{O}(\gamma^{d_{ij}+1}) \\ &= \gamma^{d_{ij}} \Big[(\boldsymbol{A}^{d_{ij}})_{ij} + \mathcal{O}(\gamma) \Big]\end{aligned}$$

  Hence

$$\begin{aligned}\lim_{\gamma \downarrow 0} \frac{\log[(\mathbb{I} - \gamma \boldsymbol{A})^{-1}]_{ij}}{\log \gamma} &= \lim_{\gamma \to 0} \frac{\log \gamma^{d_{ij}} + \log \Big[(\boldsymbol{A}^{d_{ij}})_{ij} + \mathcal{O}(\gamma) \Big]}{\log \gamma} \\ &= d_{ij} + \lim_{\gamma \to 0} \frac{\log \Big[(\boldsymbol{A}^{d_{ij}})_{ij} + \mathcal{O}(\gamma) \Big]}{\log \gamma} \\ &= d_{ij} - \lim_{x \to \infty} \frac{1}{x} \log \Big[(\boldsymbol{A}^{d_{ij}})_{ij} + \mathcal{O}(\mathrm{e}^{-x}) \Big] = d_{ij} \qquad []\end{aligned}$$

The matrix inversion is usually impossible analytically, so is done numerically (see exercises).

Node centrality. To quantify how important an individual node i may be to sustain traffic or flow of information over a graph, two measures of 'centrality' have been defined: the *closeness centrality* and the *betweenness centrality*.

- Definition: the average distance $d_i(\boldsymbol{A})$ to node i is $d_i(\boldsymbol{A}) = N^{-1} \sum_{j=1}^N d_{ij}(\boldsymbol{A})$.
- Definition: the closeness centrality $x_i(\boldsymbol{A})$ of node i is defined as $x_i(\boldsymbol{A}) = 1/d_i(\boldsymbol{A})$.‡
- Definition: the betweenness centrality $y_i(\boldsymbol{A})$ of a node is defined as the number of node pairs (k, ℓ), with $k \neq \ell \neq i$, such that i lies on a shortest path between k and ℓ.§

Nodes with a high closeness centrality have small typical distances to the other nodes, and are hence relatively close to any area of the graph. Nodes with a high betweenness centrality are apparently important relay stations that reduce the shortest path lengths between node pairs in the graph. They need not be on average close to the other nodes, but tend to be the pivotal nodes that connect otherwise separate parts of the graph.

Similarity between node pairs. The functional role of any node i in an N-node graph with adjacency matrix $\boldsymbol{A}$ is defined strictly by the specification of the links that flow into or out

‡ This definition is helpful and makes sense (and is therefore used) only for connected graphs, since otherwise $d_i(\boldsymbol{A}) = \infty$ (due to the appearance of node pairs that give $d_{ij}(\boldsymbol{A}) = \infty$).
§ Sometimes this quantity is normalised by the total number $\frac{1}{2}(N-1)(N-2)$ (in non directed graphs) of node pairs not including i.

of it, i.e. by giving the two sets

$$\partial_i^{\rm in} = \{k \leq N |\ A_{ik} = 1\}, \qquad \partial_i^{\rm out} = \{k \leq N |\ A_{ki} = 1\} \tag{21}$$

In non directed graphs these two sets are identical for all i, so there we would simply speak of the neighbourhood ∂_i (without superscripts). Hence, any measure of similarity of nodes i and j will somehow quantify the differences between $(\partial_i^{\rm in}, \partial_i^{\rm out})$ and $(\partial_j^{\rm in}, \partial_j^{\rm out})$. Here we will show two common definitions for non directed graphs (possible generalisations to directed graphs are obvious), with $|S|$ denoting the number of elements in the set S:

- Definition: the cosine similarity between nodes i and j with nonzero degrees in a nondirected N-node graph with adjacency matrix $\boldsymbol{A}$ is defined as

$$\sigma_{ij}(\boldsymbol{A}) = \frac{|\partial_i \cap \partial_j|}{\sqrt{|\partial_i||\partial_j|}} = \frac{\sum_{k=1}^N A_{ik}A_{jk}}{\sqrt{k_i(\boldsymbol{A})k_j(\boldsymbol{A})}} \tag{22}$$

- Definition: the Pearson correlation similarity between nodes i and j with nonzero degrees in a nondirected N-node graph with adjacency matrix $\boldsymbol{A}$ is defined as

$$\begin{aligned}\tau_{ij}(\boldsymbol{A}) &= \frac{\frac{1}{N}\sum_{k=1}^N A_{ik}A_{jk} - \left(\frac{1}{N}\sum_{k=1}^N A_{ik}\right)\left(\frac{1}{N}\sum_{k=1}^N A_{jk}\right)}{\sqrt{\frac{1}{N}\sum_{k=1}^N A_{ik}^2 - \left(\frac{1}{N}\sum_{k=1}^N A_{ik}\right)^2}\sqrt{\frac{1}{N}\sum_{k=1}^N A_{jk}^2 - \left(\frac{1}{N}\sum_{k=1}^N A_{jk}\right)^2}} \\ &= \frac{\sum_{k=1}^N A_{ik}A_{jk} - \frac{1}{N}k_i(\boldsymbol{A})k_j(\boldsymbol{A})}{\sqrt{k_i(\boldsymbol{A})[1-\frac{1}{N}k_i(\boldsymbol{A})]}\sqrt{k_j(\boldsymbol{A})[1-\frac{1}{N}k_j(\boldsymbol{A})]}}\end{aligned} \tag{23}$$

These measures obey $-1 \leq \sigma_{ij}(\boldsymbol{A}), \tau_{ij}(\boldsymbol{A}) \leq 1$ for all (i,j) (see exercises).

The origin of the Pearson correlation (or Pearson coefficient) definition of distance between node pairs is the following. In statistics the Pearson correlation of two variables (u,v) with joint distribution $P(u,v)$ measures the degree of linear relationship between u and v, and is defined as follows (see also 8.2):

$$\mathrm{PC} = \frac{\langle uv\rangle - \langle u\rangle\langle v\rangle}{\sqrt{(\langle u^2\rangle - \langle u\rangle^2)(\langle v^2\rangle - \langle v\rangle^2)}} \tag{24}$$

One obtains formula (23) above by choosing $P(u,v) = \frac{1}{N}\sum_k \delta_{u,A_{ik}}\delta_{v,A_{jk}}$, see exercises for the proof.

4. Macroscopic structural characteristics of graphs

The previous characteristics describe the topology of the graph in the neighbourhood of a specified node. A global characterisation could be giving the full sequence of these local numbers, as in the *degree sequence* $\boldsymbol{k}(\boldsymbol{A}) = (k_1(\boldsymbol{A}), \ldots, k_N(\boldsymbol{A}))$. However, as networks get larger, it is increasingly inconvenient to draw conclusions and derive insight from large sequences. To arrive at quantitative characteristics that are less sensitive to the number of nodes, one has two simple options that continue to build on single-node features.

4.1. Average values of single-node features

The first is to consider the *averages* of the single-node quantities.

- Definition: the average in-degree of an N-node graph with adjacency matrix $\boldsymbol{A}$ is given by $\bar{k}^{\rm in}(\boldsymbol{A}) = N^{-1}\sum_{i=1}^{N} k_i^{\rm in}(\boldsymbol{A})$
- Definition: the average out-degree of an N-node graph with adjacency matrix $\boldsymbol{A}$ is given by $\bar{k}^{\rm out}(\boldsymbol{A}) = N^{-1}\sum_{i=1}^{N} k_i^{\rm out}(\boldsymbol{A})$

The average in-degree and the average out-degree in any graph are *always* identical (see exercises), which reflects the simple fact that all arrows flowing out of a node will inevitably flow into another node. So we can use in both cases the simpler notation $\bar{k}(\boldsymbol{A})$.

- Definition: the density $\rho(\boldsymbol{A}) \in [0,1]$ of an N-node graph with adjacency matrix $\boldsymbol{A}$ is the number of edges of a graph divided by the maximum possible number of edges.

We note that in a directed graph the number of links is $L = \sum_{ij} A_{ij}$, in a non directed graph it is $L = \sum_{i<j} A_{ij} + \sum_i A_{ii}$, and in a simple nondirected graph $L = \sum_{i<j} A_{ij}$. Hence

$$\textit{directed graphs}: \qquad \rho(\boldsymbol{A}) = \frac{\sum_{ij} A_{ij}}{N^2} \tag{25}$$

$$\textit{nondirected graphs}: \qquad \rho(\boldsymbol{A}) = \frac{\sum_{i<j} A_{ij} + \sum_i A_{ii}}{\frac{1}{2}N(N-1)+N} = \frac{\sum_{i<j} A_{ij} + \sum_i A_{ii}}{\frac{1}{2}N(N+1)} \tag{26}$$

$$\textit{simple nondirected graphs}: \qquad \rho(\boldsymbol{A}) = \frac{\sum_{i<j} A_{ij}}{\frac{1}{2}N(N-1)} \tag{27}$$

Note: in these definitions we do not count the links (i,j) and (j,i) in non-directed graphs twice, and the number of non-diagonal entries in a symmetric matrix is $\frac{1}{2}N(N-1)$. These densities can be written in terms of the average degree of a graph as follows (see exercises):

$$\textit{directed graphs}: \qquad \rho(\boldsymbol{A}) = \bar{k}(\boldsymbol{A})/N \tag{28}$$

$$\textit{nondirected graphs}: \qquad \rho(\boldsymbol{A}) = \bar{k}(\boldsymbol{A})/(N+1) + \sum_i A_{ii}/N(N+1) \tag{29}$$

$$\textit{simple nondirected graphs}: \qquad \rho(\boldsymbol{A}) = \bar{k}(\boldsymbol{A})/(N-1) \tag{30}$$

- Definition: the average shortest path length in a graph with adjacency matrix $\boldsymbol{A}$

$$non-directed\ graphs: \quad \overline{d}(\boldsymbol{A}) = \frac{2}{N(N-1)} \sum_{i<j} d_{ij}(\boldsymbol{A}) \tag{31}$$

$$directed\ graphs: \quad \overline{d}(\boldsymbol{A}) = \frac{1}{N(N-1)} \sum_{i\neq j} d_{ij}(\boldsymbol{A}) \tag{32}$$

- Definition: the diameter of a graph with adjacency matrix $\boldsymbol{A}$ is defined as $d(\boldsymbol{A}) = \max_{i\neq j} d_{ij}(\boldsymbol{A})$ (i.e. the distance between the pair of nodes that are furthest from one another in the graph).
- Definition: the average local clustering coefficient of an N-node graph with adjacency matrix $\boldsymbol{A}$ is defined as $\bar{C}(\boldsymbol{A}) = N^{-1} \sum_{i=1}^{N} C_i(\boldsymbol{A})$.

Note: a graph is considered 'small-world', if $\bar{C}(\boldsymbol{A})$ is significantly higher than for a random graph constructed on the same vertex set, and if the graph has approximately the same mean-shortest path length as its corresponding random graph.

- Definition: the number of links L in a directed graph is $L = \sum_{i,j\leq N} A_{ij}$. In a non-directed graph we do not count $A_{ij} = 1$ and $A_{ji} = 1$ separately, so here the number of links would be $L = \frac{1}{2} \sum_{i\neq j\leq N} A_{ij} + \sum_{i\leq N} A_{ii}$.

4.2. Distributions of single node quantities

Degree statistics. For large graphs, or when comparing graphs of different sizes, we need quantities that are intrinsically macroscopic in nature but more informative than just average values of single-node features. The simplest of these are histograms of the observed values of the N previously defined local features. If we divide, for each possible value, how often this value is observed by the total number of observations (i.e. the number of nodes, we obtain the empirical distribution of the given feature in the graph:

- Definition: the degree distribution of a non-directed N-node graph with adjacency matrix $\boldsymbol{A}$ is defined as

$$\forall k \in \mathbb{N}: \quad p(k|\boldsymbol{A}) = \frac{1}{N} \sum_{i} \delta_{k,k_i(\boldsymbol{A})} \tag{33}$$

  It gives for each k the fraction of nodes i in the graph that have degree $k_i(\boldsymbol{A}) = k$.

- Definition: the joint in- and out-degree distribution of a directed N-node graph with adjacency matrix $\boldsymbol{A}$ is defined as

$$\forall (k^{\rm in}, k^{\rm out}) \in \mathbb{N}^2: \quad p(k^{\rm in}, k^{\rm out}|\boldsymbol{A}) = \frac{1}{N} \sum_{i} \delta_{k^{\rm in},k_i^{\rm in}(\boldsymbol{A})} \delta_{k^{\rm out},k_i^{\rm out}(\boldsymbol{A})} \tag{34}$$

  It gives for each value of the pair $(k^{\rm in}, k^{\rm out})$ the fraction of nodes i in the graph that have $k_i^{\rm in}(\boldsymbol{A}) = k^{\rm in}$ and $k_i^{\rm out}(\boldsymbol{A}) = k^{\rm out}$. Often we abbreviate $\delta_{k^{\rm in},k_i^{\rm in}(\boldsymbol{A})} \delta_{k^{\rm out},k_i^{\rm out}(\boldsymbol{A})}$ as $\delta_{\vec{k},\vec{k}_i(\boldsymbol{A})}$, with $\vec{k} = (k^{\rm in}, k^{\rm out})$.

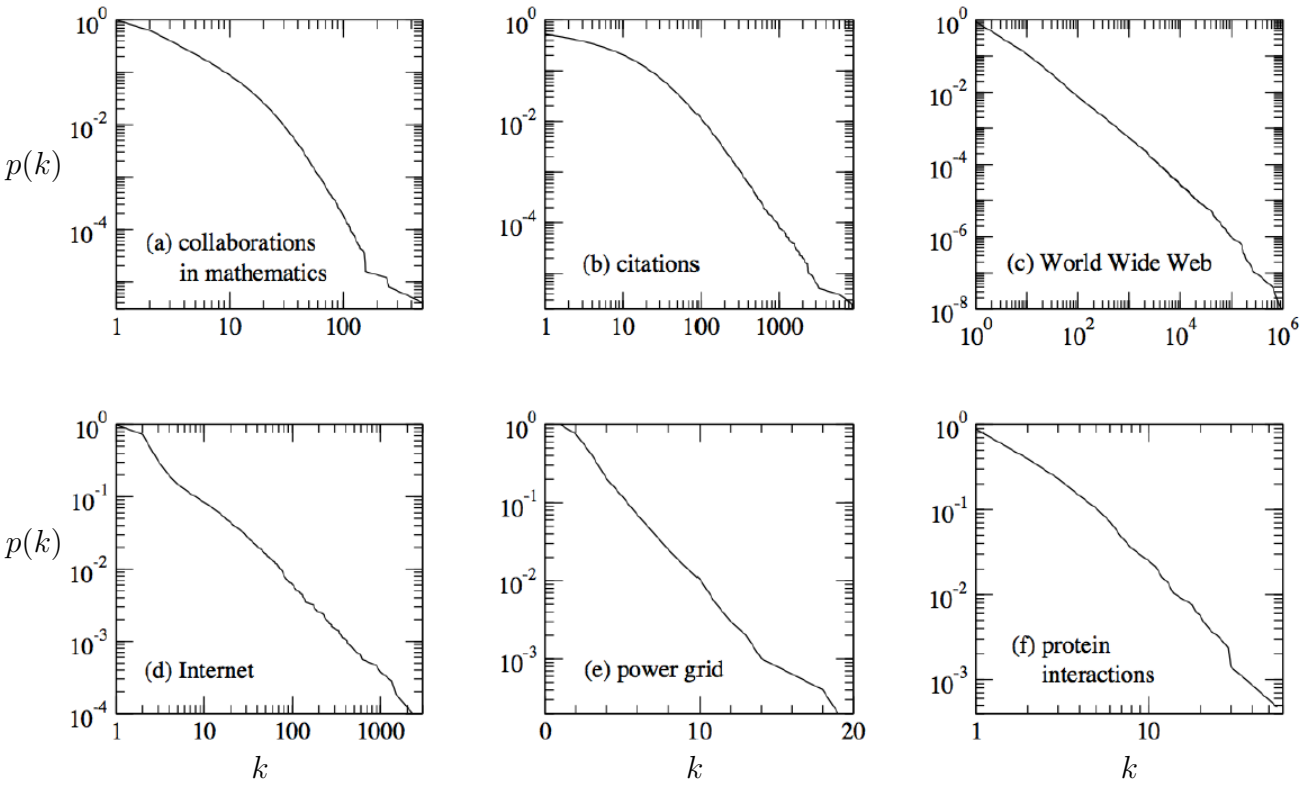


Figure 18. Degree distributions as observed in several non-directed real-world graphs, suggesting a tendency for these networks to have power-law distributions of the form $p(k) \sim k^{-\gamma}$, with powers in the range $2 < \gamma < 3$. The evidence for this is somewhat weak in the first two examples, but the last four do indeed resemble lines in a log-log plot.

In many large real-world networks one observes degree distributions of a power-law form, see e.g. Figure 18. These are also called 'scale-free' networks, since there is apparently no 'typical' scale for the degrees in such systems. See exercises. Most nodes typically have small degrees, but there is a small number of nodes (the so-called 'hubs') with very large degrees. This reflects organisation principles to which we will come back later.

Other statistics. Similarly we can define the joint distribution $p(k, T|\boldsymbol{A})$ of degrees and triangle numbers in non-directed graphs:

$$\forall k, T \in \mathbb{N}: \qquad p(k, T|\boldsymbol{A}) = \frac{1}{N} \sum_i \delta_{k,k_i(\boldsymbol{A})} \delta_{T,T_i(\boldsymbol{A})} \tag{35}$$

Now $p(k, T|\boldsymbol{A})$ is the fraction of nodes that have degree k and that participate in T triangles. It will be clear how to generalise these ideas and define similar distributions for in- and out-degrees and triangles in directed graphs, or distributions of generalised degrees.

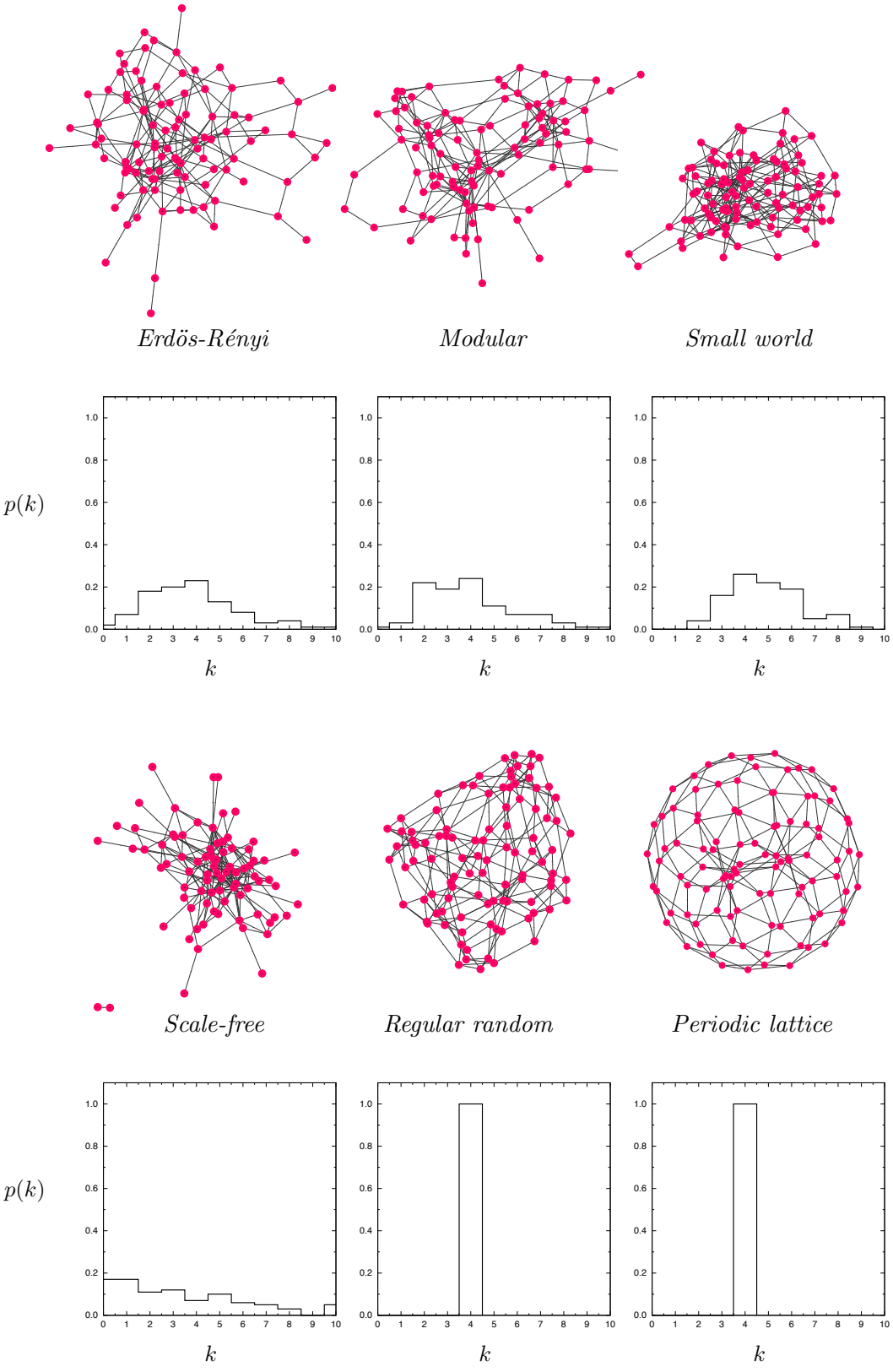


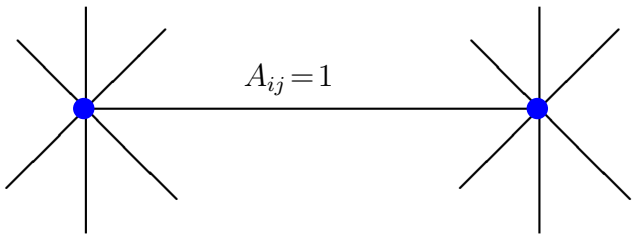
Figure 19. Examples of small numerically generated non-directed graphs, all with $N = 100$ and $\overline{k}(\boldsymbol{A}) = 4$ (their precise definitions will be given in subsequent sections of these notes). Clearly, size and average degree do not specify topologies sufficiently – there are still too many ways to generate graphs with the same size and the same number of links. The degree distribution provides additional information, but one would still like to go further.

4.3. Distributions of multi-node quantities

A logical next step, after having focused on the statistics of features that characterise nodes, is to turn to features of links. For instance, for simple nondirected graphs $\boldsymbol{A}$ we can define

- Definition: the joint distribution of degrees of *connected* node pairs in a non-directed N-node graph with adjacency matrix $\boldsymbol{A}$ is

$$\forall k,k' \geq 0: \quad W(k,k'|\boldsymbol{A}) = \frac{\sum_{i\neq j}\delta_{k,k_i(\boldsymbol{A})}A_{ij}\delta_{k',k_j(\boldsymbol{A})}}{\sum_{i\neq j}A_{ij}} \tag{36}$$



$k_i(\boldsymbol{A})=k?$ $\qquad$ $k_j(\boldsymbol{A})=k'?$

$W(k,k'|\boldsymbol{A})$ gives the fraction of non-self links in the network that connect a node of degree k to a node of degree k'. Clearly $W(k,k'|\boldsymbol{A}) = W(k',k|\boldsymbol{A})$ for all (k,k'), and $W(k,k'|\boldsymbol{A}) = 0$ if $k=0$ or $k'=0$ (or both). From (36) follows also

- Definition: the degree assortativity $a(\boldsymbol{A})$ in a non-directed graph is the Pearson correlation between the degrees of connected nonidentical node pairs,

$$a(\boldsymbol{A}) = \frac{\sum_{k,k'>0}W(k,k'|\boldsymbol{A})kk' - \left(\sum_{k>0}W(k|\boldsymbol{A})k\right)^2}{\sum_{k>0}W(k|\boldsymbol{A})k^2 - \left(\sum_{k>0}W(k|\boldsymbol{A})k\right)^2} \in [-1,1] \tag{37}$$

  with the marginal distribution $W(k|\boldsymbol{A}) = \sum_{k'>0}W(k,k'|\boldsymbol{A})$.

If $a(\boldsymbol{A}) > 0$ there is a preference in the graph for linking high-degree nodes to high-degree nodes and low-degree nodes to low-degree nodes; if $a(\boldsymbol{A}) < 0$ the preference is for linking high-degree nodes to low-degree ones. Upon summing the definition (36) over k' we see that the marginal $W(k|\boldsymbol{A})$ follows directly from the degree distribution, for simple graphs the relation is

$$W(k|\boldsymbol{A}) = \sum_{k'>0}W(k,k'|\boldsymbol{A}) = \frac{1}{N\bar{k}(\boldsymbol{A})}\sum_{i=1}^{N}\delta_{k,k_i(\boldsymbol{A})}k_i(\boldsymbol{A}) = \frac{k}{\bar{k}(\boldsymbol{A})}p(k|\boldsymbol{A}) \tag{38}$$

(for graphs with self-links we would replace $\bar{k}(\boldsymbol{A}) \to \bar{k}(\boldsymbol{A}) - N^{-1}\sum_i A_{ii}$). The reason why $W(k|\boldsymbol{A}) \neq p(k|\boldsymbol{A})$ is that in $W(k|\boldsymbol{A})$ the degree likelihood of nodes is conditioned on these nodes coming up when picking links at random; this favours nodes with more links over those with less. In those graphs where there are no correlations between the degrees of connected

nodes one would find that the joint distribution (36) is simply the product of the respective marginals (38), $W(k,k'|\boldsymbol{A}) = W(k|\boldsymbol{A})W(k'|\boldsymbol{A})$ for all $k,k' > 0$. Hence, a useful quantity to characterise correlations is

- Definition: the degree correlation ratio in a simple non-directed graph with adjacency matrix $\boldsymbol{A}$ is

$$\Pi(k,k'|\boldsymbol{A}) = \frac{W(k,k'|\boldsymbol{A})}{W(k|\boldsymbol{A})W(k'|\boldsymbol{A})} = \frac{\bar{k}^2(\boldsymbol{A})}{kk'}\frac{W(k,k'|\boldsymbol{A})}{p(k|\boldsymbol{A})p(k'|\boldsymbol{A})} \tag{39}$$

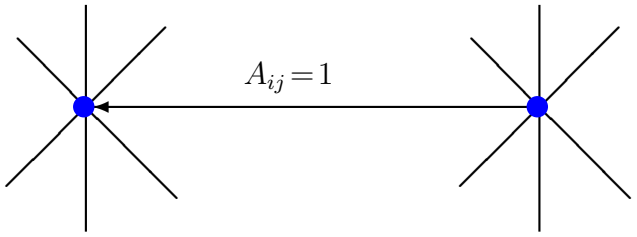
  This quantity is by definition equal to 1 for graphs without degree correlations. Any deviation from $\Pi(k,k'|\boldsymbol{A}) = 1$ will signal the presence of degree correlations.

The degree correlations captured by $\Pi(k,k'|\boldsymbol{A})$ can provide valuable new information that is not contained in the degree distribution $p(k|\boldsymbol{A})$. For instance, in Fig. 20 we show two networks with nearly identical degree distributions, that are nevertheless seen to be profoundly different at the level of degree correlations. The result of calculating the macroscopic characteristics $p(k|\boldsymbol{A})$ and $\Pi(k,k'|\boldsymbol{A})$ for the example protein interaction data of Fig. 2 is shown in Fig. 21.

For directed networks the degree correlations are described by a function $W(\vec{k},\vec{k}'|\boldsymbol{A})$, where $\vec{k} = (k^{\rm in},k^{\rm out})$ and $\vec{k}' = (k'^{\rm in},k'^{\rm out})$, since in directed graphs we must distinguish between in-in degree correlations, out-out degree correlations, and in-out degree correlations:

- Definition: the joint distribution of in- and out-degrees of *connected* node pairs in a simple directed N-node graph with adjacency matrix $\boldsymbol{A}$ is

$$\forall k,k' \geq 0: \quad W(\vec{k},\vec{k}'|\boldsymbol{A}) = \frac{\sum_{ij}\delta_{\vec{k},\vec{k}_i(\boldsymbol{A})}A_{ij}\delta_{\vec{k}',\vec{k}_j(\boldsymbol{A})}}{\sum_{ij}A_{ij}} = \frac{1}{N\bar{k}(\boldsymbol{A})}\sum_{ij}\delta_{\vec{k},\vec{k}_i(\boldsymbol{A})}A_{ij}\delta_{\vec{k}',\vec{k}_j(\boldsymbol{A})} \tag{40}$$



$\vec{k}_i(\boldsymbol{A}) = \vec{k}?$ $\qquad$ $\vec{k}_j(\boldsymbol{A}) = \vec{k}'?$

  with $\vec{k}_i(\boldsymbol{A}) = (k_i^{\rm in}(\boldsymbol{A}),k_i^{\rm out}(\boldsymbol{A}))$.

$W(\vec{k},\vec{k}'|\boldsymbol{A})$ gives the fraction of links in the network that connect a node with in- and out degrees $\vec{k}$ to a node with in- and out-degrees $\vec{k}'$. Clearly $W(\vec{k},\vec{k}'|\boldsymbol{A}) = 0$ if $\vec{k} = (0,\star)$ or $\vec{k}' = (\star,0)$ (or both), but now we may find that $W(\vec{k},\vec{k}'|\boldsymbol{A}) \neq W(\vec{k}',\vec{k}|\boldsymbol{A})$.

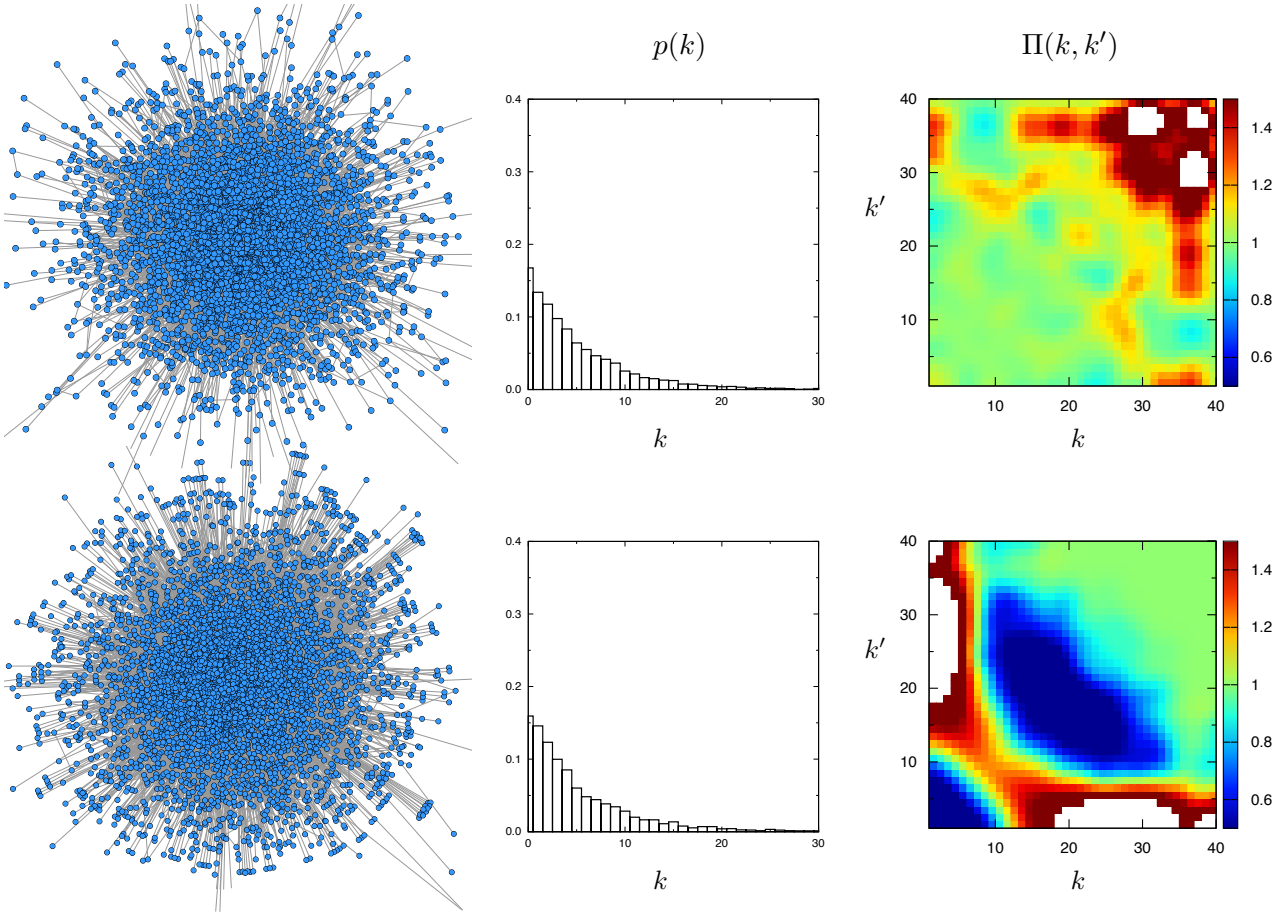


Figure 20. Illustration of the limitations of using only degree statistics to characterise graphs. The two non-directed $N = 5000$ graphs shown here look similar and have nearly indistinguishable degree distributions $p(k)$ (shown as histograms). However, they differ profoundly at the level of degree correlations, which is visible only after calculating the functions $\Pi(k, k')$ for the two graphs, shown as heat maps on the right. In the top graph, high degree nodes tend to be connected more to other high degree nodes. In the bottom graph there is a strong tendency for high degree nodes to connect to low degree nodes.

The left and right marginals of $W(\vec{k}, \vec{k}'|\boldsymbol{A})$ need not be identical (in contract to non-directed graphs). For simple directed graphs we find (see exercises):

$$W_1(\vec{k}|\boldsymbol{A}) = p(\vec{k}|\boldsymbol{A})k^{\text{in}}/\bar{k}(\boldsymbol{A}), \qquad W_2(\vec{k}'|\boldsymbol{A}) = p(\vec{k}'|\boldsymbol{A})k^{\text{out}\prime}/\bar{k}(\boldsymbol{A})$$

- Definition: the degree correlation ratio in a simple directed graph with adjacency matrix $\boldsymbol{A}$ is

$$\Pi(\vec{k}, \vec{k}'|\boldsymbol{A}) = \frac{W(\vec{k}, \vec{k}'|\boldsymbol{A})}{W_1(\vec{k}|\boldsymbol{A})W_2(\vec{k}'|\boldsymbol{A})} = \frac{\bar{k}^2(\boldsymbol{A})}{k^{\text{in}}k^{\text{out}\prime}} \frac{W(\vec{k}, \vec{k}'|\boldsymbol{A})}{p(\vec{k}|\boldsymbol{A})p(\vec{k}'|\boldsymbol{A})} \tag{41}$$

  This quantity is by definition equal to 1 for directed graphs without degree correlations. Any deviation from $\Pi(\vec{k}, \vec{k}'|\boldsymbol{A}) = 1$ will signal the presence of degree correlations.

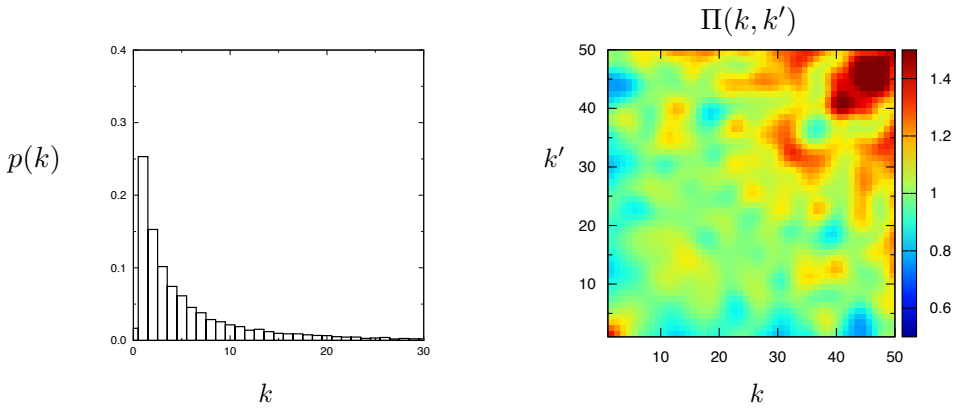


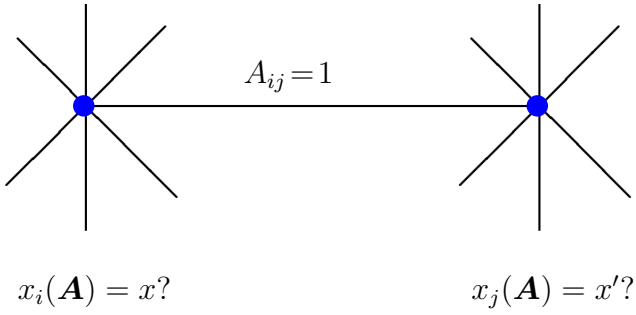
Figure 21. The degree distribution $p(k)$ (33) (left), and the normalised degree correlation kernel $\Pi(k,k')$ (39) (right, shown as a heatmap) for the protein interaction network of Fig. 2. Here $N \approx 9000$ and $\bar{k}(\boldsymbol{A}) \approx 7.5$. Significant deviations from $\Pi(k,k') \approx 1$, i.e. deviations from green in the heat map, imply nontrivial structural properties beyond those captured by the degree distribution.

4.4. Generalisation to node features other than degrees

In addition to inspecting the joint statistics of degrees of connected nodes, we can generalise the idea to arbitrary discrete features $x_i \in X$ of a node i (which could be the degree of i, but could also indicate its colour, gender, physical location, etc).

- Definition: the joint distribution of discrete features x of *connected* node pairs in a non-directed N-node graph with adjacency matrix $\boldsymbol{A}$ is

$$\forall x, x' \in X: \quad W(x,x'|\boldsymbol{A}) = \frac{\sum_{i\neq j} \delta_{x,x_i(\boldsymbol{A})} A_{ij} \delta_{x',x_j(\boldsymbol{A})}}{\sum_{i\neq j} A_{ij}} \tag{42}$$



$W(x,x'|\boldsymbol{A})$ gives the fraction of non-self links in the network that connect a node with features x to a node with features x'. Clearly $W(x,x'|\boldsymbol{A}) = W(x',x|\boldsymbol{A})$ for all (x,x'). From (42) follows also

- Definition: the assortativity $a(\boldsymbol{A})$ relative to the discrete feature x in a non-directed graph is the Pearson correlation between features of connected nonidentical node pairs,

$$a(\boldsymbol{A}) = \frac{\sum_{x,x' \in X} W(x,x'|\boldsymbol{A}) x x' - (\sum_{x \in X} W(x|\boldsymbol{A}) x)^2}{\sum_{x \in X} W(x|\boldsymbol{A}) x^2 - (\sum_{x \in X} W(x|\boldsymbol{A}) x)^2} \in [-1,1] \tag{43}$$

with the marginal distribution $W(x|\boldsymbol{A}) = \sum_{x' \in X} W(x,x'|\boldsymbol{A})$.

If $a(\boldsymbol{A}) > 0$ the linked nodes tend to have positively correlated features; if $a(\boldsymbol{A}) < 0$ they tend to have negatively correlated features. Note that this definition (43) therefore is sensible only for features x whose values are ordered in a meaningful way – like height or age, but in contrast to e.g. colour.

Upon summing (42) over x' we see that the marginal $W(x|\boldsymbol{A})$ is

$$W(x|\boldsymbol{A}) = \sum_{x' \in X} W(x,x'|\boldsymbol{A}) = \frac{1}{N \bar{k}(\boldsymbol{A})} \sum_{i=1}^{N} \delta_{x,x_i(\boldsymbol{A})} k_i(\boldsymbol{A}) \tag{44}$$

Using the joint distribution $p(x,k|\boldsymbol{A}) = N^{-1} \sum_i \delta_{x,x_i(\boldsymbol{A})} \delta_{k,k_i(\boldsymbol{A})}$ of features and degrees of nodes we can simplify the marginal of W to

$$W(x|\boldsymbol{A}) = \sum_{k>0} \frac{k}{\bar{k}(\boldsymbol{A})} p(x,k|\boldsymbol{A}) \tag{45}$$

4.5. Modularity

Sometimes the prominent structure of a network is modularity, see e.g. Fig 22. In such graphs nodes connect preferentially to other nodes that have the same module label – in fact finding the optimal modules, i.e. the optimal assignment of a string $(x_1, \ldots, x_N)$ of module labels to the nodes in the network, is a common problem in network applications. We can now use the module membership label of each node as its feature in the sense above.

To quantify the extent to which a simple non-directed graph is modular, we compare the number of 'like-connects-to-like' connections (or intra-modular links) in the graph $\boldsymbol{A}$ to what we would have found if the wiring had been completely random:

$$nr\ of\ intra{-}modular\ links\ in\ \boldsymbol{A}: \quad L_{\text{intra}}(\boldsymbol{A}) = \frac{1}{2} \sum_{i \neq j} A_{ij} \delta_{x_i,x_j} \tag{46}$$

(where the factor $\frac{1}{2}$ reflects the non-directed nature of the graph, we don't want to count the same link twice). In contrast, in a random graph $\boldsymbol{A'}$ (which we will study more rigorously later) with the same degree sequence $\boldsymbol{k} = (k_1, \ldots, k_N)$ as the graph $\boldsymbol{A}$ we would calculate the expectation value of the above quantity as $\langle L_{\text{intra}}(\boldsymbol{A'}) \rangle = \frac{1}{2} \sum_{i \neq j} \langle A'_{ij} \delta_{x_i,x_j} \rangle = \frac{1}{2} \sum_{i \neq j} \langle A'_{ij} \rangle \delta_{x_i,x_j}$, since there is assumed to be no relation between the labels x and the adjacency matrix. Now

$$\langle A'_{ij} \rangle = P(A'_{ij} = 1|\boldsymbol{k}).1 + P(A'_{ij} = 0|\boldsymbol{k}).0 = P(A'_{ij} = 1|\boldsymbol{k}) \tag{47}$$

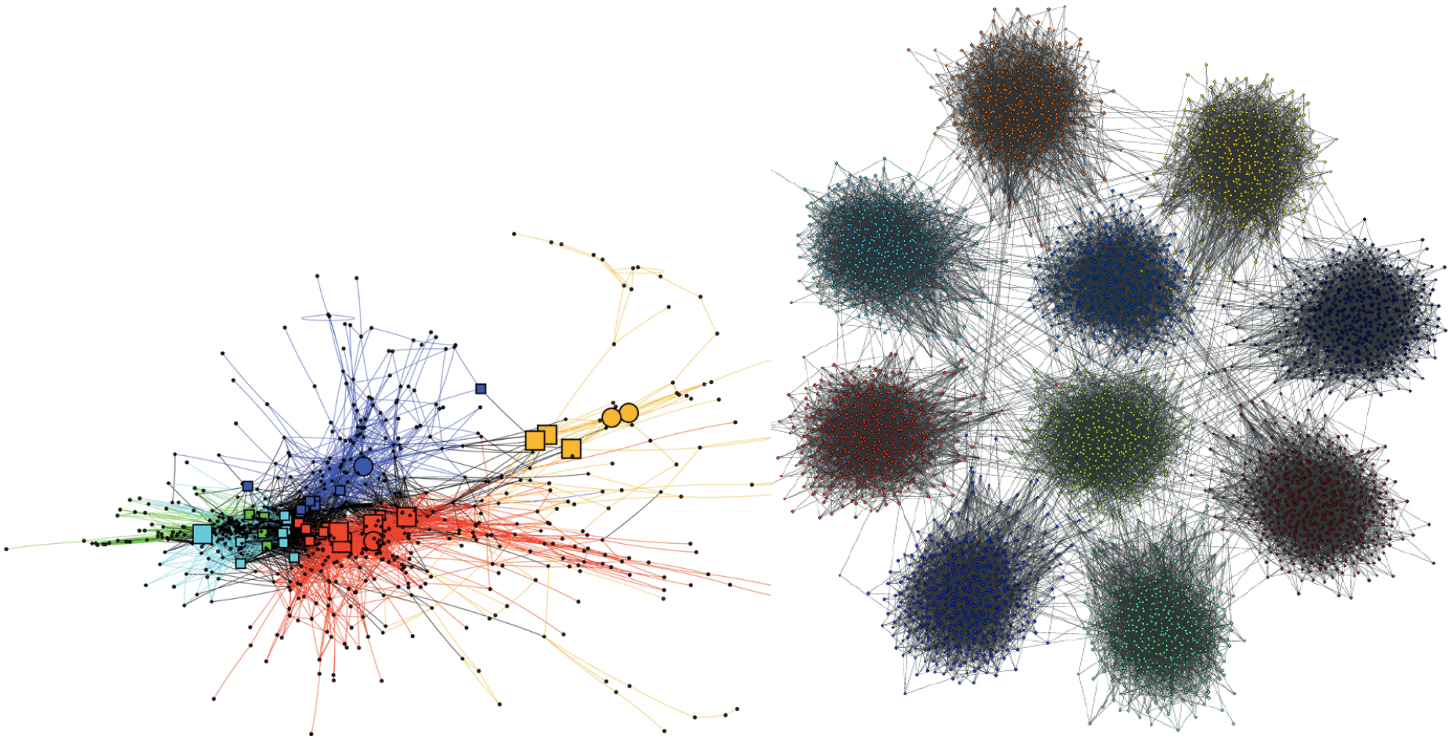


Figure 22. Examples of modular graphs. Here each node has a feature x_i that represents membership of a specific subset of nodes (its 'module', here the modules are shown colour-coded). Nodes are more frequently connected to partners within the same module, as opposed to partners from another module.

Here $P(A'_{ij} = 1|\boldsymbol{k})$ is the probability that in a random graph with degree sequence $\boldsymbol{k} = (k_1, \ldots, k_N)$ one will find nodes i and j connected. This must be proportional to k_i and k_j, so we can estimate that $\langle A'_{ij}\rangle \approx k_i k_j / C$. The value of C then follows upon summing both sides over i and j, giving (since $\boldsymbol{A}$ and $\boldsymbol{A}'$ have the same degrees):

$$\sum_{ij}\langle A'_{ij}\rangle = (\sum_i k_i)(\sum_j k_j)/C \quad \textit{hence} \quad N\bar{k} = (N\bar{k})^2/C \quad \textit{so} \quad C = N\bar{k} \tag{48}$$

This leads to our estimate $\langle A'_{ij}\rangle = k_i k_j / N\bar{k}$, and hence

$$\textit{nr of intra}-\textit{modular links in } \boldsymbol{A}' : \quad L_{\text{intra}}(\boldsymbol{A}') \approx \frac{1}{2}\sum_{i\neq j}\frac{k_i k_j}{N\bar{k}}\delta_{x_i,x_j} \tag{49}$$

We can then define (apart from an overall scaling factor) our measure of modularity in terms of the difference between the number of intra-modular links seen in $\boldsymbol{A}$ and the number we would expect to find by accident in a random non-modular graph $\boldsymbol{A}'$ with the same degrees:

- Definition: the modularity of a non-directed graph with adjacency matrix $\boldsymbol{A}$ is

$$Q(\boldsymbol{A}) = \frac{1}{2N\bar{k}(\boldsymbol{A})}\sum_{i\neq j}\Big(A_{ij} - \frac{k_i(\boldsymbol{A})k_j(\boldsymbol{A})}{N\bar{k}(\boldsymbol{A})}\Big)\delta_{x_i,x_j} \tag{50}$$

The modularity obeys $-\frac{1}{2} \leq Q(\boldsymbol{A}) \leq \frac{1}{2}$ (see exercises).

5. Processes on networks and their relation to spectral features

We often study networks because they are the infrastructure of some process. Here we inspect simple dynamical processes for variables placed on the nodes of graphs, to find out which network aspects impact on the processes that they support. This leads us to the eigenvalue spectrum of the matrix $\boldsymbol{A}$, and of the so-called Laplacian matrix of the graph.

5.1. Spin and voter models on networks

First definition – simple linear model. Imaging having a simple non-directed N-node graph with adjacency matrix $\boldsymbol{A}$. Each node i represents an individual, and carries a continuous variable s_i, which represents e.g. a voting opinion ($s_i > 0$: vote for party A; $s_i < 0$: vote for party B; $s_i = 0$: undecided). Alternatively, we could think of the s_i representing the orientations of magnetic atoms ($s_i > 0$: north pole up; $s_i < 0$: north pole down).

- *Dynamical equations*:

  Each individual i gathers opinions (or magnetic forces) from his/her social circle ∂_i, which is its neighbourhood on the graph: $\partial_i = \{j \leq N |\ A_{ij} = 1\}$, and has his/her opinion s_i driven by social pressure (the cumulative opinions) from the environment ∂_i:

  $$\frac{\mathrm{d}}{\mathrm{d}t} s_i(t) = \sum_{j \in \partial_i} s_j(t) - \lambda s_i(t) = \sum_j (\boldsymbol{A} - \lambda \mathbb{1})_{ij} s_j(t) \tag{51}$$

  $\lambda > 0$ represents a decay parameter – in the absence of peer pressure, i.e. for nodes without neighbours, so $\partial_i = \emptyset$, one would find $s_i(t) = s_i(0)\mathrm{e}^{-\lambda t}$. In vector form, with $\boldsymbol{s}(t) = (s_1(t), \ldots, s_N(t))$ equation (51) reads

  $$\frac{\mathrm{d}}{\mathrm{d}t} \boldsymbol{s}(t) = (\boldsymbol{A} - \lambda \mathbb{1}) \boldsymbol{s}(t) \tag{52}$$

- *Solution of dynamical equations*:

  Equation (52) is linear, so easily solved. The matrix $\boldsymbol{A}$ is symmetric, so it has a complete set of N orthogonal eigenvectors $\hat{\boldsymbol{e}}^k$, with $k = 1 \ldots N$, which can be normalised such that

  $$\forall k: \quad \boldsymbol{A}\hat{\boldsymbol{e}}^k = \mu_k \hat{\boldsymbol{e}}^k, \qquad \forall k, k': \quad \hat{\boldsymbol{e}}^k \cdot \hat{\boldsymbol{e}}^{k'} = \delta_{kk'} \tag{53}$$

  Here $\{\mu_1, \ldots, \mu_N\}$ are the N (not necessarily distinct) real-valued eigenvalues of $\boldsymbol{A}$; since they depend on $\boldsymbol{A}$ we should write $\mu_k(\boldsymbol{A})$, but if there is no risk of ambiguity we will drop the argument $\boldsymbol{A}$ to reduce clutter in formulae. We can use the N eigenvectors as our new basis in $\mathbb{R}^N$, and write for any $\boldsymbol{s} \in \mathbb{R}^N$: $\boldsymbol{s} = \sum_{k=1}^N \sigma_k \hat{\boldsymbol{e}}^k$. Hence also

  $$\boldsymbol{s}(t) = \sum_{k=1}^N \sigma_k(t) \hat{\boldsymbol{e}}^k \tag{54}$$

Inserting this into (52) gives

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\sum_{k=1}^{N}\sigma_k(t)\hat{\boldsymbol{e}}^k &= (\boldsymbol{A}-\lambda\mathbb{1})\sum_{k=1}^{N}\sigma_k(t)\hat{\boldsymbol{e}}^k \\
\sum_{k=1}^{N}\Big(\frac{\mathrm{d}}{\mathrm{d}t}\sigma_k(t)\Big)\hat{\boldsymbol{e}}^k &= \sum_{k=1}^{N}\sigma_k(t)(\boldsymbol{A}-\lambda\mathbb{1})\hat{\boldsymbol{e}}^k \\
\sum_{k=1}^{N}\Big(\frac{\mathrm{d}}{\mathrm{d}t}\sigma_k(t)\Big)\hat{\boldsymbol{e}}^k &= \sum_{k=1}^{N}\sigma_k(t)(\mu_k-\lambda)\hat{\boldsymbol{e}}^k
\end{aligned} \tag{55}$$

Taking the inner product on both sides of (55) with $\hat{\boldsymbol{e}}^\ell$, using (53), gives

$$\frac{\mathrm{d}}{\mathrm{d}t}\sigma_\ell(t) = (\mu_\ell-\lambda)\sigma_\ell(t) \tag{56}$$

The solution is evidently

$$\sigma_\ell(t) = \sigma_\ell(0)\mathrm{e}^{(\mu_\ell-\lambda)t} \tag{57}$$

Eigenvalues of $\boldsymbol{A}$ with $\mu_k < \lambda$ will have $\sigma_k(t) \to 0$ and those with $\mu_k > \lambda$ will have $\sigma_k(t) \to \pm\infty$. Hence either $|\boldsymbol{s}(t)| \to 0$ or $|\boldsymbol{s}(t)| \to \infty$ as $t \to \infty$; the dynamical variables evolve either to zero or to infinity. Hence, although this model (51) is well-defined mathematically, it is not a good description of social or magnetic interactions.

Revised definition – spherical model. To cure the maladies of the previous model without sacrificing its linearity, we can replace the constant λ in (51) by a *time dependent* decay rate $\lambda(t)$. If we define this $\lambda(t)$ by the requirement that $\boldsymbol{s}^2(t) = N$ for all $t \geq 0$, so that the divergencies of the previous laws can no longer occur, we obtain a so-called spherical model:

- *Dynamical equations*:

$$\forall t \geq 0: \quad \frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{s}(t) = \boldsymbol{A}\boldsymbol{s}(t) - \lambda(t)\boldsymbol{s}(t), \qquad \lambda(t) = \frac{1}{N}\boldsymbol{s}(t)\cdot\boldsymbol{A}\boldsymbol{s}(t) \tag{58}$$

The second equations follows upon setting $\boldsymbol{s}^2(0) = N$ and demanding that $\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{s}^2(t) = 0$ for all $t \geq 0$. Again we switch to the new basis of eigenvectors $\{\hat{\boldsymbol{e}}^k\}$, by substituting $\boldsymbol{s}(t) = \sum_{k=1}^{N}\sigma_k(t)\hat{\boldsymbol{e}}^k$ into (58), which now leads to the following equations, from which we need to solve both the $\{\sigma_k(t)\}$ and $\lambda(t)$:

$$\frac{\mathrm{d}}{\mathrm{d}t}\sigma_k(t) = [\mu_k-\lambda(t)]\sigma_k(t), \qquad \lambda(t) = \frac{1}{N}\sum_{k=1}^{N}\mu_k\sigma_k^2(t) \tag{59}$$

- *Solution of dynamical equations*:

The first equation in (59) is solved easily:

$$\sigma_k(t) = \sigma_k(0)\mathrm{e}^{\mu_k t - \int_0^t \mathrm{d}t'\ \lambda(t')} \tag{60}$$

Substitution into the second equation of (59) then gives an equation for $\lambda(t)$:

$$\lambda(t)\mathrm{e}^{2\int_0^t \mathrm{d}t'\ \lambda(t')} = \frac{1}{N}\sum_{k=1}^N \mu_k \mathrm{e}^{2\mu_k t}\sigma_k^2(0) \tag{61}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{e}^{2\int_0^t \mathrm{d}t'\ \lambda(t')} = \frac{\mathrm{d}}{\mathrm{d}t}\frac{1}{N}\sum_{k=1}^N e^{2\mu_k t}\sigma_k^2(0) \tag{62}$$

$$\mathrm{e}^{2\int_0^t \mathrm{d}t'\ \lambda(t')} = \frac{1}{N}\sum_{k=1}^N e^{2\mu_k t}\sigma_k^2(0) + C \tag{63}$$

$$\int_0^t \mathrm{d}t'\ \lambda(t') = \log\sqrt{\frac{1}{N}\sum_{k=1}^N e^{2\mu_k t}\sigma_k^2(0) + C} \tag{64}$$

We find the constant C by evaluating the above for $t=0$, using $\sum_k \sigma_k^2(0) = N$:

$$1 = \frac{1}{N}\sum_{k=1}^N \sigma_k^2(0) + C \qquad \textit{hence} \qquad C = 0 \tag{65}$$

We thus obtain the following solution:

$$\lambda(t) = \frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\log\Big[\frac{1}{N}\sum_{k=1}^N e^{2\mu_k t}\sigma_k^2(0)\Big] = \frac{\sum_{k=1}^N \mu_k e^{2\mu_k t}\sigma_k^2(0)}{\sum_{k=1}^N e^{2\mu_k t}\sigma_k^2(0)} \tag{66}$$

- *Asymptotic behaviour*:

  Let us assume, for simplicity, that we do not have the pathological case where the initial vector $\boldsymbol{s}(0)$ was strictly orthogonal to one or more of the $\boldsymbol{A}$-eigenvectors. So $\sigma_k(0) = \hat{\boldsymbol{e}}^k \cdot \boldsymbol{s}(0) \neq 0$ for all k. Let us also define the largest eigenvalue $\mu_{\max}$:

$$\mu_{\max} = \max_{k\leq N} \mu_k, \qquad S = \{k \leq N |\ \mu_k = \mu_{\max}\} \tag{67}$$

  Multiply numerator and denominator of (66) by $\exp(-2\mu_{\max}t)$ and take $t \to \infty$:

$$\begin{aligned}\lim_{t\to\infty}\lambda(t) &= \lim_{t\to\infty}\frac{\sum_{k\in S}\mu_{\max}\sigma_k^2(0) + \sum_{k\notin S}\mu_k e^{2(\mu_k-\mu_{\max})t}\sigma_k^2(0)}{\sum_{k\in S}\sigma_k^2(0) + \sum_{k\notin S}^N e^{2(\mu_k-\mu_{\max})t}\sigma_k^2(0)} \\ &= \frac{\sum_{k\in S}\mu_{\max}\sigma_k^2(0)}{\sum_{k\in S}\sigma_k^2(0)} = \mu_{\max}\end{aligned} \tag{68}$$

  We also know from (58) that, using standard linear algebra,

$$\lambda(t) = \frac{\boldsymbol{s}(t)\cdot\boldsymbol{A}\boldsymbol{s}(t)}{\boldsymbol{s}^2(t)} \leq \max_{\boldsymbol{x}\in\mathbb{R}^N}\frac{\boldsymbol{x}\cdot\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}^2} = \mu_{\max} \tag{69}$$

  with equality if and only if $\boldsymbol{s}(t)$ is in the eigenspace of $\boldsymbol{A}$ with eigenvalue $\mu_{\max}$. We see in (68) that our process (58) evolves towards an eigenvector of $\boldsymbol{A}$ with eigenvalue $\mu_{\max}$.

The second largest eigenvalue of $\boldsymbol{A}$ controls the timescale over which the evolution towards stationarity takes place. If we make the simple choice $\sigma_k(0) = 1\ \forall k$ for the initial conditions,

we can write the solution $\lambda(t)$ at any time in terms of the so-called eigenvalue spectrum $\{\mu_1, \ldots, \mu_N\}$ of $\boldsymbol{A}$, since when $\sigma_k(0) = 1\ \forall k$ our equation (66) simplifies to

$$\lambda(t) = \frac{\frac{1}{N}\sum_{k=1}^{N} \mu_k e^{2\mu_k t}}{\frac{1}{N}\sum_{k=1}^{N} e^{2\mu_k t}} \tag{70}$$

Apparently, the dynamics of such spherical collective processes on graphs depends critically on the eigenvalue spectra of their adjacency matrices.

5.2. Diffusion processes and random walks - the Laplacian matrix of a graph

Diffusion processes on networks. Imaging again having a simple non-directed N-node graph with adjacency matrix $\boldsymbol{A}$. Each node i of the graph now contains a conserved resource $z_i \in \mathbb{R}$ (e.g. energy, water, food, money, etc), which can diffuse (or 'leak') away to its neighbours, always from high to low levels. With 'conserved' we mean that, in contrast to the variables in our previous dynamical models, if the amount of the variable increases at one node it has to decrease somewhere else in compensation. The rate of diffusion between two nodes is larger when their differences in resource levels are larger, as would be the case with e.g. heat or water pressure.

$$\frac{\mathrm{d}}{\mathrm{d}t} z_i(t) = \sum_{j \in \partial_i} [z_j(t) - z_i(t)] = \sum_j A_{ij} z_j(t) - k_i(\boldsymbol{A}) z_i(t) \tag{71}$$

This process can be written in terms of the $N \times N$ so-called Laplacian matrix $\boldsymbol{L}$ with entries L_{ij}, as $\frac{\mathrm{d}}{\mathrm{d}t} z_i(t) = -\sum_j L_{ij} z_j(t)$, where

$$L_{ij} = k_i(\boldsymbol{A})\ \delta_{ij} - A_{ij} \tag{72}$$

In vector form, with $\boldsymbol{z}(t) = (z_1(t), \ldots, z_N(t))$, this becomes

$$\frac{\mathrm{d}}{\mathrm{d}t} \boldsymbol{z}(t) = -\boldsymbol{L}\boldsymbol{z}(t) \tag{73}$$

This is again a linear equation, which can be solved similar to earlier examples by transformation to the (complete) basis of eigenvectors of the symmetric matrix $\boldsymbol{L}$, and ends up giving a solution expressed in terms of eigenvalues and eigenvectors of $\boldsymbol{L}$. We could again worry about the possibility of exponentially diverging solutions, but we will see below that all eigenvalues of $\boldsymbol{L}$ are nonnegative, so here this cannot happen. In fact we can show easily from (71) that the total amount $Z(t) = \sum_i z_i(t)$ is conserved over time:

$$\frac{\mathrm{d}}{\mathrm{d}t} Z(t) = \sum_i \Big(\sum_j A_{ij} z_j(t) - k_i(\boldsymbol{A}) z_i(t) \Big) = \sum_j k_j(\boldsymbol{A}) z_j(t) - \sum_i k_i(\boldsymbol{A}) z_i(t) = 0 \tag{74}$$

Random walks on networks. Random walks on non-directed graphs are discrete versions of diffusion processes. We define $p_j(t) \in [0, 1]$ as the probability that the walker is at site j at time $t \in \mathbb{N}$. At each time step he moves to a new site i, selected randomly and with

equal probabilities from the neighbours of j. Since there are $k_j(\boldsymbol{A})$ sites to choose from, the probability to go to each one of these is $k_j^{-1}(\boldsymbol{A})$. Hence the dynamical equations are

$$p_i(t+1) = \sum_{j\in\partial_i} \frac{1}{k_j(\boldsymbol{A})} p_j(t) = \sum_j \frac{A_{ij}}{k_j(\boldsymbol{A})} p_j(t) \tag{75}$$

With the definition $D_{ij}(\boldsymbol{A}) = k_i(\boldsymbol{A})\delta_{ij}$ this can be rewritten as $p_i(t+1) = \sum_j (\boldsymbol{A}\boldsymbol{D}^{-1})_{ij} p_j(t)$, and hence, upon defining $\boldsymbol{p}(t) = (p_1(t), \ldots, p_N(t))$, we can write the solution at any time as

$$\boldsymbol{p}(t) = (\boldsymbol{A}\boldsymbol{D}^{-1})^t \boldsymbol{p}(0) \tag{76}$$

We see that the stationary state solution $\boldsymbol{p} = (p_1, \ldots, p_N)$ of this equation can be written as $\boldsymbol{p} = \boldsymbol{D}\boldsymbol{x}$, with $(\boldsymbol{D} - \boldsymbol{A})\boldsymbol{x} = \boldsymbol{0}$. Since $\boldsymbol{D} - \boldsymbol{A}$ is the Laplacian matrix, we see that again the Laplacian is the relevant matrix to describe the process.

The dynamical and asymptotic features of diffusion-type processes on graphs are apparently controlled by the eigenvalue spectrum of the Laplacian matrix, rather than that of the adjacency matrix. Note: the entries of the Laplacian matrix of a simple non-directed graph are no longer binary, $L_{ij} \in \{0, -1\}$ if $i \neq j$, and $L_{ii} \in \mathbb{N}$.

5.3. Spectra of adjacency matrices

Properties of eigenvalue spectra of adjacency matrices. The previous pages showed why the spectra of adjacency matrices and Laplacians are important, from the point of view of the impact of topological features of graphs on the processes for which they are the infrastructure. We now investigate the properties of these spectra, and get some feeling for which spectra we might expect to find for real networks. We know from linear algebra that all eigenvalues of symmetric matrices are real-valued, hence the above restriction to non-directed graphs. Adjacency matrices of directed graphs will indeed normally have complex eigenvalues.

In the remainder of this subsection, let $\boldsymbol{A}$ be the adjacency matrix of a non-directed N-node graph, and let $\mu_{\min}(\boldsymbol{A})$ and $\mu_{\max}(\boldsymbol{A})$ denote the smallest and the largest eigenvalue in the set $\{\mu_1(\boldsymbol{A}), \ldots, \mu_N(\boldsymbol{A})\}$. Let $\boldsymbol{u}$ be the N-dimensional vector $\boldsymbol{u} = (1, 1, \ldots, 1)$:

- Claim: $\mu_{\min}(\boldsymbol{A}) \leq \bar{k}(\boldsymbol{A})$

  Proof: $$\mu_{\min}(\boldsymbol{A}) = \min_{\boldsymbol{x}\in\mathbb{R}^N} \frac{\boldsymbol{x}\cdot\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}^2} \leq \frac{\boldsymbol{u}\cdot\boldsymbol{A}\boldsymbol{u}}{\boldsymbol{u}^2} = \frac{1}{N}\sum_{ij} A_{ij} = \bar{k}(\boldsymbol{A})$$

- Claim: $\mu_{\max}(\boldsymbol{A}) \geq \bar{k}(\boldsymbol{A})$

  Proof: $$\mu_{\max}(\boldsymbol{A}) = \max_{\boldsymbol{x}\in\mathbb{R}^N} \frac{\boldsymbol{x}\cdot\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}^2} \geq \frac{\boldsymbol{u}\cdot\boldsymbol{A}\boldsymbol{u}}{\boldsymbol{u}^2} = \frac{1}{N}\sum_{ij} A_{ij} = \bar{k}(\boldsymbol{A})$$

- Claim: $\mu_{\max}(\boldsymbol{A}) \leq \max_{j=1\ldots N}\, k_j(\boldsymbol{A})$.

  Proof:

  Let μ be an eigenvalue of $\boldsymbol{A}$ and $\boldsymbol{x} \in \mathbb{R}^N$ the corresponding eigenvector. Define $x^\star = \max_i x_i$. If $x^\star \leq 0$ (so all components of $\boldsymbol{x}$ are non-positive) we replace $\boldsymbol{x} \to -\boldsymbol{x}$, so that we *will* have $x^\star > 0$ (note: since $\boldsymbol{x} \neq \boldsymbol{0}$ this is always possible to achieve). Now choose i to be a site with $x_i = x^\star$, and use $x_j \leq x^\star$ for all j:

$$\begin{aligned} \mu x^\star &= \sum_j A_{ij} x_j \\ \mu \quad &= \sum_j A_{ij} \frac{x_j}{x^\star} \;\leq\; \sum_j A_{ij} \frac{x^\star}{x^\star} \;=\; k_i(\boldsymbol{A}) \;\leq\; \max_{j=1\ldots N}\, k_j(\boldsymbol{A}) \end{aligned}$$

  Since this result holds for any eigenvalue μ, we can indeed also state the above claim.

We can combine the above three inequalities into the following corollary:

$$\mu_{\min}(\boldsymbol{A}) \;\leq\; \bar{k}(\boldsymbol{A}) \;\leq\; \mu_{\max}(\boldsymbol{A}) \;\leq\; \max_{j=1\ldots N}\, k_j(\boldsymbol{A}) \tag{77}$$

- Definition: a non-directed N-node graph with adjacency matrix $\boldsymbol{A}$ in which all degrees $k_i(\boldsymbol{A})$ are identical to q, i.e. $k_i(\boldsymbol{A}) = q$ for all i, is called a *q-regular graph.*

- Claim: the largest eigenvalue of the adjacency matrix $\boldsymbol{A}$ of a q-regular non-directed N-node graph with adjacency matrix $\boldsymbol{A}$ is $\mu_{\max}(\boldsymbol{A}) = q$.

  Proof: this follows directly from (77).

- Claim: the eigenvalue spectrum $\{\mu_1(\boldsymbol{A}), \ldots, \mu_N(\boldsymbol{A})\}$ of a simple non-directed N-node graph obeys

$$\frac{1}{N} \sum_{k=1}^{N} \mu_k(\boldsymbol{A}) = 0 \tag{78}$$

  Proof:

  We use the fact that for each symmetric matrix $\boldsymbol{A}$ there exists a unitary $N \times N$ matrix $\boldsymbol{U}$, i.e. one such that $\boldsymbol{U}\boldsymbol{U}^\dagger = \boldsymbol{U}^\dagger\boldsymbol{U} = \mathbb{1}$, such that $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{D}(\boldsymbol{\mu})\boldsymbol{U}^\dagger$, where $\boldsymbol{D}(\boldsymbol{\mu})$ is the diagonal matrix with entries $D(\boldsymbol{\mu})_{ij} = \mu_i(\boldsymbol{A})\delta_{ij}$. Now

$$\begin{aligned} \sum_k \mu_k^\ell(\boldsymbol{A}) \quad &= \sum_k \Big[\boldsymbol{D}^\ell(\boldsymbol{\mu})\Big]_{kk} = \sum_k \Big[(\boldsymbol{U}^\dagger \boldsymbol{A}\boldsymbol{U})^\ell\Big]_{kk} \\ &= \sum_k \Big[\boldsymbol{U}^\dagger \Big(\boldsymbol{A}\boldsymbol{U}\boldsymbol{U}^\dagger\Big)^{\ell-1} \boldsymbol{A}\boldsymbol{U}\Big]_{kk} = \sum_k \Big[\boldsymbol{U}^\dagger \boldsymbol{A}^{\ell-1} \boldsymbol{A}\boldsymbol{U}\Big]_{kk} \\ &= \sum_k \Big[\boldsymbol{U}^\dagger \boldsymbol{A}^{\ell} \boldsymbol{U}\Big]_{kk} = \sum_k \sum_{ij} (\boldsymbol{U}^\dagger)_{ki} (\boldsymbol{A}^\ell)_{ij} U_{jk} = \sum_k \sum_{ij} (\boldsymbol{A}^\ell)_{ij} U_{jk} (\boldsymbol{U}^\dagger)_{ki} \\ &= \sum_{ij} (\boldsymbol{A}^\ell)_{ij} (\boldsymbol{U}\boldsymbol{U}^\dagger)_{ji} = \sum_{ij} (\boldsymbol{A}^\ell)_{ij} \delta_{ij} = \sum_i (\boldsymbol{A}^\ell)_{ii} \end{aligned}$$

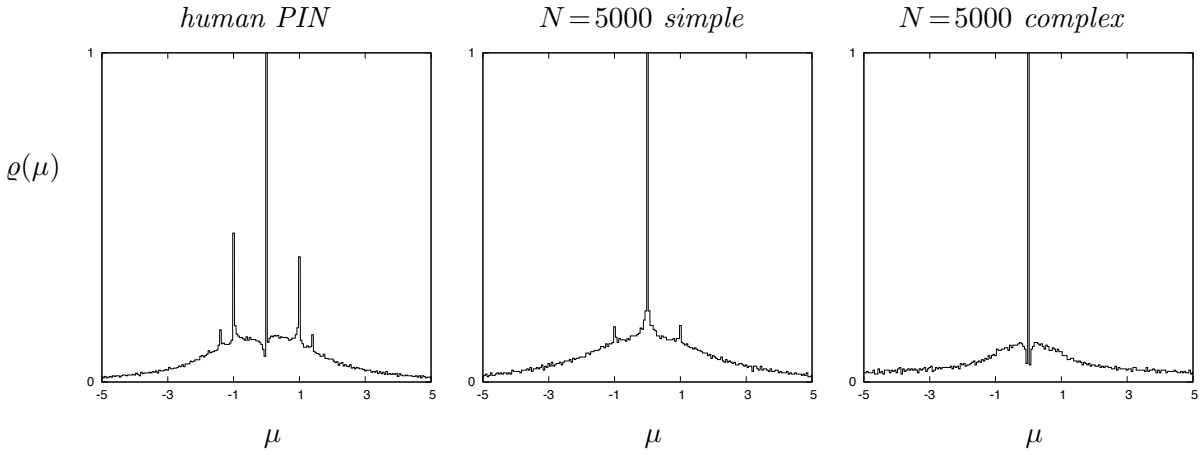


Figure 23. Adjacency matrix eigenvalue distributions (i.e. histograms of the numbers $\mu_1(\boldsymbol{A}),\ldots,\mu_N(\boldsymbol{A})$) of three graphs shown and quantified in previous figures. Left: eigenvalue distribution for the human protein interaction graph (see Figures 2 and 21). Middle and right: eigenvalue distributions for the two graphs in Fig. 20. The middle histogram refers to the graph with only weak degree correlations (top line in Fig. 20) and the right histogram refers to the graph with strong degree correlations (bottom line in Fig. 20). The eigenvalue spectra of the last two graphs are seen to be significantly different, in spite of their nearly identical degree distributions.

For $\ell = 1$ this gives $\frac{1}{N}\sum_{k=1}^{N}\mu_k(\boldsymbol{A}) = \frac{1}{N}\sum_i A_{ii} = 0$.

- Claim: the eigenvalue spectrum $\{\mu_1(\boldsymbol{A}),\ldots,\mu_N(\boldsymbol{A})\}$ of a non-directed N-node graph obeys

$$\frac{1}{N}\sum_{k=1}^{N}\mu_k^2(\boldsymbol{A}) = \bar{k}(\boldsymbol{A}) \tag{79}$$

Proof:

From the previous proof we know that for any integer $\ell > 0$: $\frac{1}{N}\sum_{k=1}^{N}\mu_k^\ell(\boldsymbol{A}) = \frac{1}{N}\sum_i(\boldsymbol{A}^\ell)_{ii}$. Upon choosing $\ell = 2$ we find, using $A_{ij}^2 = A_{ij}$ for $A_{ij}\in\{0,1\}$:

$$\begin{aligned}\frac{1}{N}\sum_{k=1}^{N}\mu_k^2(\boldsymbol{A}) &= \frac{1}{N}\sum_i(\boldsymbol{A}^2)_{ii} = \frac{1}{N}\sum_{ij}A_{ij}A_{ji}\\ &= \frac{1}{N}\sum_{ij}A_{ij} = \frac{1}{N}\sum_i k_i(\boldsymbol{A}) = \bar{k}(\boldsymbol{A})\end{aligned} \tag{80}$$

Link between adjacency matrix spectra and closed path statistics. From the eigenvalue spectrum of the adjacency matrix $\boldsymbol{A}$ of a non-directed graph one can also obtain the numbers $L_\ell(\boldsymbol{A})$ of closed paths of all possible lengths ℓ in this graph, since

- Claim: for any integer $\ell > 2$ the eigenvalue spectrum $\{\mu_1(\boldsymbol{A}),\ldots,\mu_N(\boldsymbol{A})\}$ of a non-

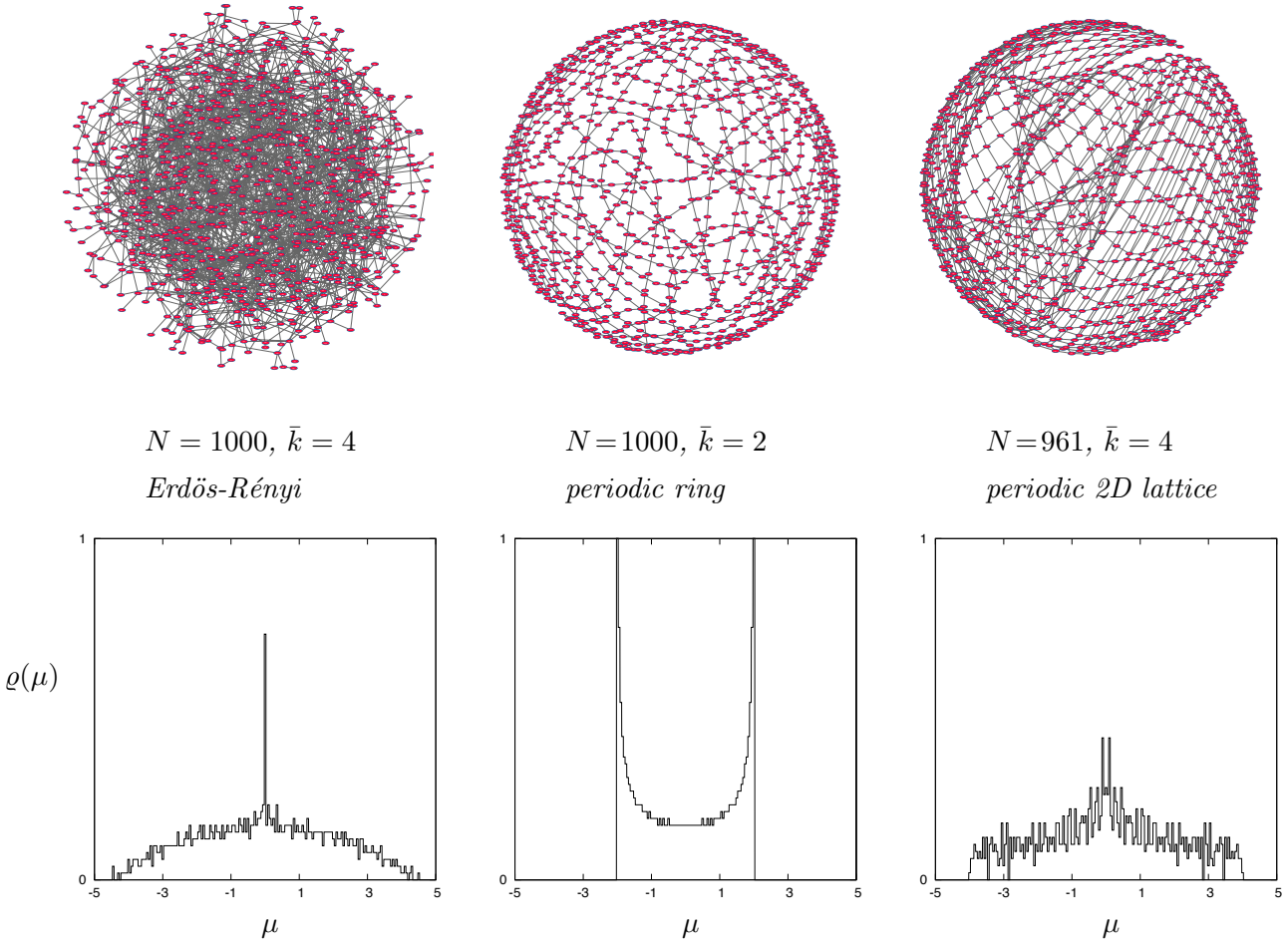


Figure 24. Adjacency matrix eigenvalue distributions (i.e. histograms of the numbers $\mu_1(\boldsymbol{A}), \ldots, \mu_N(\boldsymbol{A})$) of the three further graphs, to emphasise that the shape of these distributions can vary considerably, which, via the quantities $N^{-1}\sum_k \mu_k^\ell(\boldsymbol{A})$, reflects the different statistics of closed paths in these graphs. The Erdös-Rényi random graphs will be the subject of a subsequent section of these notes.

directed graph with adjacency matrix $\boldsymbol{A}$ obeys

$$\frac{1}{N}\sum_{k=1}^{N}\mu_k^\ell(\boldsymbol{A}) = \frac{1}{N}L_\ell(\boldsymbol{A}) \tag{81}$$

where $L_\ell(\boldsymbol{A})$, defined in (15), gives the number of closed paths of length ℓ in the graph.

Proof:

From the proof of (78) we also know that for any integer $\ell > 2$: $\frac{1}{N}\sum_k \mu_k^\ell(\boldsymbol{A}) = \frac{1}{N}\sum_i(\boldsymbol{A}^\ell)_{ii} = \frac{1}{N}L_\ell(\boldsymbol{A})$.

- Claim: if the eigenvalue spectrum of the adjacency matrix $\boldsymbol{A}$ of a non-directed N-node graph is *symmetric*, i.e. the histogram of eigenvalues is symmetrix with respect to

reflection in the line $\mu = 0$, and hence $N^{-1}\sum_k f(\mu_k(\boldsymbol{A})) = 0$ for any anti-symmetric function $f(x)$, then this graph has no closed paths of odd length (no triangles, no pentagons, etc).

Proof:

We use the previous result with $\ell = 2m+1$ and $m \in \mathbb{N}$, and use the fact that the function $f(x) = x^{2m+1}$ is anti-symmetric:

$$L_{2m+1}(\boldsymbol{A}) = \sum_k \mu_k^{2m+1}(\boldsymbol{A}) = 0$$

Some examples of adjacency matrix eigenvalue spectra for non-directed graphs that we have already inspected earlier are shown in Fig. 23. Further examples are shown in Fig. 24, to emphasise the large variability in spectra one should expect.

5.4. Spectra of Laplacian matrices

We saw that an alternative spectral characterisation of nondirected graphs, especially relevant for graphs describing diffusive processes, is based on the eigenvalues of the so-called Laplacian $N \times N$ matrix $\boldsymbol{L} = \{L_{ij}\}$, rather than those of the adjacency matrix $\boldsymbol{A}$.

- Definition: the Laplacian matrix $\boldsymbol{L}$ of an N-node graph with adjacency matrix $\boldsymbol{A}$ is defined by the entries

$$L_{ij} = k_i(\boldsymbol{A})\delta_{ij} - A_{ij} \tag{82}$$

  Since the Laplacian is a symmetric matrix, it must have real-valued eigenvalues.

- Claim: all eigenvalues of a Laplacian matrix $\boldsymbol{L}$ of a graph are nonnegative.

  Proof:

  We show that for any $\boldsymbol{x} \in \mathbb{R}^N$ one will find $\boldsymbol{x}\cdot\boldsymbol{L}\boldsymbol{x} \geq 0$:

$$\begin{aligned}\boldsymbol{x}\cdot\boldsymbol{L}\boldsymbol{x} &= \sum_{ij} x_i\Big(k_i(\boldsymbol{A})\delta_{ij} - A_{ij}\Big)x_j = \sum_i x_i^2 k_i(\boldsymbol{A}) - \sum_{ij} A_{ij}x_i x_j \\ &= \sum_{ij} x_i^2 A_{ij} - \sum_{ij} A_{ij}x_i x_j = \frac{1}{2}\sum_{ij} A_{ij}(x_i^2+x_j^2) - \sum_{ij} A_{ij}x_i x_j \\ &= \frac{1}{2}\sum_{ij} A_{ij}(x_i^2+x_j^2-2x_i x_j) = \frac{1}{2}\sum_{ij} A_{ij}(x_i-x_j)^2 \;\geq\; 0\end{aligned}$$

  Any eigenvector $\boldsymbol{x}$ of $\boldsymbol{L}$ with eigenvalue $\mu < 0$ would have given $\boldsymbol{x}\cdot\boldsymbol{L}\boldsymbol{x} = \mu\boldsymbol{x}^2 < 0$, in contradiction with the above. Hence $\boldsymbol{L}$ cannot have negative eigenvalues.

- Claim: the Laplacian matrix $\boldsymbol{L}$ of a graph always has at least one eigenvalue $\mu = 0$.

  Proof:

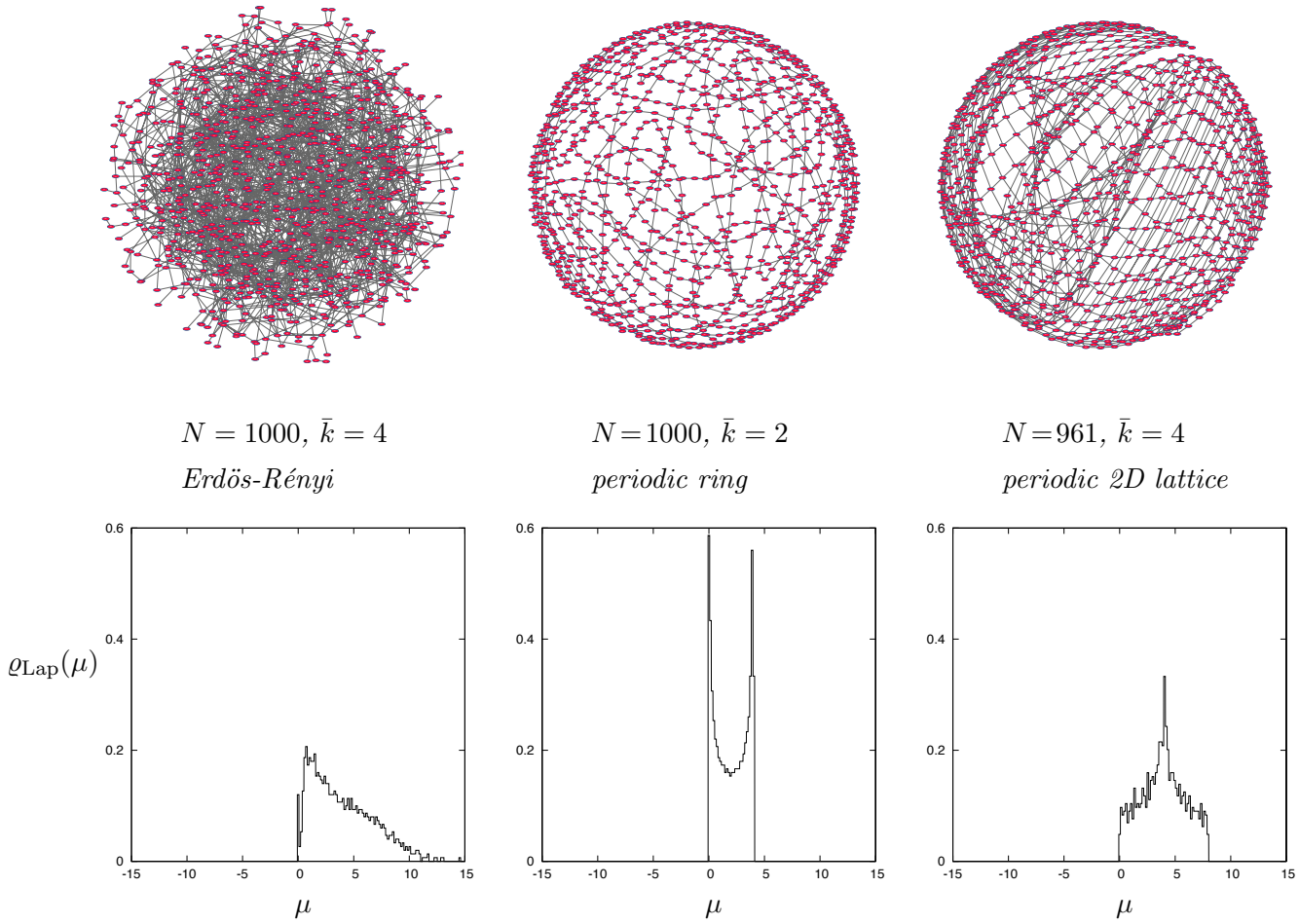


Figure 25. Laplacian matrix eigenvalue distributions $\varrho_{\rm Lap}(\mu)$ of the three graphs of Figure 24, which indeed show nonnegative eigenvalues only. It is clear the the eigenvalue spectra of the adjacency matrix and of the Laplacian matrix sometimes will and sometimes will not be similar. In the exercises we will find out why the Laplacian spectra in the middle and right graph are of the same shape as their adjacency matrix spectra in Fig. 24.

Define $\boldsymbol{u} = (1, 1, \ldots, 1)$, and show that it is an eigenvector with eigenvalue zero:

$$(\boldsymbol{L}\boldsymbol{u})_i = \sum_j L_{ij} u_j = \sum_j \Big(k_i(\boldsymbol{A})\delta_{ij} - A_{ij} \Big).1 = k_i(\boldsymbol{A}) - k_i(\boldsymbol{A}) = 0 \qquad (83)$$

- Claim: the multiplicity of the kernel of a Laplacian matrix $\boldsymbol{L}$ of a graph (i.e. the dimension of the eigenspace corresponding to eigenvalue zero) equals the number of disconnected components in the graph.

  Proof:

  Consider a vector $\boldsymbol{x}$ with eigenvalue zero. Using the identity $\boldsymbol{x} \cdot \boldsymbol{L}\boldsymbol{x} = \frac{1}{2} \sum_{ij} A_{ij} (x_i - x_j)^2$

derived in the previous proof, it follows that for such an eigenvector

$$0 = \sum_{ij} A_{ij}(x_i - x_j)^2$$

Hence $\forall(i,j):\ A_{ij} = 0$ or $x_i = x_j$. For each connected component $V' \subseteq \{1,\ldots,N\}$ of our graph we have thereby found an eigenvector $\vec{x}^{V'} \in \mathbb{R}^N$ with eigenvalue 0:

$$\textit{connected component } V': \quad \begin{cases} x_i^{V'} = 1 & if \quad i \in V' \\ x_i^{V'} = 0 & if \quad i \notin V' \end{cases}$$

Imagine there was a further zero eigenvalue, with an eigenvector $\boldsymbol{x}$ that is not one of the above. Again we would find $0 = \sum_{ij} A_{ij}(x_i - x_j)^2$. We can now decompose

$$0 = \sum_{V'} \sum_{i,j \in V'} A_{ij}(x_i - x_j)^2$$

Hence we would again get, for any connected component V': $x_i = x_j$ for all $i,j \in V'$. But that implies that $\boldsymbol{x}$ is a linear combination of the eigenvectors above, which is not possible. Hence the dimension of the kernel of $\boldsymbol{L}$, i.e. the number of independent eigenvectors with eigenvalue zero, is exactly the number of connected components.

The above features of the Laplacian matrix allow us to predict immediately the stationary state of diffusion processes. For instance, from expression (**??**) we may now conclude that in the stationary state $\boldsymbol{z}(\infty) = \lim_{t\to\infty} \boldsymbol{z}(t)$:

$$\textit{for each connected component } V': \quad \forall i \in V': \quad z_i(\infty) = \frac{1}{|V'|} \sum_{j \in V'} z_j(0) \tag{84}$$

(see exercises for proof). In Fig 25 we show the eigenvalue spectra $\varrho_{\mathrm{Lap}}(\mu|\boldsymbol{A})$ of the matrices shown earlier (with their adjacency matrix eigenvalue spectra) in Fig. 24.

6. Random and pseudo-random graphs

6.1. Random graphs as 'null models'

The need for 'null models'. We have seen many ways to quantify network topologies. The values we find for these quantifiers in a network, however, need to be interpreted. We need to know what values we would have expected to find *by default* or *typically*. If we observe that a graph has a Poissonian degree distribution, should we be excited? If we find that the number of triangles in an N-node graph equals $N/5$, is this a large or a small number? Which features of adjacency matrix spectra are common to most networks, and which are informative and special? We lack a yardstick against which to measure what we see.

We can define 'typical' values as those that we would find in a 'null model', which we define as a *random* graph that is otherwise similar to the network at hand. But how do we define 'similar'? Observations in a null model will depend on which features of the real network we imposed upon its random counterpart – the devil is in the detail. For instance, in constructing a measure for modularity, we compared observations in a network to what we would expect from a randomly generated graph with the same degrees as the observed one. We could have chosen other quantities than degrees to be copied to our null model ...

- Definition: a random graph ensemble $\{\mathcal{G}, p\}$ is defined as a set $\mathcal{G}$ of adjacency matrices $\boldsymbol{A}$, together with a measure p that specifies a probability $p(\boldsymbol{A})$ for each $\boldsymbol{A}$ in $\mathcal{G}$.

- Definition: ensemble averages of observable quantitative features $f(\boldsymbol{A})$ of random graphs are defined as
$$\langle f \rangle = \sum_{\boldsymbol{A} \in \mathcal{G}} p(\boldsymbol{A}) f(\boldsymbol{A}) \tag{85}$$

In this section we first define and study the simplest nontrivial random graph ensemble, the Erdös-Rènyi model. Later we turn to more systematic ways of defining and constructing random graph ensembles to serve as null models.

6.2. The Erdös-Rènyi model

Definition and basic properties. The Erdös-Rényi (ER) model is the random graph ensemble in which $\mathcal{G}$ is the set of all simple nondirected N-node graphs, and all links are drawn independently, according to $p(A_{ij} = 1) = p^\star$ and $p(A_{ij} = 0) = 1 - p^\star$, with $p^\star \in [0, 1]$:
$$\mathcal{G} \quad = \{\boldsymbol{A} \in \{0,1\}^{N \times N} |\ A_{ij} = A_{ji} \textit{ and } A_{ii} = 0 \ \forall i, j \leq N\} \tag{86}$$
$$p(\boldsymbol{A}) = \prod_{i<j=1}^{N} \left[p^\star \delta_{A_{ij},1} + (1 - p^\star) \delta_{A_{ij},0} \right], \tag{87}$$
We have to be careful to distinguish between averages that are defined for a single graph, such as $\bar{k}(\boldsymbol{A})$, and averages over the ensemble, to be written as $\langle \ldots \rangle$, which are average values of graph features calculated over randomly generated graph instances $\boldsymbol{A}$.

- Claim: for graphs generated in the ER ensemble , the *average* value of the average degree $\bar{k}(\boldsymbol{A}) = N^{-1}\sum_{ij} A_{ij}$ equals $p^\star(N-1)$.

  Proof:

$$\begin{aligned}
\langle \bar{k}(\boldsymbol{A})\rangle &= \sum_{\boldsymbol{A}\in\mathcal{G}} p(\boldsymbol{A})\frac{1}{N}\sum_{rs} A_{rs} = \sum_{\boldsymbol{A}\in\mathcal{G}} p(\boldsymbol{A})\frac{2}{N}\sum_{r<s} A_{rs} = \frac{2}{N}\sum_{r<s}\sum_{\boldsymbol{A}\in\mathcal{G}} p(\boldsymbol{A})A_{rs} \\
&= \frac{2}{N}\sum_{r<s}\sum_{\boldsymbol{A}\in\mathcal{G}} A_{rs}\prod_{i<j=1}^{N}\left[p^\star\delta_{A_{ij},1}+(1-p^\star)\delta_{A_{ij},0}\right] \\
&= \frac{2}{N}\sum_{r<s}\Big(\sum_{A_{rs}=0}^{1} A_{rs}\left[p^\star\delta_{A_{rs},1}+(1-p^\star)\delta_{A_{rs},0}\right]\Big)\prod_{i<j,\ (i,j)\neq(r,s)}\ \sum_{A_{ij}=0}^{1}\left[p^\star\delta_{A_{ij},1}+(1-p^\star)\delta_{A_{ij},0}\right] \\
&= \frac{2}{N}\sum_{r<s} p^\star.1 = \frac{2p^\star}{N}\sum_{r<s} = \frac{2p^\star}{N}.\frac{1}{2}N(N-1) = p^\star(N-1)
\end{aligned}$$

  Note: $\langle k\rangle$ is the average *over the ensemble* of the average degree $\bar{k}(\boldsymbol{A})$ of its graphs, i.e. $\langle k\rangle = \sum_{\boldsymbol{A}\in\mathcal{G}} p(\boldsymbol{A})\bar{k}(\boldsymbol{A})$. Individual random graphs $\boldsymbol{A}$ generated according to (87) will generally have $\bar{k}(\boldsymbol{A}) \neq \langle k\rangle$.

- Claim: the Erdös-Rényi ensemble assigns equal probabilities to all graphs with the same number of links.

  Proof:

  Since $A_{ij}\in\{0,1\}$, the probabilities (87) can be written in the alternative form:

$$p(\boldsymbol{A}) = \prod_{i<j}\left[(p^\star)^{A_{ij}}(1-p^\star)^{1-A_{ij}}\right] = (p^\star)^{\sum_{i<j}A_{ij}}(1-p^\star)^{\frac{1}{2}N(N-1)-\sum_{i<j}A_{ij}} \tag{88}$$

  Hence the dependence of $p(\boldsymbol{A})$ on $\boldsymbol{A}$ can indeed be expressed fully in terms of the number $L(\boldsymbol{A}) = \sum_{i<j} A_{ij}$ of links in $\boldsymbol{A}$, via

$$p(\boldsymbol{A}) = (p^\star)^{L(\boldsymbol{A})}(1-p^\star)^{N(N-1)/2-L(\boldsymbol{A})} \tag{89}$$

- Claim: the graph probabilities (87) of the ER ensemble can equivalently be written as

$$p(\boldsymbol{A}) = \prod_{i<j=1}^{N}\left[\frac{\langle k\rangle}{N-1}\delta_{A_{ij},1} + \left(1-\frac{\langle k\rangle}{N-1}\right)\delta_{A_{ij},0}\right] \tag{90}$$

  Proof: This follows directly from the above result $\langle k\rangle = p^\star(N-1)$.

In the ER ensemble we control the likelihood of graphs via just one graph observable, which can either be $\bar{k}(\boldsymbol{A})$ or the number of links $L(\boldsymbol{A})$ (one follows from the other), and all graphs with the same value for this parameter are equally probable. In spite of this superficial simplicity, analysing this model turns out to be less than trivial.

- Claim: the average clustering coefficient $C_i = \langle C_i(\boldsymbol{A})\rangle$ of any node i in graphs generated from the ER ensemble (87), with the definition of $C_i(\boldsymbol{A})$ given in (14), is

$$\langle C_i(\boldsymbol{A})\rangle = p^\star[1-(1-p^\star)^{N-1} - p^\star(N-1)(1-p^\star)^{N-2}] \tag{91}$$

Proof:

We use definition (14), and have to be careful to distinguish between $k_i(\boldsymbol{A}) < 2$ and $k_i(\boldsymbol{A}) \geq 2$. To handle this implicit conditioning on the degree value we use the integral representation of the Kronecker δ-symbol (see 8.4), $\delta_{nm} = (2\pi)^{-1}\int_{-\pi}^{\pi}\mathrm{d}\omega\ \mathrm{e}^{\mathrm{i}(n-m)\omega}$:

$$\begin{aligned}
\langle C_i(\boldsymbol{A})\rangle &= \Big\langle \sum_{k\geq 0}\delta_{k,k_i(\boldsymbol{A})}C_i(\boldsymbol{A})\Big\rangle = \Big\langle \sum_{k\geq 2}\delta_{k,k_i(\boldsymbol{A})}\frac{\sum_{r\neq s}A_{ir}A_{rs}A_{si}}{k_i(\boldsymbol{A})(k_i(\boldsymbol{A})-1)}\Big\rangle \\
&= \sum_{k\geq 2}\frac{1}{k(k-1)}\sum_{r\neq s}\Big\langle \delta_{k,k_i(\boldsymbol{A})}A_{ir}A_{rs}A_{si}\Big\rangle \\
&= \sum_{k\geq 2}\frac{2}{k(k-1)}\sum_{r<s}\Big\langle \int_{-\pi}^{\pi}\frac{\mathrm{d}\omega}{2\pi}\mathrm{e}^{\mathrm{i}\omega(k-\sum_j A_{ij})}A_{ir}A_{rs}A_{si}\Big\rangle \\
&= \sum_{k\geq 2}\frac{2}{k(k-1)}\sum_{r<s,\ r,s\neq i}\int_{-\pi}^{\pi}\frac{\mathrm{d}\omega}{2\pi}\mathrm{e}^{\mathrm{i}\omega k}\Big\langle A_{ir}A_{rs}A_{si}\prod_{j\neq i}\mathrm{e}^{-\mathrm{i}\omega A_{ij}}\Big\rangle \\
&= \sum_{k\geq 2}\frac{2}{k(k-1)}\sum_{r<s,\ r,s\neq i}\int_{-\pi}^{\pi}\frac{\mathrm{d}\omega}{2\pi}\mathrm{e}^{\mathrm{i}\omega k}\Big\langle A_{rs}\Big(A_{ir}\mathrm{e}^{-\mathrm{i}\omega A_{ir}}\Big)\Big(A_{is}\mathrm{e}^{-\mathrm{i}\omega A_{is}}\Big)\prod_{j\notin\{i,r,s\}}\mathrm{e}^{-\mathrm{i}\omega A_{ij}}\Big\rangle
\end{aligned}$$

So far we have only substituted definitions, and rearranged factors such that entries of the adjacency matrix are grouped together. Now we do the actual ensemble averages. The measure $p(\boldsymbol{A})$ in the ER ensemble (87) factorises over the links, reflecting the fact that they are indeed generated independently, which means that the average over $p(\boldsymbol{A})$ above simplifies to so the above can be reduced to the product of ensemble averages:

$$\begin{aligned}
\Big\langle A_{rs}\Big(A_{ir}\mathrm{e}^{-\mathrm{i}\omega A_{ir}}\Big)\Big(A_{is}\mathrm{e}^{-\mathrm{i}\omega A_{is}}\Big)\prod_{j\notin\{i,r,s\}}\mathrm{e}^{-\mathrm{i}\omega A_{ij}}\Big\rangle &= \langle A_{rs}\rangle\langle A_{ir}\mathrm{e}^{-\mathrm{i}\omega A_{ir}}\rangle\langle A_{is}\mathrm{e}^{-\mathrm{i}\omega A_{is}}\rangle\prod_{j\notin\{i,r,s\}}\langle \mathrm{e}^{-\mathrm{i}\omega A_{ij}}\rangle \\
&= p^\star(p^\star\mathrm{e}^{-\mathrm{i}\omega})^2(p^\star\mathrm{e}^{-\mathrm{i}\omega}+1-p^\star)^{N-3} = (p^\star)^3\mathrm{e}^{-2\mathrm{i}\omega}(p^\star\mathrm{e}^{-\mathrm{i}\omega}+1-p^\star)^{N-3}
\end{aligned}$$

Next we use Newton's binomium formula to work out the quantity $(p^\star\mathrm{e}^{-\mathrm{i}\omega}+1-p^\star)^{N-3}$:

$$\begin{aligned}
\Big\langle A_{rs}\Big(A_{ir}\mathrm{e}^{-\mathrm{i}\omega A_{ir}}\Big)\Big(A_{is}\mathrm{e}^{-\mathrm{i}\omega A_{is}}\Big)\prod_{j\notin\{i,r,s\}}\mathrm{e}^{-\mathrm{i}\omega A_{ij}}\Big\rangle &= (p^\star)^3\mathrm{e}^{-2\mathrm{i}\omega}\sum_{\ell=0}^{N-3}\binom{N-3}{\ell}(p^\star)^\ell\mathrm{e}^{-\ell\mathrm{i}\omega}(1-p^\star)^{N-3-\ell} \\
&= \sum_{\ell=0}^{N-3}\binom{N-3}{\ell}(p^\star)^{\ell+3}\mathrm{e}^{-(\ell+2)\mathrm{i}\omega}(1-p^\star)^{N-3-\ell}
\end{aligned}$$

We insert this into our expression for $\langle C_i(\boldsymbol{A})\rangle$, use $\sum_{r<s,\ r,s\neq i}1 = \frac{1}{2}(N-1)(N-2)$, and do some simple cleaning up:

$$\langle C_i(\boldsymbol{A})\rangle = \sum_{k\geq 2}\frac{(N-1)(N-2)}{k(k-1)}\int_{-\pi}^{\pi}\frac{\mathrm{d}\omega}{2\pi}\mathrm{e}^{\mathrm{i}\omega k}\sum_{\ell=0}^{N-3}\binom{N-3}{\ell}(p^\star)^{\ell+3}\mathrm{e}^{-(\ell+2)\mathrm{i}\omega}(1-p^\star)^{N-3-\ell}$$

$$= \sum_{k\geq 2} \frac{(N-1)(N-2)}{k(k-1)} \sum_{\ell=0}^{N-3} \Big(\begin{array}{c} N-3 \\ \ell \end{array} \Big) (p^\star)^{\ell+3} (1-p^\star)^{N-3-\ell} \int_{-\pi}^{\pi} \frac{\mathrm{d}\omega}{2\pi} \mathrm{e}^{\mathrm{i}\omega(k-\ell-2)}$$

At this stage we use the integral representation of the Kronecker delta-symbol to get rid of the ω-integral:

$$\begin{aligned} \langle C_i(\boldsymbol{A}) \rangle &= \sum_{k\geq 2} \frac{(N-1)(N-2)}{k(k-1)} \sum_{\ell=0}^{N-3} \Big(\begin{array}{c} N-3 \\ \ell \end{array} \Big) (p^\star)^{\ell+3} (1-p^\star)^{N-3-\ell} \; \delta_{k,\ell+2} \\ &= \sum_{k=2}^{N-1} \frac{(N-1)(N-2)}{k(k-1)} \Big(\begin{array}{c} N-3 \\ k-2 \end{array} \Big) (p^\star)^{k+1} (1-p^\star)^{N-1-k} \end{aligned}$$

We now write explicitly the combinatorial factor, and clean up the various quantities where possible:

$$\begin{aligned} \langle C_i(\boldsymbol{A}) \rangle &= \sum_{k=2}^{N-1} \frac{(N-1)(N-2)}{k(k-1)} \frac{(N-3)!}{(k-2)!(N-1-k)!} (p^\star)^{k+1} (1-p^\star)^{N-1-k} \\ &= \sum_{k=2}^{N-1} \frac{(N-1)!}{k!(N-1-k)!} (p^\star)^{k+1} (1-p^\star)^{N-1-k} \\ &= p^\star \sum_{k=2}^{N-1} \Big(\begin{array}{c} N-1 \\ k \end{array} \Big) (p^\star)^{k} (1-p^\star)^{N-1-k} \\ &= p^\star \sum_{k=0}^{N-1} \Big(\begin{array}{c} N-1 \\ k \end{array} \Big) (p^\star)^{k} (1-p^\star)^{N-1-k} - p^\star (1-p^\star)^{N-1} - p^\star (N-1) p^\star (1-p^\star)^{N-2} \end{aligned}$$

We then recognise that Newton's binomial formula can be used to do the sum over ℓ, and proceed to our final result:

$$\langle C_i(\boldsymbol{A}) \rangle = p^\star \Big\{ 1 - (1-p^\star)^{N-1} - (N-1) p^\star (1-p^\star)^{N-2} \Big\} \tag{92}$$

The above proof is a useful exercise in the use of various bookkeeping tools, such us summation formulae from Calculus, Newton's binomial formula, and the integral representation of the Kronecker delta-symbol. These tools will continue to serve us.

6.3. The Erdös-Renyi model in the finite connectivity regime

We are usually interested in large networks with a finite average degree – these tend to be found in the real world. Therefore many properties of the ER ensemble have been studied in the so-called finite connectivity regime, starting from (90), where: $N \to \infty$ with $\langle k \rangle$ finite. It follows from a relation found earlier, namely $\langle \bar{k}(\boldsymbol{A}) \rangle = p^\star (N-1)$, that in this regime $p^\star = \mathcal{O}(N^{-1})$. The probability for an individual link to be present must indeed scale as $\mathcal{O}(N^{-1})$ in order to have on average a finite number of partners per node in the system. We now investigate properties of the ER ensemble in this *finite connectivity* limit.

- Claim: in the finite connectivity limit, i.e. for $N \to \infty$ with $\langle k \rangle$ fixed, the degree distribution of the Erdös-Rènyi ensemble has the Poissonnian form

$$\lim_{N\to\infty} \langle p(k|\boldsymbol{A})\rangle = \mathrm{e}^{-\langle k\rangle}\langle k\rangle^k/k! \tag{93}$$

Proof:

$$\lim_{N\to\infty} p(k|\boldsymbol{A}) = \lim_{N\to\infty} \sum_{\boldsymbol{A}\in\mathcal{G}} p(k|\boldsymbol{A})p(\boldsymbol{A}|\langle k\rangle) \tag{94}$$

As always we manipulate all ensemble averages until they factorise over the bond variables, since the entries of $\boldsymbol{A}$ are in the ER model distributed independently. Again we use the symmetry and absence of diagonal elements of $\boldsymbol{A}$, and the integral representation of the Kronecker δ-symbol to achieve this:

$$\begin{aligned}
\lim_{N\to\infty}\langle p(k|\boldsymbol{A})\rangle &= \lim_{N\to\infty}\langle \frac{1}{N}\sum_i \delta_{k,\sum_j A_{ij}}\rangle = \lim_{N\to\infty}\frac{1}{N}\sum_i \int_{-\pi}^{\pi}\frac{\mathrm{d}\omega}{2\pi}\mathrm{e}^{\mathrm{i}\omega k}\langle \mathrm{e}^{-\mathrm{i}\omega\sum_j A_{ij}}\rangle \\
&= \lim_{N\to\infty}\frac{1}{N}\sum_i \int_{-\pi}^{\pi}\frac{\mathrm{d}\omega}{2\pi}\mathrm{e}^{\mathrm{i}\omega k}\langle \mathrm{e}^{-\mathrm{i}\omega\sum_{js} A_{sj}\delta_{is}}\rangle \\
&= \lim_{N\to\infty}\frac{1}{N}\sum_i \int_{-\pi}^{\pi}\frac{\mathrm{d}\omega}{2\pi}\mathrm{e}^{\mathrm{i}\omega k}\langle \mathrm{e}^{-\frac{1}{2}\mathrm{i}\omega\sum_{js} A_{sj}(\delta_{is}+\delta_{ij})}\rangle \\
&= \lim_{N\to\infty}\frac{1}{N}\sum_i \int_{-\pi}^{\pi}\frac{\mathrm{d}\omega}{2\pi}\mathrm{e}^{\mathrm{i}\omega k}\langle \mathrm{e}^{-\mathrm{i}\omega\sum_{j<s} A_{sj}(\delta_{is}+\delta_{ij})}\rangle \\
&= \lim_{N\to\infty}\frac{1}{N}\sum_i \int_{-\pi}^{\pi}\frac{\mathrm{d}\omega}{2\pi}\mathrm{e}^{\mathrm{i}\omega k}\prod_{j<s}\langle \mathrm{e}^{-\mathrm{i}\omega A_{sj}(\delta_{is}+\delta_{ij})}\rangle \\
&= \lim_{N\to\infty}\frac{1}{N}\sum_i \int_{-\pi}^{\pi}\frac{\mathrm{d}\omega}{2\pi}\mathrm{e}^{\mathrm{i}\omega k}\prod_{j<s}\sum_{A\in\{0,1\}}\Big(\frac{\langle k\rangle}{N-1}\delta_{A,1}+(1-\frac{\langle k\rangle}{N-1})\delta_{A,0}\Big)\mathrm{e}^{-\mathrm{i}\omega A(\delta_{is}+\delta_{ij})} \\
&= \lim_{N\to\infty}\frac{1}{N}\sum_i \int_{-\pi}^{\pi}\frac{\mathrm{d}\omega}{2\pi}\mathrm{e}^{\mathrm{i}\omega k}\prod_{j<s}\Big[1+\frac{\langle k\rangle}{N-1}(\mathrm{e}^{-\mathrm{i}\omega(\delta_{is}+\delta_{ij})}-1)\Big]
\end{aligned}$$

Next, since $N\to\infty$ and $\langle k\rangle$ is finite, we can use the expansion $1+x=\exp(x-\frac{1}{2}x^2+\mathcal{O}(x^3))$. We also note that $\mathrm{e}^{-\mathrm{i}(\delta_{is}+\delta_{ij})}-1=0$ unless either $s=i$ or $j=i$ or both, so that $\sum_{j<s}(\mathrm{e}^{-\mathrm{i}(\delta_{is}+\delta_{js})}-1)=\mathcal{O}(N)$ and $\sum_{j<s}(\mathrm{e}^{-\mathrm{i}(\delta_{is}+\delta_{js})}-1)^2=\mathcal{O}(N)$ as $N\to\infty$. Hence

$$\begin{aligned}
\lim_{N\to\infty}\langle p(k|\boldsymbol{A})\rangle &= \lim_{N\to\infty}\frac{1}{N}\sum_i \int_{-\pi}^{\pi}\frac{\mathrm{d}\omega}{2\pi}\mathrm{e}^{\mathrm{i}\omega k}\prod_{j<s}\mathrm{e}^{\frac{\langle k\rangle}{N-1}(\mathrm{e}^{-\mathrm{i}(\delta_{is}+\delta_{ij})}-1)-\frac{1}{2}\frac{\langle k\rangle^2}{(N-1)^2}(\mathrm{e}^{-\mathrm{i}\omega(\delta_{is}+\delta_{ij})}-1)^2+\mathcal{O}(N^{-3})} \\
&= \lim_{N\to\infty}\frac{1}{N}\sum_i \int_{-\pi}^{\pi}\frac{\mathrm{d}\omega}{2\pi}\mathrm{e}^{\mathrm{i}\omega k+\frac{\langle k\rangle}{N}\sum_{j<s}(\mathrm{e}^{-\mathrm{i}\omega(\delta_{is}+\delta_{ij})}-1)+\mathcal{O}(N^{-1})} \\
&= \lim_{N\to\infty}\frac{1}{N}\sum_i \int_{-\pi}^{\pi}\frac{\mathrm{d}\omega}{2\pi}\mathrm{e}^{\mathrm{i}\omega k+\frac{\langle k\rangle}{2N}\sum_{j\neq s}(\mathrm{e}^{-\mathrm{i}\omega(\delta_{is}+\delta_{ij})}-1)+\mathcal{O}(N^{-1})} \\
&= \lim_{N\to\infty}\frac{1}{N}\sum_i \int_{-\pi}^{\pi}\frac{\mathrm{d}\omega}{2\pi}\mathrm{e}^{\mathrm{i}\omega k+\frac{\langle k\rangle}{N}\sum_{j\neq i}(\mathrm{e}^{-\mathrm{i}\omega}-1)+\mathcal{O}(N^{-1})}
\end{aligned}$$

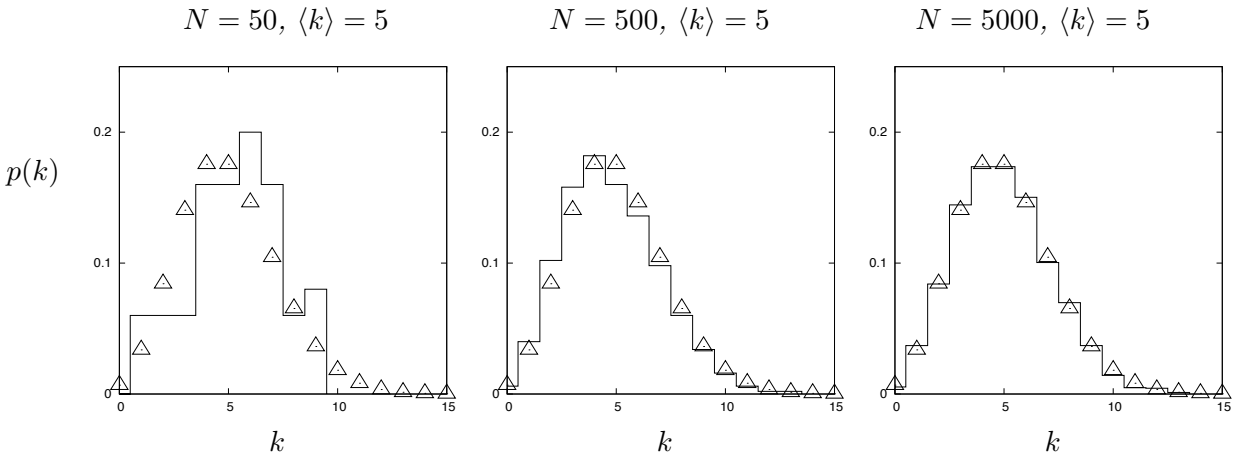


Figure 26. Comparison between the theoretical prediction (93) for the average degree distribution of infinitely large ER graphs (triangles) and the observed distributions in random samples from the ER ensemble, for $N = 50,\ 500,\ 5000$ (histograms). Clearly, for $N = 500$ the differences between actual degree distributions and the $N \to \infty$ formula are already small, and for $N = 5000$ they are more or less negligible. This means that asymptotic results are useful descriptions of large but finite graphs.

$$= \lim_{N\to\infty} \frac{1}{N} \sum_i \int_{-\pi}^{\pi} \frac{\mathrm{d}\omega}{2\pi} \mathrm{e}^{\mathrm{i}\omega k + \langle k\rangle(\mathrm{e}^{-\mathrm{i}\omega}-1) + \mathcal{O}(N^{-1})} = \mathrm{e}^{-\langle k\rangle} \int_{-\pi}^{\pi} \frac{\mathrm{d}\omega}{2\pi} \mathrm{e}^{\mathrm{i}\omega k + \langle k\rangle \mathrm{e}^{-\mathrm{i}\omega}}$$

The last step is the remaining integral over ω. It looks nasty, but is in fact simple. Just expand $\exp(\langle k\rangle \mathrm{e}^{-\mathrm{i}\omega})$ as a power series, and use the integral representation of the Kronecker delta again:

$$\lim_{N\to\infty} \langle p(k|\boldsymbol{A})\rangle = \mathrm{e}^{-\langle k\rangle} \int_{-\pi}^{\pi} \frac{\mathrm{d}\omega}{2\pi} \mathrm{e}^{\mathrm{i}\omega k} \sum_{\ell\geq 0} \frac{\langle k\rangle^\ell}{\ell!} \mathrm{e}^{-\mathrm{i}\ell\omega} = \mathrm{e}^{-\langle k\rangle} \sum_{\ell\geq 0} \frac{\langle k\rangle^\ell}{\ell!} \int_{-\pi}^{\pi} \frac{\mathrm{d}\omega}{2\pi} \mathrm{e}^{\mathrm{i}\omega(k-\ell)}$$

$$= \mathrm{e}^{-\langle k\rangle} \sum_{\ell\geq 0} \frac{\langle k\rangle^\ell}{\ell!} \delta_{k\ell} = \mathrm{e}^{-\langle k\rangle} \langle k\rangle^k / k!$$

In Fig 26 we see that this asymptotic result (i.e. a formula derived for $N \to \infty$) is accurate already for large but finite graphs of size $N \sim 10^2 - 10^3$. In other words: in the finite connectivity regime, size $N \sim 10^2 - 10^3$ graphs appear to behave like infinite ones.

- Claim: in the finite connectivity limit, i.e. for $N \to \infty$ with $\langle k\rangle$ fixed, the average clustering coefficients of the Erdös-Rènyi ensemble become

$$\langle C_i(\boldsymbol{A})\rangle = \frac{\langle k\rangle}{N}\left[1 - \mathrm{e}^{-\langle k\rangle} - \langle k\rangle \mathrm{e}^{-\langle k\rangle}\right] + \mathcal{O}(N^{-2}) \tag{95}$$

Proof:

We substitute $p^\star = \langle k\rangle/(N-1)$ into (91) and expand the result for $N \to \infty$, using

$\log(1+x) = x + \mathcal{O}(x^2)$ and $\mathrm{e}^x = 1 + x + \mathcal{O}(x^2)$:

$$\begin{aligned}
\langle C_i(\boldsymbol{A})\rangle &= \frac{\langle k\rangle}{N-1}\Big[1-(1-\frac{\langle k\rangle}{N-1})^{N-1} - \langle k\rangle(1-\frac{\langle k\rangle}{N-1})^{N-2}\Big] \\
&= \frac{\langle k\rangle}{N-1}\Big[1-\mathrm{e}^{(N-1)\log(1-\frac{\langle k\rangle}{N-1})} - \langle k\rangle \mathrm{e}^{(N-2)\log(1-\frac{\langle k\rangle}{N-1})}\Big] \\
&= \frac{\langle k\rangle}{N-1}\Big[1-\mathrm{e}^{-(N-1)\frac{\langle k\rangle}{N-1}+\mathcal{O}(N^{-1})} - \langle k\rangle \mathrm{e}^{-(N-2)\frac{\langle k\rangle}{N-1}+\mathcal{O}(N^{-1})}\Big] \\
&= \frac{\langle k\rangle}{N-1}\Big[1-\mathrm{e}^{-\langle k\rangle+\mathcal{O}(N^{-1})} - \langle k\rangle \mathrm{e}^{-\langle k\rangle+\mathcal{O}(N^{-1})}\Big] \\
&= \frac{\langle k\rangle}{N}\Big[1-\mathrm{e}^{-\langle k\rangle} - \langle k\rangle \mathrm{e}^{-\langle k\rangle}\Big] + \mathcal{O}(N^{-2})
\end{aligned}$$

So in the finite connectivity scaling regime all clustering coefficients of typical Erdös-Rényi graphs vanish for $N\to\infty$, and the number of triangles per node is order N^{-1}. Using in principle similar tools (but involving calculations that are more tedious), one can show that large random graphs generated from the Erdös-Rényi ensemble (87) with fixed $\langle k\rangle$ will be *locally tree-like* (i.e. have a vanishing number of short loops per node) and will on average have *vanishing degree correlations*,

$$\text{for all } (k,k'): \quad \lim_{N\to\infty}\langle W(k,k'|\boldsymbol{A})\rangle = [\lim_{N\to\infty}\langle W(k|\boldsymbol{A})\rangle][\lim_{N\to\infty}\langle W(k'|\boldsymbol{A})\rangle] \tag{96}$$

with $W(k,k'|\boldsymbol{A})$ as defined in (36).

6.4. Generating functions

Many averages over the degree distribution $p(k)$ of a graph or an ensemble of graphs can be expressed in terms of the following generating function, calculation of which will reduce the amount of work (and the likelihood or error) in our calculations:

- Definition: the generating function of the degree distribution is defined for $x\in[0,1]$ as

$$G(x) = \sum_{k\geq 0} p(k)x^k \tag{97}$$

  We see that it obeys: $\frac{\mathrm{d}}{\mathrm{d}x}G(x)\geq 0$, with $G(0)=p(0)$ and $G(1)=1$.

- Claim: the degree distribution follows from its generating function via

$$p(k) = \lim_{x\to 0}\frac{1}{k!}\frac{\mathrm{d}^k G(x)}{\mathrm{d}x^k} \tag{98}$$

  Proof:

  This follows directly from application of the Taylor expansion to the function $G(x)$, which tells us that $G(x)=\sum_{\ell\geq 0}\frac{1}{\ell!}G^{(\ell)}(0)x^\ell$.

- Claim: all moments $\langle k^m \rangle$ of the degree distribution, with $m \in \mathbb{N}$, follow from the generating function via

$$\langle k^m \rangle = \lim_{x \to 1} \left(x \frac{\mathrm{d}}{\mathrm{d}x} \right)^m G(x) \tag{99}$$

  Proof:

  For $m = 0$ the claim holds trivially. We just work out the recipe on the right for $m > 0$:

$$\begin{aligned} \left(x \frac{\mathrm{d}}{\mathrm{d}x} \right)^m G(x) &= \left(x \frac{\mathrm{d}}{\mathrm{d}x} \right)^m \sum_{k \geq 0} p(k) x^k = \left(x \frac{\mathrm{d}}{\mathrm{d}x} \right)^{m-1} \sum_{k \geq 0} p(k) x \frac{\mathrm{d}}{\mathrm{d}x} x^k \\ &= \left(x \frac{\mathrm{d}}{\mathrm{d}x} \right)^{m-1} \sum_{k \geq 0} p(k) k x^k = \left(x \frac{\mathrm{d}}{\mathrm{d}x} \right)^{m-2} \sum_{k \geq 0} p(k) k^2 x^k \\ &= \ldots\ldots \\ &= \left(x \frac{\mathrm{d}}{\mathrm{d}x} \right)^{0} \sum_{k \geq 0} p(k) k^m x^k = \sum_{k \geq 0} p(k) k^m x^k \end{aligned}$$

  Setting $x \to 1$ then leads to the above claim.

Let us work out the generating function (97) for some simple degree distributions, with average degree $\langle k \rangle = q \geq 0$:

- Claim: the generating function of the degree distribution for regular random graphs, where $p(k) = \delta_{qk}$, is

$$G(x) = x^q \tag{100}$$

  Proof: this is trivial, in the sum over k we retain only the term $k = q$.

- Claim: the generating function of the degree distribution for Poissonnian random graphs, i.e. where $p(k) = \mathrm{e}^{-q} q^k / k!$ with $q \geq 0$ (like finitely connected ER ones) is

$$G(x) = \mathrm{e}^{-q(1-x)} \tag{101}$$

  Proof:

  We do the sum over k in (106), using $\sum_{k \geq 0} y^k / k! = \mathrm{e}^y$:

$$G(x) = \mathrm{e}^{-q} \sum_{k \geq 0} \frac{q^k x^k}{k!} = \mathrm{e}^{-q} \mathrm{e}^{qx} = \mathrm{e}^{-q(1-x)} \tag{102}$$

- Claim: the generating function of the degree distribution for random graphs with exponential degree distributions, i.e. where $p(k) = (\frac{q}{1+q})^k / (1+q)$ with $q \geq 0$, is

$$G(x) = \frac{1}{1 + q(1-x)} \tag{103}$$

  (see exercises for proofs that this distribution obeys $\sum_k p(k) = 1$ and $\sum_k p(k) k = q$).

Proof:

We do the sum over k in (106), using $\sum_{k\geq 0} y^k = 1/(1-y)$:

$$G(x) = \frac{1}{1+q}\sum_{k\geq 0}\Big(\frac{qx}{1+q}\Big)^k = \frac{1}{1+q}\frac{1}{1-qx/(1+q)}$$
$$= \frac{1}{1+q-qx} = \frac{1}{1+q(1-x)} \tag{104}$$

In the exercises we confirm for all three examples that they indeed obey $G(0)=0$, $G(1)=1$, and $\lim_{x\to 1} x\frac{\mathrm{d}}{\mathrm{d}x}G(x) = q$.

6.5. The giant component in random tree-like graphs

Definition and general formula. We saw that in the finite connectivity regime the ER model generates locally tree-like graphs. These have convenient mathematical features. The main one is this: given the statistical features of a node i, those of the k_i nodes in its environment ∂_i can in leading order be taken as statistically independent, since each 'branch' of the tree centred at i is connected to each other branch *only via* node i. See Figure 27. Here we investigate locally tree-like random graphs with a given degree distribution $p(k)$.

We consider the largest connected component LCC of a graph G (see definition in section 2), and define f as the fraction of the N nodes in the graph that is *in the LLC*, so

$$size\ of\ the\ largest\ component: \quad |LCC| = fN \tag{105}$$

Note that f is then also the probability that a randomly drawn node from G is in the LCC.

- Definition: a graph G has a 'giant component' if $|LCC| = \mathcal{O}(N)$, i.e. if $f > 0$ and of order $\mathcal{O}(1)$.

Clearly, having a giant component will have a significant impact on processes running on the graph. Therefore we would like to calculate f. Since each randomly drawn node has degree distribution $p(k)$, we may use the following simple argument

$$1-f = \mathrm{Prob}(randomly\ drawn\ node\ \underline{not}\ in\ LCC)$$
$$= \sum_k p(k).\mathrm{Prob}(neighbours\ of\ randomly\ drawn\ node\ with\ degree\ k\ \underline{not}\ in\ LCC)$$

Since the graph is tree-like, the only possible correlations among local topological features of nodes in a generation g are those caused by their common ancestor (if any), so $\mathrm{Prob}(k\ neighbours\ are\ not\ in\ LCC) = (1-f)^k$, and hence we find that f is the solution of the simple equation $1-f = \sum_{k\geq 0} p(k)(1-f)^k$, or equivalently

$$f = 1 - G(1-f) \tag{106}$$

where $G(x)$ is the generating function (97). Since $G(1) = 1$ we see that $f = 0$ is always a solution of (106), describing a graph with a non-extensive largest component, since the

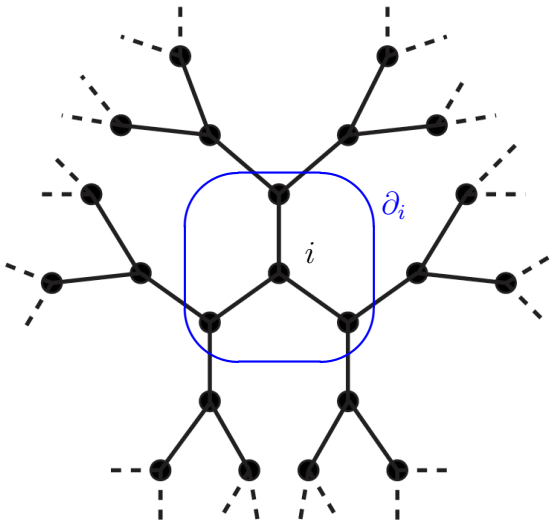


Figure 27. In locally tree-like graphs the number of short loops is vanishingly small, for the vast majority of the nodes the local topology is that of a tree. Hence, starting from a node i, we would find that the tree branches descending from i are nearly disconnected – they connect only at site i. In this example $p(k) = \delta_{k,3}$. With non-regular degree distributions we will see local randomness in this environment.

relative size of the giant component is fN. The next question is whether there are solutions of (106) with $f > 0$.

The solution of equation (106) can be found graphically by intersecting the function $F(f)$ on the interval $f \in [0, 1]$ with the diagonal, since this intersection has $f = F(f)$, in which

$$
\begin{aligned}
&F(f) = 1 - G(1 - f) \\
&\textit{regular random graphs},\ \langle k \rangle = q: \qquad F(f) = 1 - (1 - f)^q \\
&\textit{finitely connected ER graphs},\ \langle k \rangle = q: \qquad F(f) = 1 - \mathrm{e}^{-fq} \\
&\textit{exponential random graphs},\ \langle k \rangle = q: \qquad F(f) = qf/(1 + qf)
\end{aligned}
\tag{107}
$$

(where we used the generating functions calculated in the previous subsection). We observe that regular graphs are a special case. Here for $q = 1$ *all* values $f \in [0, 1]$ obey $f = F(q)$, and for $q > 1$ the value $f = 1$ is always a solution of $f = F(q)$.

The result of plotting the functions $F(f)$ on $f \in [0, 1]$ for the above choices for $p(k)$, and of solving numerically the equation $f = F(f)$ for $\langle k \rangle = q \in [0, 6]$ is shown in Figure 28. In all three cases the system undergoes a *percolation phase transition* at $\langle k \rangle = q = 1$, where the system switches from $f = 0$ (for $q < 1$) to $f > 0$ (for $q > 1$). Exactly at $\langle k \rangle = 1$ the networks develop a giant component. The size of this component is largest for regular graphs, and smallest for the exponential ones.

Universality of the percolation transition. It is no accident that in all three examples above

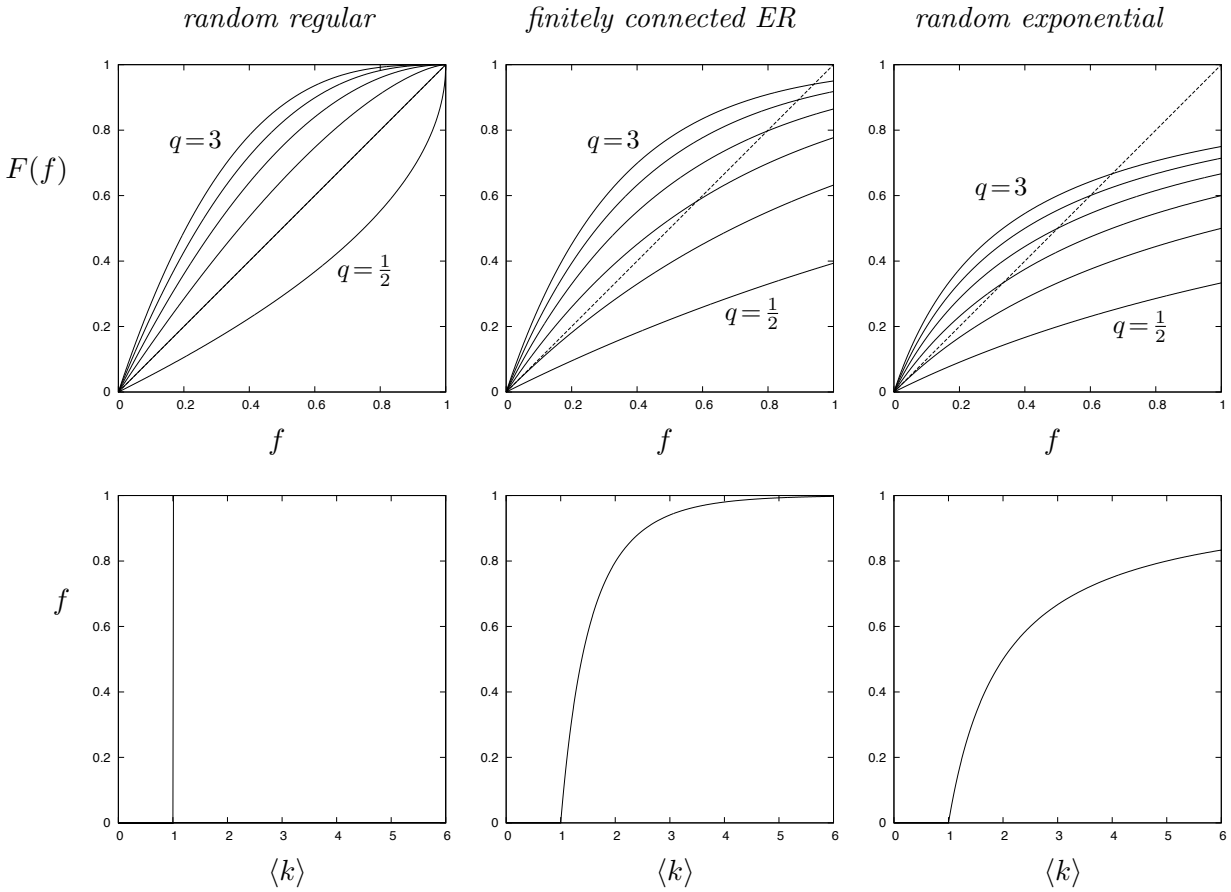


Figure 28. Top row: shapes of the function $F(f)$ for three random graph types with average degree $\langle k\rangle = q$: regular graphs with $p(k) = \delta_{kq}$, finitely connected Erdös-Rèyi graphs (for large N) with $p(k) = \mathrm{e}^{-q}q^k/k!$, and exponential graphs with $p(k) = (\frac{q}{1+q})^k/(1+q)$. Solid lines in each panel: $F(f)$ for $q \in \{\frac{1}{2}, 1, \frac{3}{2}, 2, \frac{5}{2}, 3\}$, from bottom to top. Dotted: the diagonal, whose intersection point with $F(f)$ marks the value f that gives the relative size of the largest connected component in the graph. Bottom row: corresponding expected relative sizes of the largest connected component (i.e. solutions of $F(f) = f$) in each of the there random graphs, shown as a function of the average degree $\langle k\rangle$. In all cases a giant component appears in the graphs at $\langle k\rangle = 1$, which is interpreted as a *percolation transition point*.

the percolation transition occurs at the same point $\langle k\rangle = 1$. The transition point correspond to the value of $\langle k\rangle$ where the trivial solution $f = 0$ of $F(f) = f$ ceases to be unique, this event is called a *bifurcation* of a new solution of (106). Figure 28 (bottom row) shows that, except for the degenerate case of regular graphs, the transition is generally *continuous*, i.e. there is no jump in the solution f when we vary $\langle k\rangle$. This prompts us to do a simple continuous bifurcation analysis of (106) near $f = 0$, by expansion in powers of f:

$$f = 1 - \sum_{k\geq 0} p(k)(1-f)^k = 1 - \sum_{k\geq 0} p(k) \sum_{\ell=0}^{k} \binom{k}{\ell} (-f)^\ell$$

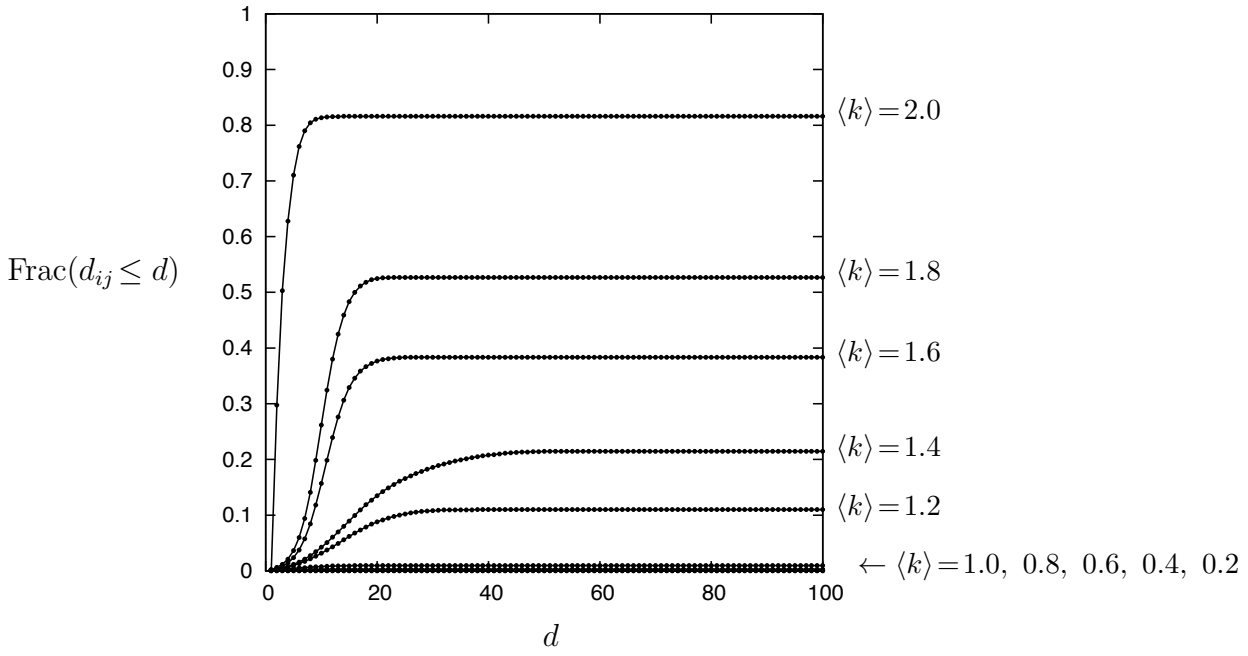


Figure 29. The fraction of node pairs (i, j) in randomly generated ER graphs that are found to have a distance $d_{ij} \leq d$, shown versus d. Here $N = 1000$ and $\langle k\rangle \in [0, 2]$. One observes clearly that there is a sharp transition at $\langle k\rangle \approx 1$. For $\langle k\rangle < 1$ the fraction of node pairs at a finite distance from each other is vanishing small. For $\langle k\rangle > 1$ this fraction all of sudden becomes finite, which marks the emergence of the giant component in the graph.

$$
\begin{aligned}
&= 1 - \sum_{\ell\geq 0} (-f)^\ell \sum_{k\geq \ell} p(k) \binom{k}{\ell} \\
&= 1 - \Big\{ \sum_{k\geq \ell} p(k) - f \sum_{k\geq 1} p(k) \binom{k}{1} + f^2 \sum_{k\geq 2} p(k) \binom{k}{2} + \mathcal{O}(f^3) \Big\} \\
&= f \sum_{k\geq 1} p(k) k - f^2 \sum_{k\geq 2} p(k) \frac{k!}{2(k-2)!} + \mathcal{O}(f^3) \\
&= f\langle k\rangle - \frac{1}{2} f^2 \sum_{k\geq 2} p(k) k(k-1) + \mathcal{O}(f^3) \\
&= f\langle k\rangle - \frac{1}{2} f^2 \sum_{k\geq 0} p(k) k(k-1) + \mathcal{O}(f^3) \\
&= f\langle k\rangle - \frac{1}{2} f^2 \langle k(k-1)\rangle + \mathcal{O}(f^3) \qquad (108)
\end{aligned}
$$

So our equation for f can be written as

$$
f\Big(1 - \langle k\rangle + \frac{1}{2} f \langle k(k-1)\rangle\Big) + \mathcal{O}(f^3) = 0 \qquad (109)
$$

For sufficiently small f we may neglect the cubic term and conclude

$$f = 0 \quad \text{or} \quad f = 2\frac{\langle k\rangle - 1}{\langle k^2\rangle - \langle k\rangle} \tag{110}$$

It can be shown (see exercises) that always $\langle k^2\rangle > \langle k\rangle$ unless we have a graph without links. Hence, as soon as $\langle k^2\rangle - \langle k\rangle < \infty$ a new solution $f > 0$ bifurcates exactly when $\langle k\rangle = 1$, irrespective of the shape of the degree distribution $p(k)$. Only for power law degree distributions, where $\langle k^2\rangle = \infty$, this bifurcation does not occur, and we have to evaluate (106) in full to study whether and when a giant component may appear.

In Figure 29 we show that the predicted emergence of a giant component at a sharply defined value of $\langle k\rangle$ is real, for numerically generated random Erdös-Rènyi graphs of size $N = 1000$, and with $\langle k\rangle \in [0, 2]$. We plot the fraction of node pairs (i, j) (with $i \neq j$) that have a distance $d_{ij} \leq d$, as a function of d. Clearly, there is a *phase transition* at $\langle k\rangle \approx 1$, where all of a sudden the typical distances between nodes in the graph become finite, marking the creation of the giant component.

Note that all calculations in this subsection were based on the assumption that the neighbours of any randomly picked node have degrees that are distributed *independently*. Just giving the degree distribution of a graph is not enough, as we have already seen in e.g. Figure 20; we need to specify the full ensemble probabilities $p(\boldsymbol{A})$ to check when this assumed degree independence holds. It is easy to construct graphs with $\langle k\rangle > 1$ (see exercises for an example of a regular graph with $\langle k\rangle = 2$) without a giant component. For the Erdös-Renyi ensemble, where we indeed have a fully specified $p(\boldsymbol{A})$, a more precise calculation can be done, and here (106) is indeed found to be exact in the limit $N \to \infty$.

6.6. Tailoring topologies via maximum entropy random graph ensembles

The ensemble (87,90) is the simplest nontrivial graph ensemble, and many of its properties are consequently rather different from those observed in real networks. Hence using the graphs produced by (87,90) as a statistical 'null models' to quantify the relevance of observations that we make in real networks is pointless – if our null model is too simple nearly anything we measure will seem special. So we need to think about this in a more sophisticated manner. Given an *observed* network with adjacency matrix $\boldsymbol{A}^\star$, we want to build random graph ensembles $\{\mathcal{G}, p\}$ that generate graphs with adjacency matrices $\boldsymbol{A}$ that are similar to $\boldsymbol{A}^\star$. How do we define similar? We choose specific measurements $\Omega_\mu(\boldsymbol{A}^\star)$, with $\mu = 1 \ldots K$, that we want our graphs to inherit from $\boldsymbol{A}^\star$, and demand that their values are either reproduced by *all* graphs in the ensemble (via a hard constraint), or that they are reproduced *on average* (via a soft constraint). See Figure 30. Abbreviating the set of chosen features as $\boldsymbol{\Omega}(\boldsymbol{A}) = (\Omega_1(\boldsymbol{A}), \ldots, \Omega_K(\boldsymbol{A}))$, this gives

hard constraints: graphs must have the features $\boldsymbol{\Omega}(\boldsymbol{A}^\star)$ *individually*,

$$\boldsymbol{\Omega}(\boldsymbol{A}^\star) = \boldsymbol{\Omega}(\boldsymbol{A}) \quad \text{for } \underline{\text{all}}\ \boldsymbol{A} \in \mathcal{G} \ \text{ with } \ p(\boldsymbol{A}) > 0 \tag{111}$$

soft constraints: graphs must have the features $\boldsymbol{\Omega}(\boldsymbol{A}^\star)$ *on average*,

$$\boldsymbol{\Omega}(\boldsymbol{A}^\star) = \sum_{\boldsymbol{A}\in\mathcal{G}} p(\boldsymbol{A})\boldsymbol{\Omega}(\boldsymbol{A}) \tag{112}$$

Both (111) and (112) give us conditions, but they do not yet specify the probabilities $p(\boldsymbol{A})$ in full, for that we need information-theoretic arguments. We want to construct an ensemble $\{\mathcal{G}, p\}$ that builds in the chosen features $\boldsymbol{\Omega}(\boldsymbol{A})$ but is otherwise strictly *noninformative and unbiased*. Given a choice for the set $\mathcal{G}$, this will be true only for the distribution $p(\boldsymbol{A})$ that, subject to the chosen constraints (111) or (112), maximises the Shannon entropy

$$S[p, \mathcal{G}] = -\sum_{\boldsymbol{A}\in\mathcal{G}} p(\boldsymbol{A}) \log p(\boldsymbol{A}) \tag{113}$$

The Shannon entropy defines the information content of a typical sample from the distribution $p(\boldsymbol{A})$. If we choose any $p(\boldsymbol{A})$ other than the one for which the entropy is maximal (subject to the chosen constraints), we would already know something about our graphs in addition to the information contained in our constraints. In other words: we would introduce an ad hoc bias into our graph ensembles. We now show how maximisation of (140), subject to the constraints, leads us to unique formulas for our tailored graph ensembles.

Tailored random graphs with hard constraints. If we go for (111), we need only define $p(\boldsymbol{A})$ for all graphs $\boldsymbol{A}$ that obey $\boldsymbol{\Omega}(\boldsymbol{A}) = \boldsymbol{\Omega}(\boldsymbol{A}^\star)$. All other graphs are forbidden. We choose the collection of allowed graphs to be our set $\mathcal{G}$, and we take the functions $\Omega_\mu(\boldsymbol{A})$ to be integer-valued. Maximisation of (140) subject to the constraint $\sum_{\boldsymbol{A}\in\mathcal{G}} p(\boldsymbol{A}) = 1$, using Lagrange's method (see Appendix **??**), then gives the following equations that are to be solved simultaneously for $\{p(\boldsymbol{A})\}$:

$$\begin{aligned}
\forall \boldsymbol{A} \in \mathcal{G}: \quad 0 &= \frac{\partial}{\partial p(\boldsymbol{A})}\Big[S[p,\mathcal{G}] + \lambda\Big(1 - \sum_{\boldsymbol{A}'\in\mathcal{G}} p(\boldsymbol{A}')\Big)\Big] \\
&= \frac{\partial}{\partial p(\boldsymbol{A})}\Big[-\sum_{\boldsymbol{A}'\in\mathcal{G}} p(\boldsymbol{A}')\log p(\boldsymbol{A}') - \lambda \sum_{\boldsymbol{A}'\in G} p(\boldsymbol{A}')\Big] \\
&= -\frac{\partial}{\partial p(\boldsymbol{A})}\Big[p(\boldsymbol{A})\log p(\boldsymbol{A} + \lambda p(\boldsymbol{A})\Big] = -\log p(\boldsymbol{A}) - 1 - \lambda
\end{aligned} \tag{114}$$

Hence $\log p(\boldsymbol{A}) = -1 - \lambda$ for all $\boldsymbol{A} \in \mathcal{G}$, or $p(\boldsymbol{A}) = \mathrm{e}^{-1-\lambda}$. Normalisation tells us that $\sum_{\boldsymbol{A}\in\mathcal{G}} p(\boldsymbol{A}) = 1$, hence $\mathrm{e}^{-1-\lambda} = |\mathcal{G}|^{-1}$. The end result is the following maximum entropy ensemble, in which all graphs $\boldsymbol{A}$ that meet our constraints are equally probable

$$\forall \boldsymbol{A} \in \mathcal{G}: \quad p(\boldsymbol{A}|\boldsymbol{\Omega}) = \frac{1}{Z(\boldsymbol{\Omega})}, \qquad Z(\boldsymbol{\Omega}) = |\mathcal{G}| \tag{115}$$

$Z(\boldsymbol{\Omega}) \geq 1$ is the number of graphs in $\{0,1\}^{N^2}$ that exhibit $\boldsymbol{\Omega}(\boldsymbol{A}) = \boldsymbol{\Omega}(\boldsymbol{A}^\star)$. We could also choose to work with the set $\mathcal{G} = \{0,1\}^{N^2}$ of all N-node graphs, and forbid those graphs $\boldsymbol{A}$

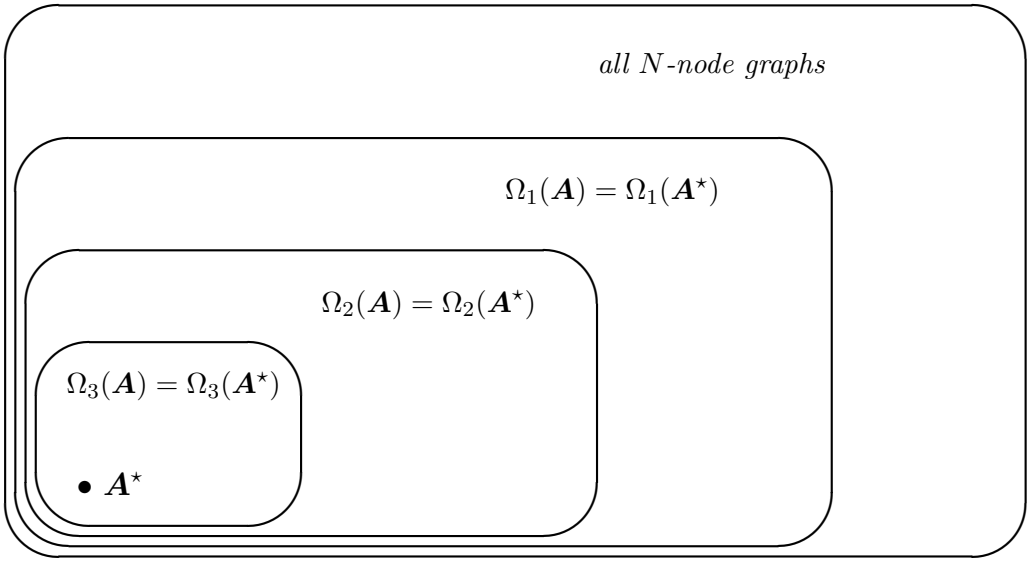


Figure 30. Systematic tailoring of random graph ensembles, in order to generate random graphs that resemble a given graph $\boldsymbol{A}^\star$, by prescribing that a series of K measurements $\Omega_\mu(\boldsymbol{A})$ give values identical to those found for $\boldsymbol{A}^\star$. The larger K, the smaller the 'box' of graphs that meet all K imposed conditions, and the more similar the graphs in the remaining box will be to $\boldsymbol{A}^\star$. The probabilities $p(\boldsymbol{A})$ assigned to all graphs $\boldsymbol{A}$ in each box are determined via entropy maximisation.

that do not satisfy the constraints by assigning to them $p(\boldsymbol{A}) = 0$. This gives‖

$$\forall \boldsymbol{A} \in \mathcal{G}: \quad p(\boldsymbol{A}|\boldsymbol{\Omega}) = \frac{\delta_{\boldsymbol{\Omega}(\boldsymbol{A}),\boldsymbol{\Omega}(\boldsymbol{A}^\star)}}{Z(\boldsymbol{\Omega})} \qquad Z(\boldsymbol{\Omega}) = \sum_{\boldsymbol{A}\in\mathcal{G}} \delta_{\boldsymbol{\Omega}(\boldsymbol{A}),\boldsymbol{\Omega}(\boldsymbol{A}^\star)} \tag{116}$$

Global constraints such as nondirectionality or simplicity can in principle be incorporated via the quantities $\Omega_\mu(\boldsymbol{A})$, but it is often more transparent to build such constraints into the definition of $\mathcal{G}$. Ensembles of the type (116) are also called *microcanonical* ensembles.

Tailored random graphs with soft constraints. If we choose the soft constraints (112), we have to build any global constraints into the set $\mathcal{G}$, and maximise the entropy (140) subject to $\sum_{\boldsymbol{A}\in\mathcal{G}} p(\boldsymbol{A}) = 1$ and $\sum_{\boldsymbol{A}\in\mathcal{G}} p(\boldsymbol{A})\Omega_\mu(\boldsymbol{A}) = \Omega_\mu(\boldsymbol{A}^\star)$ for all $\mu = 1\ldots K$. The Lagrange maximisation methods will now involve $K+1$ Lagrange parameters $\{\lambda_\mu\}$, K of which corresponding to the imposed constraints and with λ_0 representing the normalisation requirement $\sum_{\boldsymbol{A}\in\mathcal{G}} p(\boldsymbol{A}) = 1$. Our equations for $\{p(\boldsymbol{A})\}$ become

$$\forall \boldsymbol{A} \in \mathcal{G}: \quad 0 = \frac{\partial}{\partial p(\boldsymbol{A})}\Big[S[p,G] + \lambda_0\Big(1 - \sum_{\boldsymbol{A}'\in\mathcal{G}} p(\boldsymbol{A}')\Big) + \sum_{\mu=1}^{K} \lambda_\mu\Big(\Omega_\mu(\boldsymbol{A}^\star) - \sum_{\boldsymbol{A}'\in\mathcal{G}} p(\boldsymbol{A}')\Omega_\mu(\boldsymbol{A}')\Big)\Big]$$

‖ We use the standard notation convention $\delta_{\boldsymbol{x},\boldsymbol{y}} = \prod_\mu \delta_{x_\mu y_\mu}$.

$$= \frac{\partial}{\partial p(\boldsymbol{A})}\Big[- \sum_{\boldsymbol{A}'\in\mathcal{G}} p(\boldsymbol{A}')\log p(\boldsymbol{A}') - \lambda_0 \sum_{\boldsymbol{A}'\in\mathcal{G}} p(\boldsymbol{A}') - \sum_{\mu=1}^{K}\lambda_\mu \sum_{\boldsymbol{A}'\in\mathcal{G}} p(\boldsymbol{A}')\Omega_\mu(\boldsymbol{A}')\Big]$$

$$= -\frac{\partial}{\partial p(\boldsymbol{A})}\Big[p(\boldsymbol{A})\log p(\boldsymbol{A}) + \lambda_0 p(\boldsymbol{A}) + \sum_{\mu=1}^{K}\lambda_\mu p(\boldsymbol{A})\Omega_\mu(\boldsymbol{A})\Big]$$

$$= -1 - \log p(\boldsymbol{A}) - \lambda_0 - \sum_{\mu=1}^{K}\lambda_\mu\Omega_\mu(\boldsymbol{A}) \tag{117}$$

Thus the solution of Lagrange's equations takes the following form, after having determined the value of λ_0 from the normalisation requirement $\sum_{\boldsymbol{A}\in\mathcal{G}} p(\boldsymbol{A}) = 1$:

$$\forall \boldsymbol{A}\in\mathcal{G}: \qquad p(\boldsymbol{A}) = \frac{\mathrm{e}^{-\sum_{\mu=1}^{K}\lambda_\mu\Omega_\mu(\boldsymbol{A})}}{Z(\boldsymbol{\Omega})}, \qquad Z(\boldsymbol{\Omega}) = \sum_{\boldsymbol{A}\in\mathcal{G}} \mathrm{e}^{-\sum_{\mu=1}^{K}\lambda_\mu\Omega_\mu(\boldsymbol{A})} \tag{118}$$

in which the parameters $\{\lambda_1,\ldots,\lambda_K\}$ are found by solving the coupled equations

$$\forall \mu\in\{1,\ldots,K\}: \quad \Omega_\mu(\boldsymbol{A}^\star) = \sum_{\boldsymbol{A}\in\mathcal{G}} p(\boldsymbol{A})\Omega_\mu(\boldsymbol{A})$$
$$= \frac{1}{Z(\boldsymbol{\Omega})}\sum_{\boldsymbol{A}\in\mathcal{G}} \mathrm{e}^{-\sum_{\nu=1}^{K}\lambda_\nu\Omega_\nu(\boldsymbol{A})}\Omega_\mu(\boldsymbol{A}). \tag{119}$$

One can show that equations (119) have a unique solution, but solving them numerically is nontrivial. Even for simple nondirected graphs it involves sampling a space of size $|\mathcal{G}| = 2^{N(N-1)/2}$. Sometimes one can proceed analytically, if one chooses the $\Omega_\mu(\boldsymbol{A})$ wisely and N is large. The ensembles (118) are also called *exponential* or *canonical* ensembles.

In the above derivations of the hard-constrained and the soft-constrained ensembles we should in principle have included also the inequality constraints $p(\boldsymbol{A}) \geq 0$ for all $\boldsymbol{A}\in\mathcal{G}$ when maximising the Shannon entropy (140). However, it turned out in both cases that even without imposing them explicitly, the inequality constraints are satisfied automatically by the maximum entropy distributions. Hence they are obsolete.

Ensembles with constrained average degree. We first inspect ensembles of simple non-directed graphs in which the only information we choose to carry over from the given graph $\boldsymbol{A}^\star$ is the value $\bar{k} = N^{-1}\sum_{ij} A^\star_{ij}$ of the average degree, i.e. $K = 1$ and $\Omega_1(\boldsymbol{A}) = N^{-1}\sum_{ij} A_{ij}$. It turns out that the maximum entropy ensemble of simple nondirected graphs in which this average degree is imposed via a *soft constraint* is the Erdös-Rényi ensemble (87). See exercises. Imposing the average degree as a hard constraint, in contrast, leads to the following ensemble defined on the set $\mathcal{G}$ of all simple non directed N-node graphs:

$$p(\boldsymbol{A}) = \frac{1}{Z(\bar{k})}\,\delta_{\sum_{ij}A_{ij},N\bar{k}}, \qquad Z(\bar{k}) = \sum_{\boldsymbol{A}\in\mathcal{G}} \delta_{\sum_{ij}A_{ij},N\bar{k}} \tag{120}$$

Each graph with average degree $\bar{k}$ is equally likely, but for this ensemble not to be empty we need $N\bar{k}\in\mathbb{N}$; this will automatically be the case if $\bar{k} = N^{-1}\sum_i k_i(\boldsymbol{A}^\star)$ for some N-node

graph $\boldsymbol{A}^\star$. The factor $Z(\bar{k})$ gives the number of graphs in $\mathcal{G}$ with average degree $\bar{k}$.

Ensembles with hard-constrained degree sequences. The natural next step up is to impose instead of only the average degree the values of all N degrees $\boldsymbol{k} = (k_1, \ldots, k_N)$; typically these would be the degrees of an observed graph $\boldsymbol{A}^\star$. If the degrees $\boldsymbol{k}$ are imposed as hard constraints, so we allow only for graphs $\boldsymbol{A}$ for which $\boldsymbol{k}(\boldsymbol{A}) = \boldsymbol{k}$, we obtain an ensemble on the set $\mathcal{G}$ of simple non directed graphs that is called the 'configuration model':

$$p(\boldsymbol{A}|\boldsymbol{k}) = \frac{1}{Z(\boldsymbol{k})}\ \delta_{\boldsymbol{k},\boldsymbol{k}(\boldsymbol{A})}, \qquad Z(\boldsymbol{k}) = \sum_{\boldsymbol{A}\in\mathcal{G}} \delta_{\boldsymbol{k},\boldsymbol{k}(\boldsymbol{A})}. \tag{121}$$

All graphs $\boldsymbol{A}$ that have the degree sequence $\boldsymbol{k}$ are equally probable, and all graphs that do not have this degree sequence are assigned probability zero.

Ensembles with soft-constrained degree sequences. We obtain ensembles of simple nondirected graphs with soft-constrained degree sequences from the general formulae (118,119) by inserting the N degrees $k_i(\boldsymbol{A})$ for our observables $\Omega_\mu(\boldsymbol{A})$. We then find

$$p(\boldsymbol{A}) = \frac{\mathrm{e}^{-\sum_{i=1}^N \lambda_i \sum_{j=1}^N A_{ij}}}{Z(\boldsymbol{k})}, \qquad Z(\boldsymbol{k}) = \sum_{\boldsymbol{A}\in\mathcal{G}} \mathrm{e}^{-\sum_{i=1}^N \lambda_i \sum_{j=1}^N A_{ij}} \tag{122}$$

We can rewrite the probability distribution, using the symmetry of $\boldsymbol{A}$:

$$\begin{aligned} p(\boldsymbol{A}) &= \frac{\mathrm{e}^{-\sum_{i,j=1}^N A_{ij}\lambda_i}}{Z(\boldsymbol{k})} = \frac{\mathrm{e}^{-\frac{1}{2}\sum_{i,j=1}^N A_{ij}(\lambda_i+\lambda_j)}}{Z(\boldsymbol{k})} = \frac{\mathrm{e}^{-\sum_{i<j} A_{ij}(\lambda_i+\lambda_j)}}{Z(\boldsymbol{k})} \\ &= \frac{1}{Z(\boldsymbol{k})}\prod_{i<j}\mathrm{e}^{-A_{ij}(\lambda_i+\lambda_j)} = \frac{1}{Z(\boldsymbol{k})}\prod_{i<j}\left[\mathrm{e}^{-\lambda_i-\lambda_j}\delta_{A_{ij},1} + \delta_{A_{ij},0}\right] \end{aligned} \tag{123}$$

The normalisation sum $Z(\boldsymbol{k})$ then becomes

$$\begin{aligned} Z(\boldsymbol{k}) &= \sum_{\boldsymbol{A}\in\mathcal{G}}\prod_{i<j}\left[\mathrm{e}^{-\lambda_i-\lambda_j}\delta_{A_{ij},1} + \delta_{A_{ij},0}\right] = \prod_{i<j}\sum_{A_{ij}\in\{0,1\}}\left[\mathrm{e}^{-\lambda_i-\lambda_j}\delta_{A_{ij},1} + \delta_{A_{ij},0}\right] \\ &= \prod_{i<j}\left[1+\mathrm{e}^{-\lambda_i-\lambda_j}\right] \end{aligned} \tag{124}$$

In combination we can now write

$$p(\boldsymbol{A}) = \prod_{i<j}\left[\frac{\mathrm{e}^{-\lambda_i-\lambda_j}}{1+\mathrm{e}^{-\lambda_i-\lambda_j}}\delta_{A_{ij},1} + \frac{1}{1+\mathrm{e}^{-\lambda_i-\lambda_j}}\delta_{A_{ij},0}\right] \tag{125}$$

Finally, the equations (119) from which to solve the $\{\lambda_i\}$ now become $k_\ell = \sum_{\boldsymbol{A}\in\mathcal{G}} p(\boldsymbol{A})k_\ell(\boldsymbol{A})$ for all ℓ. This gives

$$\begin{aligned} \forall \ell \in \{1,\ldots,N\}: \quad k_\ell &= \sum_{\boldsymbol{A}\in\mathcal{G}}\prod_{i<j}\left[\frac{\mathrm{e}^{-\lambda_i-\lambda_j}}{1+\mathrm{e}^{-\lambda_i-\lambda_j}}\delta_{A_{ij},1} + \frac{1}{1+\mathrm{e}^{-\lambda_i-\lambda_j}}\delta_{A_{ij},0}\right]k_\ell(\boldsymbol{A}) \\ &= \sum_{r=1}^N \sum_{\boldsymbol{A}\in\mathcal{G}} A_{\ell r}\prod_{i<j}\left[\frac{\mathrm{e}^{-\lambda_i-\lambda_j}}{1+\mathrm{e}^{-\lambda_i-\lambda_j}}\delta_{A_{ij},1} + \frac{1}{1+\mathrm{e}^{-\lambda_i-\lambda_j}}\delta_{A_{ij},0}\right] \end{aligned}$$

$$= \sum_{r=1}^{N} \sum_{A_{\ell r}\in\{0,1\}} A_{\ell r}\Big[\frac{\mathrm{e}^{-\lambda_\ell-\lambda_r}}{1+\mathrm{e}^{-\lambda_\ell-\lambda_r}}\delta_{A_{\ell r},1} + \frac{1}{1+\mathrm{e}^{-\lambda_\ell-\lambda_r}}\delta_{A_{\ell r},0}\Big]$$

$$= \sum_{r=1}^{N} \frac{\mathrm{e}^{-\lambda_\ell-\lambda_r}}{1+\mathrm{e}^{-\lambda_\ell-\lambda_r}} = \sum_{r=1}^{N} \frac{1}{1+\mathrm{e}^{\lambda_\ell+\lambda_r}} \qquad (126)$$

In the same manner one can impose further topological features (degree correlations, number of loops, etc), via soft constraints, hard constraints, or even a mixture of the two.

7. Further topics

7.1. Graphicality of degree sequences

Not all degree sequences that one could write down are actually 'graphical': some choices for $\{k_1,\ldots,k_N\}$ cannot be realised in nondirected networks without introducing self-loops or multiple links. The simplest example of non-graphical degree sequences are those where $\sum_i k_i$ is odd. Here we will only discuss a few results on graphicality, without proofs.

A sufficient condition for a degree sequence to be graphical is provided by the so-called Erdös-Gallai theorem, which states that a degree sequence, with degrees ranked in decreasing order, i.e. $k_1 \geq k_2 \geq \ldots \geq k_N$, is graphical if $\sum_i k_i$ is even and it satisfies the following inequality for any integer n in the range $1 \leq n \leq N-1$:

$$\sum_{i=1}^{n} k_i \;\leq\; n(n-1) + \sum_{i=n+1}^{N} \min\{n, k_i\} \qquad (127)$$

Graphicality conditions can be violated quite easily. Degree sequences that are distributed according to a power-law distribution of the form $p(k) \sim k^{-\gamma}$, over a broad range of values bounded by $k_{\min} \leq k \leq k_{\max}$, are graphical for $k_{\max} = N-1$ and $k_{\min} = 1$ only if $\gamma > 2$. For $\gamma < 2$ only those degree sequences that scale with N as $k_{\max} < N^{1/\gamma}$ are graphical.

7.2. Watts-Strogatz and shifted-Poissonnian small-world networks

Various social experiments suggest that most people in the world are connected via only a small number of 'friends' links – typically between 3 and 10. This phenomenon has been called *Six Degrees of Separation* (after an idea in a 1929 story from a Hungarian writer). We can indeed see in Figure 29 that in large *random* graphs one needs just $\langle k\rangle \sim 2$ to achieve such short typical distances. However, in real-world social networks we expect that physical distances constrain us - we expect on average to have more friends living near us than friends at the other side of the globe. The 'small world' networks aim to shed light on this question: can we achieve the above short distances also in networks where most links are 'local'?

The Watts-Strogatz construction. This non-directed network model interpolates between a regular ring of N nodes with non-random links that are all short-ranged, and a random

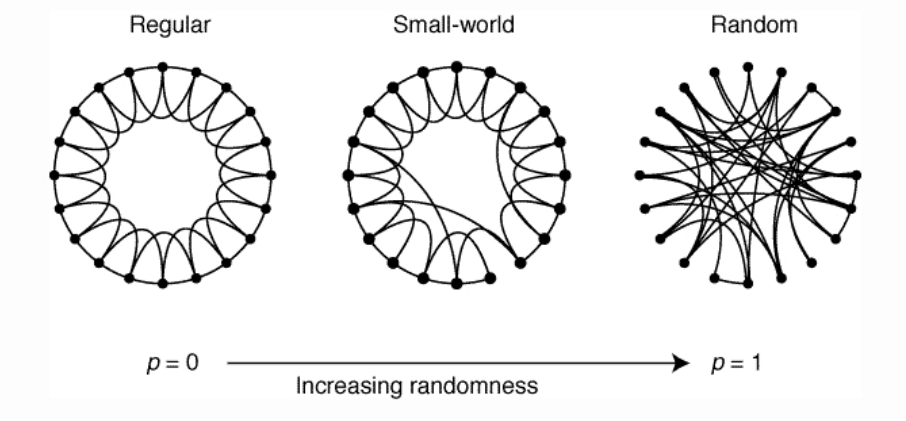


Figure 31. The Watts-Strogatz non-directed network model. Starting from a ring in which each node is connected to its $K \geq 2$ nearest neighbours (shown on the left, here with $K = 4$), one subsequently replaces a given fraction p of the original nearest neighbour edges with randomly selected alternative edges (not necessarily between neighbours in the ring). The total number of edges remains $\frac{1}{2}NK$, and the average connectivity is always $\bar{k}(\boldsymbol{A}) = K$.

graph. See Figure 31 for details of the construction. The relevant question is how the various distance-related quantities in the graph depend on the fraction p of links that have become random and long-range, as opposed to short-range.

Superposition construction. An alternative construction that is very similar but more easily analysed mathematically is the following. Again all N nodes are initially placed on a ring. We then (a) connect all nearest neighbours, and (b) superimpose on this an Erdös-Rènyi graph with average degree $\langle k \rangle = K - 2$. Provided N is large and K is small, the likelihood of a proposed random link coinciding with an existing nearest neighbour link on the ring is negligible – if it happens we simply don't add the random one. We then obtain the following adjacency matrix $\boldsymbol{A}$, built from the deterministic ring $\boldsymbol{A}^{\rm D}$ and the random ER graph $\boldsymbol{A}^{\rm R}$:

$$A_{ij} = A^{\rm D}_{ij} + A^{\rm R}_{ij} - A^{\rm D}_{ij}.A^{\rm R}_{ij} \tag{128}$$

$$A^{\rm D}_{ij} = \delta_{i,j+1 \ mod \ N} + \delta_{i,j-1 \ mod \ N} \tag{129}$$

$$p(\boldsymbol{A}^{\rm R}) = \prod_{i<j} \Big(\frac{K-2}{N-1} \delta_{A^{\rm R}_{ij},1} + (1 - \frac{K-2}{N-1}) \delta_{A^{\rm R}_{ij},0} \Big) \tag{130}$$

Inspection reveals (see e.g. Figure 32) that only a small fraction of the edges in the graph need to be non-local to achieve the 'small world' effect.

7.3. Preferential attachment networks

Not all of the panels in Figure 18 are equally convincing straight lines, but the WWW, the internet, and the power grid data do suggest power law degree distributions with powers in the range that give scale-free graphs. Preferential attachment networks are models of graphs

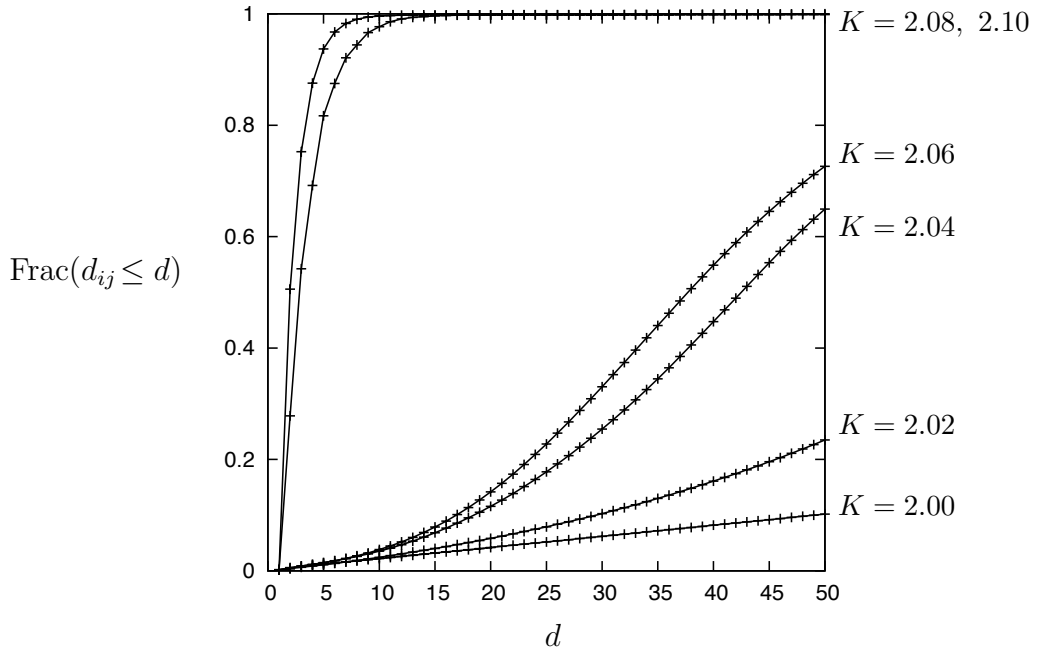


Figure 32. The fraction of node pairs (i, j) that are found to have a distance $d_{ij} \leq d$, shown versus d, measured in small world graphs created by superimposing an Erdös-Rènyi graph with average degree $K-2$ upon a ring with nearest-neighbour interactions only. The combined graph has expected average degree $\langle k \rangle = K$. For $K = 2$ we would have only short-range links. Here $N = 1000$ and $\langle k \rangle \in [2.0, 2.1]$, and each curve refers to a single graph (hence there are fluctuations). One observes a relatively sharp transition – at around $K = 2.07$ where there is still only a small fraction of 'long-range' links – to graphs with the short typical distances that characterise the small-world effect.

that seek too provide a generative explanation for power-law degree distributions, which strictly random graphs (such as Erdös-Rènyi type graphs) would not give. One possible explanation turns out to be that scale-free graphs result from growth processes in which new links are more likely to be attached to nodes that already have a high degree. Although usually attributed to Albert and Barabasi, the basic idea goes back to earlier studies.

The preferential attachment process is a stochastic growth process for a simple non-directed graph, whereby at each step we add one node and one or more links in a very specific way. There are many different versions, the one below is just the simplest:

initiation: create the first node $i = 1$, with no links.

iterate for $\ell = 1, 2, 3, \ldots$*:*

- ℓ is the present number of nodes, add <u>one</u> new node $\ell+1$

- generate randomly <u>one</u> link from the new node to one of the ℓ existing nodes. The node to connect to is selected with probabilities

$$\forall i \leq \ell : \qquad \text{Prob}(A_{i,\ell+1} = 1) = \frac{k_i(\ell-1)}{\sum_{j=1}^{\ell} k_j(\ell-1)} \tag{131}$$

Here $k_j(\ell)$ is the degree of node j <u>after</u> ℓ steps of the process. We terminate if the size of the network is what we want it to be.

Next we show that it leads to a power-law degree distribution. Let us first summarise a few facts regarding the number of nodes N and links L, and the degrees.

(i) At step ℓ: we go from $(N, L) = (\ell, \ell-1)$ to $(N, L) = (\ell+1, \ell)$.

(ii) After ℓ steps we have $N = \ell+1$ nodes and $L = \ell$ links.

(iii) The most recently created node always has degree 1, so $k_{\ell+1}(\ell) = 1$ for all ℓ

(iv) The average degree after ℓ steps is $\frac{1}{\ell+1}\sum_{j=1}^{\ell+1} k_j(\ell) = 2\ell/(\ell+1)$

From (iv) we conclude that $\sum_{j=1}^{\ell+1} k_j(\ell) = 2\ell$ for all ℓ. Therefore the selection probability (131) at step ℓ can be written as $\text{Prob}(A_{i,\ell+1} = 1) = k_i(\ell-1)/2(\ell-1)$.

The degrees evolve stochastically, so we must introduce the probability $p_\ell(k|i)$ that after step ℓ we will have $k_i(\ell) = k$. It is defined only for $i \leq \ell+1$, and we know that $p_\ell(k|\ell+1) = \delta_{k1}$. At each step ℓ of the process, the degree of each node $i \leq \ell$ either stays the same, this happens with probability $1 - \text{Prob}(A_{i,\ell+1}=1)$, or increases by one, which happens with probability $\text{Prob}(A_{i,\ell+1}=1)$. Hence

$$\forall i \leq \ell : \qquad p_\ell(k|i) = \overbrace{p_{\ell-1}(k|i)\,\Big(1 - \frac{k}{2(\ell-1)}\Big)}^{\textit{no degree increase at step } \ell} + \overbrace{p_{\ell-1}(k-1|i)\,\Big(\frac{k-1}{2(\ell-1)}\Big)}^{\textit{degree increase at step } \ell} \tag{132}$$

Next we move to the overall degree distribution after ℓ steps, defined as

$$p_\ell(k) = \frac{1}{\ell+1}\sum_{i=1}^{\ell+1} p_\ell(k|i) = \frac{1}{\ell+1}\sum_{i=1}^{\ell} p_\ell(k|i) + \frac{1}{\ell+1}\delta_{k1} \tag{133}$$

Inserting (132) into the right-hand side of this latter expression gives

$$\begin{aligned} p_\ell(k) &= \frac{1}{\ell+1}\sum_{i=1}^{\ell}\Big\{p_{\ell-1}(k|i)\Big(1 - \frac{k}{2(\ell-1)}\Big) + p_{\ell-1}(k-1|i)\frac{k-1}{2(\ell-1)}\Big\} + \frac{1}{\ell+1}\delta_{k1} \\ &= \frac{\ell}{\ell+1}\Big(1 - \frac{k}{2(\ell-1)}\Big)\frac{1}{\ell}\sum_{i=1}^{\ell} p_{\ell-1}(k|i) + \frac{\ell}{\ell+1}\frac{k-1}{2(\ell-1)}\frac{1}{\ell}\sum_{i=1}^{\ell} p_{\ell-1}(k-1|i) + \frac{1}{\ell+1}\delta_{k1} \\ &= \frac{\ell}{\ell+1}\Big(1 - \frac{k}{2(\ell-1)}\Big)p_{\ell-1}(k) + \frac{\ell}{\ell+1}\frac{k-1}{2(\ell-1)}p_{\ell-1}(k-1) + \frac{1}{\ell+1}\delta_{k1} \end{aligned} \tag{134}$$

So now we have a closed dynamical equation for the evolving overall degree distribution. Let us check that the probabilities in our equation remain properly normalised. Suppose

$\sum_{k\geq 0} p_{\ell-1}(k) = 1$. Summation over all k in our equation gives:

$$\begin{aligned}\sum_{k\geq 0} p_\ell(k) &= \frac{\ell}{\ell+1}\sum_{k\geq 0}\big(1-\frac{k}{2(\ell-1)}\big)p_{\ell-1}(k) + \frac{\ell}{\ell+1}\sum_{k\geq 0}\frac{k-1}{2(\ell-1)}p_{\ell-1}(k-1) + \frac{1}{\ell+1}\\ &= \frac{\ell}{\ell+1} - \frac{\ell}{\ell+1}\frac{1}{2(\ell-1)}\sum_{k\geq 0} k p_{\ell-1}(k) + \frac{\ell}{\ell+1}\frac{1}{2(\ell-1)}\sum_{k\geq 0}(k+1)p_{\ell-1}(k)\\ &\qquad\qquad - \frac{\ell}{\ell+1}\frac{1}{2(\ell-1)} + \frac{1}{\ell+1}\\ &= \frac{\ell}{\ell+1} + \frac{1}{\ell+1} = 1 \end{aligned} \tag{135}$$

Hence normalisation is preserved, as it should. In the same manner one can confirm form the dynamical equation (134) that the average degree after ℓ iterations is indeed given by $2\ell/(\ell+1)$ (see Exercises).

The final step is to find the asymptotic solution of (134). To do this we introduce a real-valued time variable $t_\ell = \ell\tau$, with $0 < \tau \ll 1$, and adapt our notation according to $p_\ell(k) \to p_{t_\ell}(k)$. Our equation then becomes

$$\begin{aligned} p_{t_\ell}(k) - p_{t_{\ell-1}}(k) &= \frac{\ell}{\ell+1}\frac{k-1}{2(\ell-1)}p_{t_{\ell-1}}(k-1) - \frac{\ell}{\ell+1}\frac{k}{2(\ell-1)}p_{t_{\ell-1}}(k) - \frac{1}{\ell+1}p_{t_{\ell-1}}(k) + \frac{1}{\ell+1}\delta_{k1}\\ \frac{p_{t_\ell}(k) - p_{t_{\ell-1}}(k)}{\tau} &= \frac{1}{t_\ell+\tau}\Big\{\frac{k-1}{2(1-\tau/t_\ell)}p_{t_{\ell-1}}(k-1) - \frac{k}{2(1-\tau/t_\ell)}p_{t_{\ell-1}}(k) - p_{t_{\ell-1}}(k) + \delta_{k1}\Big\}\\ \frac{p_{t_\ell}(k) - p_{t_\ell-\tau}(k)}{\tau} &= \frac{1}{t_\ell+\tau}\Big\{\frac{k-1}{2(1-\tau/t_\ell)}p_{t_\ell-\tau}(k-1) - \frac{k}{2(1-\tau/t_\ell)}p_{t_\ell-\tau}(k) - p_{t_\ell-\tau}(k) + \delta_{k1}\Big\} \end{aligned} \tag{136}$$

We now take the limit $\tau \to 0$ (assuming that it exists) and find

$$t\frac{\mathrm{d}}{\mathrm{d}t}p_t(k) = \frac{1}{2}(k-1)p_t(k-1) - \frac{1}{2}(k+2)p_t(k) + \delta_{k1} \tag{137}$$

Stationary solutions of this equation are apparently to be solved from

$$(k-1)p(k-1) - (k+2)p(k) + 2\delta_{k1} = 0 \tag{138}$$

One can confirm (see Exercises) that this equation is solved by:

$$p(0) = 0, \qquad k > 0: \quad p(k) = \frac{4}{k(k+1)(k+2)} \tag{139}$$

This is clearly a power law degree distribution, with $p(k) \sim k^{-3}$ for large k. The average degree is finite, but $\sum_k k^2 p(k)$ diverges, so the width of $p(k)$ is infinite (see Exercises). Variations on the above preferential attachment process include e.g. starting with more than one disconnected node, or making multiple links at each iteration step.

7.4. Complexity – counting graphs

Graph complexity can be quantified as follows. A graph with features $\boldsymbol{\Omega}(\boldsymbol{A})$ is more complex than one with features $\boldsymbol{\Omega}'(\boldsymbol{A})$, if there exist fewer graphs with features $\boldsymbol{\Omega}(\boldsymbol{A})$ then graphs

with features $\boldsymbol{\Omega}'(\boldsymbol{A})$. The more 'unique' are a graph's characteristics, the more complex is this graph. There is an intimate connection with the Shannon entropy of information theory. The Shannon entropy of a discrete random variable x with probability distribution $p(x)$ is defined (apart from an overall constant $\ln 2$) as $S = -\sum_x p(x) \log p(x)$. The Shannon entropy of a random graph ensemble with macroscopic characteristics $\boldsymbol{\Omega}(\boldsymbol{A})$ is

$$S(\boldsymbol{\Omega}) = -\sum_{\boldsymbol{A}} p(\boldsymbol{A}|\boldsymbol{\Omega}) \log p(\boldsymbol{A}|\boldsymbol{\Omega}) \tag{140}$$

For the hard-constrained ensembles (115) this gives $S(\boldsymbol{\Omega}) = \sum_{\boldsymbol{A}} \delta_{\boldsymbol{\Omega},\boldsymbol{\Omega}(\boldsymbol{A})} Z^{-1}(\boldsymbol{\Omega}) \log Z(\boldsymbol{\Omega}) = \log Z(\boldsymbol{\Omega})$. With the definition of $Z(\boldsymbol{\Omega})$ as given in (115) we then obtain

$$\sum_{\boldsymbol{A}} \delta_{\boldsymbol{\Omega},\boldsymbol{\Omega}(\boldsymbol{A})} = \exp[S(\boldsymbol{\Omega})] \tag{141}$$

The left-hand side is exactly the *number* of graphs that have features $\boldsymbol{\Omega}(\boldsymbol{A}) = \boldsymbol{\Omega}$, so we can define the *complexity* of typical graphs with $\boldsymbol{\Omega}(\boldsymbol{A}) = \boldsymbol{\Omega}$ simply as the logarithm of this number, i.e. simply as the Shannon entropy (140) of the ensemble. The same definition (140) can then be used also for the soft-constrained ensembles (118).

Total number of N-node graphs. The total number of N-node graphs is simply the number of ways we can choose the binary entries of the adjacency matrix. For directed graphs this $Z = 2^{N^2}$, for non-directed graphs we find $Z = 2^{\frac{1}{2}N(N+1)}$, and for simple non-directed graphs we get $Z = 2^{\frac{1}{2}N(N-1)}$. So in all cases $\log Z \sim \frac{1}{2}N^2$ in leading order for large N. In the finite connectivity regime, however, we allow only for a finite number of nonzero entries per row and per column of the adjacency matrix. This reduces the number of possibilities. The question is: by how much? Below we study this problem for simple-nondirected graphs, similar results can be derived for directed ones (see also the exercises).

Simple non-directed graphs with specified average degree. Let us calculate the number of N-node simple nondirected graphs that have average degree $\bar{k}$, given that $N\bar{k} \in \mathbb{N}$. This is simply $Z(\bar{k})$ in the ensemble (120). Upon writing the Kronecker symbol in $Z(\bar{k})$ in integral form, and upon choosing $\mathcal{G}$ to be the set of all simple non-directed graphs, we obtain

$$\begin{aligned}
Z(\bar{k}) &= \sum_{\boldsymbol{A}\in\mathcal{G}} \delta_{N\bar{k},\sum_{ij} A_{ij}} = \int_{-\pi}^{\pi} \frac{\mathrm{d}\omega}{2\pi}\, \mathrm{e}^{\mathrm{i}\omega N\bar{k}/2} \sum_{\boldsymbol{A}\in\mathcal{G}} \mathrm{e}^{-\mathrm{i}\omega \sum_{i<j} A_{ij}} \\
&= \int_{-\pi}^{\pi} \frac{\mathrm{d}\omega}{2\pi}\, \mathrm{e}^{\mathrm{i}\omega N\bar{k}/2} \prod_{i<j} \sum_{A_{ij}\in\{0,1\}} \mathrm{e}^{-\mathrm{i}\omega A_{ij}} = \int_{-\pi}^{\pi} \frac{\mathrm{d}\omega}{2\pi}\, \mathrm{e}^{\mathrm{i}\omega N\bar{k}/2} (1+\mathrm{e}^{-\mathrm{i}\omega})^{N(N-1)/2} \\
&= \int_{-\pi}^{\pi} \frac{\mathrm{d}\omega}{2\pi}\, \mathrm{e}^{\mathrm{i}\omega N\bar{k}/2} \sum_{m=0}^{N(N-1)/2} \left(\begin{array}{c} N(N-1)/2 \\ m \end{array} \right) \mathrm{e}^{-\mathrm{i}\omega m} \\
&= \sum_{m=0}^{N(N-1)/2} \left(\begin{array}{c} N(N-1)/2 \\ m \end{array} \right) \delta_{m,N\bar{k}/2} = \left(\begin{array}{c} N(N-1)/2 \\ N\bar{k}/2 \end{array} \right)
\end{aligned} \tag{142}$$

This result can be understood easily, as it gives the number of possible ways in which one can draw $N\bar{k}/2$ links from a set of $N(N-1)/2$ possible candidates.

For large N and finite $\bar{k}$ (the finite connectivity regime) we can inspect the leading orders in N, using Stirling's formula $\log n! \approx n\log n - n + \mathcal{O}(\log n)$ for $n \to \infty$. If both n and m are large (with $m \ll n$), this formula, together with $\log(1+x) = x + \mathcal{O}(x^2)$, allows us to write

$$\begin{aligned}
\log\Big(\frac{n!}{(n-m)!\ m!}\Big) &= \log n! - \log m! - \log(n-m)! \\
&\approx n\log n - (n-m)\log(n-m) - m\log m + \mathcal{O}(\log n, \log m) \\
&= n\Big\{\log n - (1-\frac{m}{n})\log(n-m) - \frac{m}{n}\log m\Big\} + \mathcal{O}(\log n, \log m) \\
&= n\Big\{\log n - (1-\frac{m}{n})\log n - (1-\frac{m}{n})\log(1-\frac{m}{n}) - \frac{m}{n}\log m\Big\} + \mathcal{O}(\log n, \log m) \\
&= n\Big\{\frac{m}{n}\log\frac{n}{m} + \frac{m}{n}\Big\} + \mathcal{O}(\log n, \log m) + \mathcal{O}(\frac{m^2}{n^3}) \\
&= m\log\frac{n}{m} + m + \mathcal{O}(\log n, \log m) + \mathcal{O}(\frac{m^2}{n^3}) \qquad (143)
\end{aligned}$$

Application of this expansion to $n = \frac{1}{2}N(N-1)$ and $m = \frac{1}{2}N\bar{k}$ then gives us

$$\begin{aligned}
N^{-1}\log Z(\bar{k}) &= \frac{1}{2}\bar{k}\log[(N-1)/\bar{k}] + \frac{1}{2}\bar{k} + \mathcal{O}(\frac{\log N}{N}) \\
&= \frac{1}{2}\bar{k}\log(N/\bar{k}) + \frac{1}{2}\bar{k} + \mathcal{O}(\frac{\log N}{N}) \qquad (144)
\end{aligned}$$

Thus for large N the leading order of the number of graphs with finite average degree $\bar{k}$ still grows super-exponentially as $Z(\bar{k}) \sim \exp(\frac{1}{2}\bar{k}N\log(N) + \ldots)$. This means that one can never hope to sample numerically the space of all such graphs, even for relatively modest sizes N. For instance, working out the formula gives

$$N = 32,\ \bar{k} = 2: \quad Z(\bar{k}) \approx 2.7\ 10^{52} \qquad (145)$$

$$N = 50,\ \bar{k} = 2: \quad Z(\bar{k}) \approx 4.1\ 10^{91} \qquad (146)$$

These are very big numbers (for comparison: the total number of atoms in the observable universe is estimated to be somewhere between 10^{78} and 10^{82}).

Simple non-directed graphs with specified degree distribution. The next stage is to impose not only the average degree $\bar{k}$, but the full degree distribution $p(k)$. The calculation is not fundamentally different from the previous one but does require some new tools (saddle-point integration), so here we mention only the result:

$$\frac{1}{N}\log Z(\{p(k)\}) = \frac{1}{2}\bar{k}\log(N/\bar{k}) + \frac{1}{2}\bar{k} - \sum_k p(k)\log[\frac{p(k)}{\pi(k)}] + \mathcal{O}(\frac{\log N}{N}) \qquad (147)$$

Here $\bar{k} = \sum_k kp(k)$, and $\pi(k) = \mathrm{e}^{-\bar{k}}\bar{k}^k/k!$ (the degree distribution of Erdös-Rènyi graphs with average degree $\bar{k}$). Comparison to (144) shows that, due to the prescribed degree distribution, $N^{-1}\log Z$ has been reduced by an amount which is the Kullback-Leibler distance (see Information Theory) between $p(k)$ and the degree distribution of an ER graph with the

same $\bar{k}$. We see that if we impose the degree statistics $p(k) = \pi(k)$ for all k, formula (147) reduces to (144). Hence

$$\frac{Z(\{p(k)\})}{Z(\{\pi(k)\})} = \mathrm{e}^{-N\sum_k p(k)\log[p(k)/\pi(k)]+\mathcal{O}(\log N)} \tag{148}$$

This tells us that, for large N, nearly all graphs with a given finite average degree $\bar{k}$ have a Poissonnian degree distribution.

Simple non-directed graphs with specified degree distribution and specified degree correlations. We can in the same way prescribe further information beyond the degree distribution, such as the degree correlation kernel $W(k,k')$. This turns out to give

$$\begin{aligned}\frac{1}{N}\log Z(\{p(k),W(k,k')\}) = \frac{1}{2}\bar{k}\log(N/\bar{k}) + \frac{1}{2}\bar{k} - \sum_k p(k)\log[\frac{p(k)}{\pi(k)}] \\ - \frac{1}{2}\bar{k}\sum_{k,k'} W(k,k')\log\big[\frac{W(k,k')}{W(k)W(k')}\big] + \mathcal{O}(\frac{\log N}{N})\end{aligned} \tag{149}$$

Comparison to (147) shows that prescribing $W(k,k')$ reduces the quantity $N^{-1}\log Z$ further by an amount which is proportional to the mutual information (see Information Theory) of the degrees of connected nodes. We see that if we impose that degrees of connected nodes are uncorrelated, i.e. $W(k,k') = W(k)W(k')$ for all (k,k'), formula (149) reduces to (147). Hence nearly all graphs with a given degree distribution $p(k)$ have uncorrelated degrees.

8. Appendices

8.1. Network software

The following software resources for imaging and/or analysis of networks, created within the academic community, are free:

- *Cytoscape*: www.cytoscape.org
- *Gephi*: http://gephi.github.io
- *R – with igraph package*: www.r-project.org
- *Pajek*: vlado.fmf.uni-lj.si/pub/networks/pajek/

8.2. The Pearson correlation

The Pearson correlation of two random variables (u, v) with joint distribution $P(u, v)$ is defined as

$$\mathrm{PC} = \frac{\langle uv\rangle - \langle u\rangle\langle v\rangle}{\sqrt{(\langle u^2\rangle - \langle u\rangle^2)(\langle v^2\rangle - \langle v\rangle^2)}} \tag{150}$$

It tests for statistical dependence in the form a (partially) linear relationship between u and v. To get some intuition for this, let us work out two simple extreme cases:

- *Statistically independent u and v*

  Now $P(u, v) = P(u)P(v)$, and hence

  $$\langle uv\rangle = \sum_{uv} P(u,v)uv = \sum_{uv}\ P(u)P(v)uv = \Big(\sum_{u}\ P(u)u\Big)\Big(\sum_{v}\ P(v)v\Big) = \langle u\rangle\langle v\rangle$$

  Hence we obtain $\mathrm{PC} = 0$.

- *Linearly related u and v*

  Suppose $u = \lambda v + c$ for all combinations (u, v). Now we obtain

  $$\begin{aligned}\langle uv\rangle &= \langle v(\lambda v + c)\rangle = \lambda\langle v^2\rangle + c\langle v\rangle\\ \langle u^2\rangle &= \langle(\lambda v + c)^2\rangle = \lambda^2\langle v^2\rangle + c^2 + 2\lambda c\langle v\rangle\\ \langle u\rangle &= \langle\lambda v + c\rangle = \lambda\langle v\rangle + c\end{aligned}$$

  Inserting all this into formula (150) then leads to

  $$\begin{aligned}\mathrm{PC} &= \frac{\langle uv\rangle - \langle u\rangle\langle v\rangle}{\sqrt{(\langle u^2\rangle - \langle u\rangle^2)(\langle v^2\rangle - \langle v\rangle^2)}}\\ &= \frac{\lambda\langle v^2\rangle + c\langle v\rangle - \lambda\langle v\rangle^2 - c\langle v\rangle}{\sqrt{\lambda^2\langle v^2\rangle + c^2 + 2\lambda c\langle v\rangle - \lambda^2\langle v\rangle^2 - c^2 - 2\lambda c\langle v\rangle}\sqrt{\langle v^2\rangle - \langle v\rangle^2}}\\ &= \frac{\lambda\Big(\langle v^2\rangle - \langle v\rangle^2\Big)}{|\lambda|\sqrt{\langle v^2\rangle - \langle v\rangle^2}\sqrt{\langle v^2\rangle - \langle v\rangle^2}} = \ \mathrm{sgn}(\lambda)\end{aligned}$$

Hence if u and v are perfectly positively linearly correlated we find PC $= 1$, and if they are perfectly negatively linearly related we find PC $= -1$.

8.3. Properties of symmetric matrices

Eigenvectors and eigenvalues. We derive some properties of real symmetric $N \times N$ matrices $\boldsymbol{A}$. The eigenvalue polynomial $\det(\boldsymbol{A} - \lambda \mathbb{I}) = 0$ is of order N, so $\boldsymbol{A}$ will have N (possibly complex) solutions λ (where some may coincide) of the eigenvalue problem

$$\boldsymbol{A}\boldsymbol{x} = \lambda \boldsymbol{x}, \qquad \boldsymbol{x} \neq \boldsymbol{0} \tag{151}$$

We denote complex conjugation of complex numbers z in the usual way: if $z = a + ib$ (where $a, b \in \mathbb{R}$), then $z^* = a - ib$ and $|z|^2 = z^* z \in \mathbb{R}$. The inner product on $\mathbb{C}^N$ is $\boldsymbol{x} \cdot \boldsymbol{y} = \sum_i x_i^* y_i$.

- Claim: all eigenvalues of the matrix $\boldsymbol{A}$ are real.

  Proof:

  Take the inner product in (151) with the conjugate vector $\boldsymbol{x}^*$, which gives

  $$\sum_{i,j=1}^{N} x_i^* A_{ij} x_j = \lambda \sum_{i=1}^{N} |x_i|^2$$

  We use the symmetry of $\boldsymbol{A}$, and substitute $A_{ij} \to \frac{1}{2}(A_{ij} + A_{ji})$:

  $$\lambda = \frac{1}{2} \frac{\sum_{ij} x_i^* (A_{ij} + A_{ji}) x_j}{\sum_{i=1}^{N} |x_i|^2} = \frac{1}{2} \frac{\sum_{ij} A_{ij} (x_i^* x_j + x_i x_j^*)}{\sum_{i=1}^{N} |x_i|^2}$$

  Since $(x_i^* x_j + x_i x_j^*)^* = x_i x_j^* + x_i^* x_j = x_i^* x_j + x_i x_j^*$, the above fraction is real-valued.

- Claim: all eigenvectors of the matrix $\boldsymbol{A}$ can be chosen real-valued.

  Proof:

  We separate real and imaginary parts of every eigenvector:

  $$\boldsymbol{x} = \text{Re } \boldsymbol{x} + \text{i} \text{Im } \boldsymbol{x} \qquad \text{Re } \boldsymbol{x} = \frac{1}{2}(\boldsymbol{x} + \boldsymbol{x}^*) \qquad \text{Im } \boldsymbol{x} = \frac{1}{2\text{i}}(\boldsymbol{x} - \boldsymbol{x}^*)$$

  with Re $\boldsymbol{x} \in \mathbb{R}^N$ and Im $\boldsymbol{x} \in \mathbb{R}^N$. Complex conjugation of (151) gives $\boldsymbol{A}\boldsymbol{x}^* = \lambda \boldsymbol{x}^*$ (since λ is real). Hence, if $\boldsymbol{x}$ is an eigenvector with eigenvalue λ, so is $\boldsymbol{x}^*$. By adding/subtracting the conjugate equation to/from (151) it follows: if $\boldsymbol{x}$ and $\boldsymbol{x}^*$ are eigenvectors, so are Re $\boldsymbol{x}$ and Im $\boldsymbol{x}$. Since the space spanned by $\boldsymbol{x}$ and $\boldsymbol{x}^*$ is the same as the space spanned by Re $\boldsymbol{x}$ and Im $\boldsymbol{x}$, we are always allowed to choose the real-valued pair Re $\boldsymbol{x}$ and Im $\boldsymbol{x}$.

- Claim: for every linear subspace $L \subseteq \mathbb{R}^N$ the following holds:

  $$\text{if} \quad \boldsymbol{A}L \subseteq L \quad \text{then also} \quad \boldsymbol{A}L^\perp \subseteq L^\perp$$

  in which $L^\perp$ denotes the orthogonal complement, i.e. $\mathbb{R}^N = L \oplus L^\perp$.

Proof:

For each $\boldsymbol{x} \in L$ and $\boldsymbol{y} \in L^\perp$ we find $(\boldsymbol{x} \cdot \boldsymbol{A}\boldsymbol{y}) = (\boldsymbol{y} \cdot \boldsymbol{A}\boldsymbol{x}) = 0$ (since $\boldsymbol{A}\boldsymbol{x} \in L$ and $\boldsymbol{y} \in L^\perp$). Therefore $\boldsymbol{A}\boldsymbol{y} \in L^\perp$, which completes the proof.

- Claim: we can construct a complete orthogonal basis in $\mathbb{R}^N$ of $\boldsymbol{A}$-eigenvectors.

  Proof:

  Consider two eigenvectors $\boldsymbol{x}_a$ and $\boldsymbol{x}_b$ of $\boldsymbol{A}$, corresponding to different eigenvalues:

  $$\boldsymbol{A}\boldsymbol{x}_a = \lambda_a \boldsymbol{x}_a \qquad \boldsymbol{A}\boldsymbol{x}_b = \lambda_b \boldsymbol{x}_b \qquad \lambda_a \neq \lambda_b$$

  We now form:

  $$\begin{aligned} 0 &= (\boldsymbol{x}_a \cdot \boldsymbol{A}\boldsymbol{x}_b) - (\boldsymbol{x}_a \cdot \boldsymbol{A}\boldsymbol{x}_b) = (\boldsymbol{x}_a \cdot \boldsymbol{A}\boldsymbol{x}_b) - (\boldsymbol{x}_b \cdot \boldsymbol{A}\boldsymbol{x}_a) \\ &= \lambda_b(\boldsymbol{x}_a \cdot \boldsymbol{x}_b) - \lambda_a(\boldsymbol{x}_b \cdot \boldsymbol{x}_a) = (\lambda_a - \lambda_b)(\boldsymbol{x}_a \cdot \boldsymbol{x}_b) \end{aligned}$$

  Since $\lambda_a \neq \lambda_b$ it follows that $\boldsymbol{x}_a \cdot \boldsymbol{x}_b = 0$. If all eigenvalues are distinct, this completes the proof, since now there will be N eigenvalues with eigenvectors $\boldsymbol{x} \neq 0$. Since these N eigenvectors are orthogonal, after normalization they form a complete orthogonal basis.

  To deal with degenerate eigenvalues we need the third property above. If $A\boldsymbol{x} = \lambda\boldsymbol{x}$, then $\forall \boldsymbol{y}$ with $\boldsymbol{x} \cdot \boldsymbol{y} = 0$: $(A\boldsymbol{y}) \cdot \boldsymbol{x} = 0$. Having found an eigenvector for eigenvalue λ (not unique in the case of a degenerate eigenvalue), a new reduced $(N-1) \times (N-1)$ matrix can be constructed by restricting ourselves to the subspace $\boldsymbol{x}^\perp$. The new matrix is again symmetric, the eigenvalue polynomial is of order $N-1$ (and contains all the previous roots except for one corresponding to the eigenvector just eliminated), and we can repeat the argument. This shows that there *must* again be N orthogonal eigenvectors.

Basis of eigenvectors and diagonal form. The final consequence of the above facts is that there exist a set of N vectors $\{\hat{\boldsymbol{e}}^i\}$, where $i = 1, \ldots, N$ and $\hat{\boldsymbol{e}}^i \in \mathbb{R}^N$ for all i, with the following properties:

$$\boldsymbol{A}\hat{\boldsymbol{e}}^i = \lambda_i \hat{\boldsymbol{e}}^i, \qquad \lambda_i \in \mathbb{R}, \qquad \lambda_i > 0, \qquad \hat{\boldsymbol{e}}^i \cdot \hat{\boldsymbol{e}}^j = \delta_{ij} \tag{152}$$

We can now bring $\boldsymbol{A}$ onto diagonal form by a simple unitary transformation $\boldsymbol{U}$, which we construct from the components of the normalised eigenvectors $\hat{\boldsymbol{e}}$: $U_{ij} = \hat{e}^j_i$. We denote the transpose of $\boldsymbol{U}$ by $\boldsymbol{U}^\dagger$, $U^\dagger_{ij} = U_{ji}$, and show that $\boldsymbol{U}$ is indeed unitary, i.e. $\boldsymbol{U}^\dagger\boldsymbol{U} = \boldsymbol{U}\boldsymbol{U}^\dagger = \mathbb{1}$:

$$\begin{aligned} \sum_j (\boldsymbol{U}^\dagger\boldsymbol{U})_{ij} x_j &= \sum_{jk} U^{ki} U_{kj} x_j = \sum_{jk} \hat{e}^i_k \hat{e}^j_k x_j = \sum_j \delta_{ij} x_j = x_i \\ \sum_j (\boldsymbol{U}\boldsymbol{U}^\dagger)_{ij} x_j &= \sum_{jk} U^{ik} U_{jk} x_j = \sum_{jk} \hat{e}^k_i \hat{e}^k_j x_j = \sum_k \hat{e}^k_i (\hat{\boldsymbol{e}} \cdot \boldsymbol{x}) = x_i \end{aligned}$$

(since $\{\hat{\boldsymbol{e}}^\ell\}$ is a complete orthogonal basis). From $\boldsymbol{U}$ being unitary it follows that $\boldsymbol{U}$ and $\boldsymbol{U}^\dagger$ leave inner products, and therefore also lengths, invariant:

$$(\boldsymbol{U}\boldsymbol{x}) \cdot (\boldsymbol{U}\boldsymbol{y}) = \boldsymbol{x} \cdot \boldsymbol{U}^\dagger\boldsymbol{U}\boldsymbol{y} = \boldsymbol{x} \cdot \boldsymbol{y} \qquad \boldsymbol{U}^\dagger\boldsymbol{x} \cdot \boldsymbol{U}^\dagger\boldsymbol{y} = \boldsymbol{x} \cdot \boldsymbol{U}\boldsymbol{U}^\dagger\boldsymbol{y} = \boldsymbol{x} \cdot \boldsymbol{y}$$

We can see explicitly that $\boldsymbol{U}$ indeed brings $\boldsymbol{A}$ onto diagonal form:

$$(\boldsymbol{U}^\dagger \boldsymbol{A}\boldsymbol{U})_{ij} = \sum_{kl=1}^{N} U^\dagger_{ik} A_{kl} U_{lj} = \sum_{kl=1}^{N} \hat{e}^i_k A_{kl} \hat{e}^j_l = \lambda_j \sum_{k=1}^{N} \hat{e}^i_k \hat{e}^j_k = \lambda_j \delta_{ij} \tag{153}$$

Note that the inverse $\boldsymbol{A}^{-1}$ of the matrix $\boldsymbol{A}$ exists, and can be written as follows:

$$(\boldsymbol{A}^{-1})_{ij} = \sum_{k=1}^{N} \lambda_k^{-1} \hat{e}^k_i \hat{e}^k_j \tag{154}$$

To prove that this is the inverse of $\boldsymbol{A}$, we work out for any $\boldsymbol{x} \in \mathbb{R}^N$ the two expressions

$$(\boldsymbol{A}\boldsymbol{A}^{-1}\boldsymbol{x})_i = \sum_{kj=1}^{N} A_{ik} \sum_{\ell=1}^{N} \lambda_\ell^{-1} \hat{e}^\ell_k \hat{e}^\ell_j x_j = \sum_{\ell=1}^{N} \hat{e}^\ell_i (\hat{\boldsymbol{e}}^\ell \cdot \boldsymbol{x}) = x_i$$

(again since $\{\hat{\boldsymbol{e}}^\ell\}$ forms a complete orthogonal basis), and

$$(\boldsymbol{A}^{-1}\boldsymbol{A}\boldsymbol{x})_i = \sum_{kj=1}^{N} \sum_{\ell=1}^{N} \lambda_\ell^{-1} \hat{e}^\ell_i \hat{e}^\ell_k A_{kj} x_j = \sum_{\ell=1}^{N} \hat{e}^\ell_i (\hat{\boldsymbol{e}}^\ell \cdot \boldsymbol{x}) = x_i$$

8.4. Integral representation of the Kronecker δ-symbol

Here we show that for any $n, m \in \mathbb{Z}$ the Kronecker δ can be written in the integral form

$$\delta_{nm} = \int_{-\pi}^{\pi} \frac{\mathrm{d}\omega}{2\pi}\, \mathrm{e}^{\mathrm{i}(n-m)\omega} \tag{155}$$

To see this one simply does the integral on the right:

$$\begin{aligned}
n = m: \quad & \int_{-\pi}^{\pi} \frac{\mathrm{d}\omega}{2\pi}\, \mathrm{e}^{\mathrm{i}(n-m)\omega} = \int_{-\pi}^{\pi} \frac{\mathrm{d}\omega}{2\pi} = 1 \\
n \neq m: \quad & \int_{-\pi}^{\pi} \frac{\mathrm{d}\omega}{2\pi}\, \mathrm{e}^{\mathrm{i}(n-m)\omega} = \int_{-\pi}^{\pi} \frac{\mathrm{d}\omega}{2\pi} \Big(\cos((n-m)\omega) + \mathrm{i}\sin((n-m)\omega) \Big) \\
& \qquad = \frac{1}{2\pi} \Big[\frac{1}{n-m} \sin((n-m)\omega) - \frac{\mathrm{i}}{n-m} \cos((n-m)\omega) \Big]_{\omega=-\pi}^{\omega=\pi} = 0
\end{aligned}$$

8.5. The Landau order symbol

Let $f(x)$ and $g(x)$ be two functions of a variable x that is taken to zero, such that $\lim_{x\to 0} f(x) = \lim_{x\to 0} g(x) = 0$. We then define the order symbol $\mathcal{O}$ as follows:

$$f(x) = \mathcal{O}(g(x)) \;\text{ for }\; x \to 0 \quad \Leftrightarrow \quad (\exists C > 0, \epsilon > 0)(\forall |x| < \epsilon): \quad |f(x)/g(x)| < C \tag{156}$$

In words: asymptotically for $x \to 0$, $f(x)$ decays to zero equally fast or faster than $g(x)$. Similarly we could use it to characterise the behaviour of functions that diverge, such as

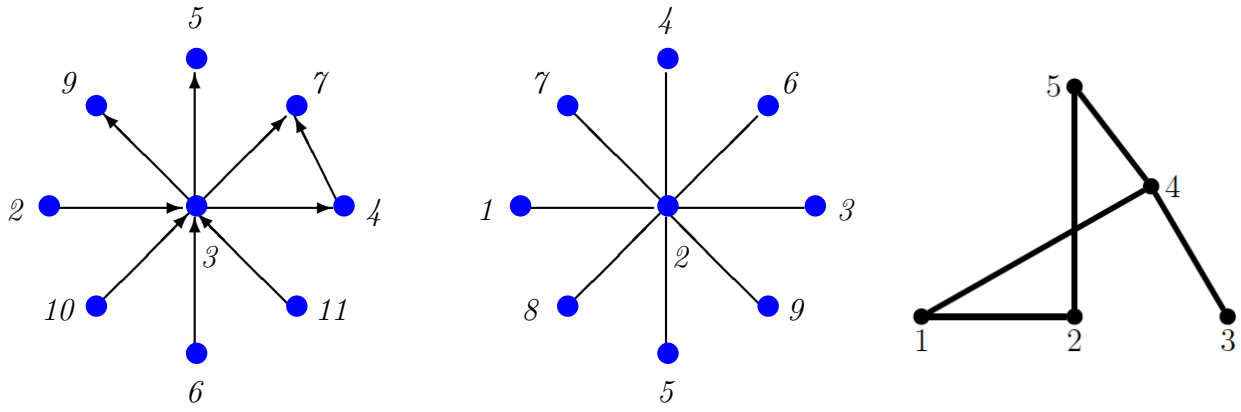
$$f(x) = \mathcal{O}(g(x)) \;\text{ for }\; x \to \infty \quad \Leftrightarrow \quad (\exists C > 0, X > 0)(\forall x > X): \quad |f(x)/g(x)| < C \tag{157}$$

In words: asymptotically for $x \to \infty$, $f(x)$ diverges equally fast or slower than $g(x)$.

9. Exercises

Tutorial 1

1. Which if the three graphs below is simple? Which of them is directed? Give for each these graphs the vertex set V and the edge set E.



2. Calculate the adjacency matrices for each of the three graphs above, upon relabelling the nodes of the first graph such that its vertex set becomes $V = \{1, \ldots, 9\}$.
3. Use the adjacency matrices calculated in the previous exercise to prove that the first of the three graphs has exactly four paths of length three and no paths of length 4 or larger. Argue why we can be sure that the middle and right graphs will contain paths of any length $\ell > 0$.
4. Calculate all in-degrees and all out-degrees of the above three graphs.
5. Prove that in nondirected graphs always $k_i^{\text{in}}(\boldsymbol{A}) = k_i^{\text{out}}(\boldsymbol{A})$.
   Prove that for non-directed graphs one always has $k_i(\boldsymbol{A}) = (\boldsymbol{A}^2)_{ii}$.
6. Show that in any simple non-directed graph with adjacency matrix $\boldsymbol{A}$ one has $C_i(\boldsymbol{A}) = 2T_i(\boldsymbol{A})/k_i(\boldsymbol{A})[k_i(\boldsymbol{A}) - 1]$, where $T_i(\boldsymbol{A})$ is the number of triangles in which node i participates, and $k_i(\boldsymbol{A})$ is its degree.
7. Calculate the clustering coefficients for all nodes in the second and the third of of the above graphs. Why would we not calculate them for the first graph?

Tutorial 2

8. Define the neighbourhood ∂_i of a node i in a nondirected graph with adjacency matrix $\boldsymbol{A}$ as follows: $\partial_i = \{j \leq N |\ A_{ij} = 1\}$. Show that the order-2 generalised degrees can be written as $k_i^{(2)}(\boldsymbol{A}) = \sum_{j\in\partial_i} k_j^{\rm in}(\boldsymbol{A})$. Calculate all order-2 generalised degrees $k_i^{(2)}(\boldsymbol{A})$ of the second and third graph above.

9. Prove that in directed graphs always $L = \sum_i k_i^{\rm in}(\boldsymbol{A}) = \sum_i k_i^{\rm out}(\boldsymbol{A})$. Show that in a simple non-directed graph one has $L = \frac{1}{2}N\bar{k}(\boldsymbol{A})$.

10. Verify that the order-ℓ generalised degrees in a nondirected graph with adjacency matrix $\boldsymbol{A}$ can be written as

$$k_i^{(\ell)}(\boldsymbol{A}) = \sum_{j=1}^{N} (\boldsymbol{A}^{\ell-1})_{ij} k_j(\boldsymbol{A})$$

Show that the corresponding expressions for directed graphs are, with $(\boldsymbol{A}^\dagger)_{ij} = A_{ji}$:

$$k_i^{(\ell){\rm in}}(\boldsymbol{A}) = \sum_{j=1}^{N} (\boldsymbol{A}^{\ell-1})_{ij} k_j^{\rm in}(\boldsymbol{A}), \qquad k_i^{(\ell){\rm out}}(\boldsymbol{A}) = \sum_{j=1}^{N} (\boldsymbol{A}^{\dagger\ \ell-1})_{ij} k_j^{\rm out}(\boldsymbol{A})$$

11. Prove the matrix identity $(\mathbf{I} - \gamma\boldsymbol{A})^{-1} = \sum_{\ell\geq 0} \gamma^\ell \boldsymbol{A}^\ell$. Given any matrix norm $|\boldsymbol{A}|$ that satisfies the usual conditions (i.e. $|\boldsymbol{A}| \in \mathbb{R}^+$, $|\lambda\boldsymbol{A}| = |\lambda||\boldsymbol{A}|$ for any $\lambda \in \mathbb{R}$, $|\boldsymbol{A}| = 0$ if and only if $\boldsymbol{A} = \mathbf{0}$, $|\boldsymbol{A}^1 + \boldsymbol{A}^2| \leq |\boldsymbol{A}^1| + |\boldsymbol{A}^2|$), show that there is always a sufficiently small but nonzero value of γ such that the series $\sum_{\ell\geq 0} \gamma^\ell \boldsymbol{A}^\ell$ converges in norm.

12. Consider the N-node graph with $N > 2$ and the following adjacency matrix entries: $A_{ij} = \delta_{i,j+1\ mod\ N} + \delta_{i,j-1\ mod\ N}$. Prove that for this graph

$$r, s \in \{1, \ldots, N\}: \qquad [(\mathbf{1} - \gamma\boldsymbol{A})^{-1}]_{rs} = \frac{1}{N} \sum_{\ell=0}^{N-1} \frac{{\rm e}^{2\pi {\rm i}\ell(r-s)/N}}{1 - 2\gamma\cos(2\pi\ell/N)}$$

(you may find the geometric series helpful in proving this, as well as the series $(1-\epsilon)^{-1} = \sum_{m\geq 0} \epsilon^m$; see Calculus lectures).

Tutorial 3

13. Calculate the closeness centrality and the betweenness centrality of nodes $i = 2$ and $i = 3$ in the second and the third of of the graphs in exercise (i).

14. Show that the definition of the Pearson correlation similarity $\tau_{ij}(\boldsymbol{A})$ can be derived from the definition of the Pearson correlation of two random variables (u, v), upon choosing $P(u, v) = \frac{1}{N}\sum_k \delta_{u,A_{ik}}\delta_{v,A_{jk}}$.

15. Let $\boldsymbol{x}\cdot\boldsymbol{y}$ denote an inner product on $\mathbb{R}^N$, so that it meets the defining criteria:

    (i) $(\forall \boldsymbol{x},\boldsymbol{y},\boldsymbol{z}\in\mathbb{R}^N):\ (\boldsymbol{x}+\boldsymbol{y})\cdot\boldsymbol{z} = \boldsymbol{x}\cdot\boldsymbol{z}+\boldsymbol{y}\cdot\boldsymbol{z}$
    (ii) $(\forall \boldsymbol{x}\cdot\boldsymbol{y}\in\mathbb{R}^N)(\forall\lambda\in\mathbb{R}):\ \boldsymbol{x}\cdot(\lambda\boldsymbol{y}) = \lambda\boldsymbol{x}\cdot\boldsymbol{y}$
    (iii) $(\forall \boldsymbol{x}\cdot\boldsymbol{y}\in\mathbb{R}^N):\ \boldsymbol{x}\cdot\boldsymbol{y} = \boldsymbol{y}\cdot\boldsymbol{x}$
    (iv) $(\forall \boldsymbol{x}\in\mathbb{R}^N):\ \boldsymbol{x}\cdot\boldsymbol{x}\geq 0$, with equality if and only if $\boldsymbol{x} = \boldsymbol{0}$

    Prove the Schwartz inequality: $|\boldsymbol{x}\cdot\boldsymbol{y}| \leq |\boldsymbol{x}||\boldsymbol{y}|$. Hint: calculate $|\boldsymbol{x}+\lambda\boldsymbol{y}|^2|\boldsymbol{y}|^2$ with $\lambda\in\mathbb{R}$, and choose a clever value for λ at the end.

16. Explain why the two expressions given for the cosine similarity $\sigma_{ij}(\boldsymbol{A})$ of two nodes i and j in a nondirected graph are identical. Show that $|\sigma_{ij}(\boldsymbol{A})| \leq 1$, and that $\sigma_{ij}(\boldsymbol{A}) = 1$ if and only if $\partial_i = \partial_j$. Hint: define for each node i the vector $\boldsymbol{a}^{(i)} = (A_{i1}, A_{i2}, \ldots, A_{iN}) \in \{0,1\}^N$, and write $\sigma_{ij}(\boldsymbol{A})$ in terms of the two vectors $\boldsymbol{a}^{(i)}$ and $\boldsymbol{a}^{(j)}$.

17. Explain why the two expressions given for the Pearson correlation similarity $\tau_{ij}(\boldsymbol{A})$ of two nodes i and j in a nondirected graph are identical. Show that $|\tau_{ij}(\boldsymbol{A})| \leq 1$. Hint: define for each node i the vector $\boldsymbol{a}^{(i)}$ with entries $a^{(i)}_k = [A_{ik} - \frac{1}{N}\sum_\ell A_{i\ell}]/\sqrt{N}$, and write $\tau_{ij}(\boldsymbol{A})$ in terms of $\boldsymbol{a}^{(i)}$ and $\boldsymbol{a}^{(j)}$.

18. Prove that the *average* in-degree $\overline{k}^{\rm in}(\boldsymbol{A}) = N^{-1}\sum_i k_i^{\rm in}(\boldsymbol{A})$ and the *average* out-degree $\overline{k}^{\rm out}(\boldsymbol{A}) = N^{-1}\sum_i k_i^{\rm out}(\boldsymbol{A})$ of any graph (directed or non-directed) are always identical. Show that the total number of links in a directed graph can be written as either $L = \sum_{i=1}^N k_i^{\rm in}(\boldsymbol{A})$ or $L = \sum_{i=1}^N k_i^{\rm out}(\boldsymbol{A})$. Show that in simple nondirected graphs the number of links is $L = \frac{1}{2}N\bar{k}(\boldsymbol{A})$, where $\bar{k}(\boldsymbol{A}) = N^{-1}\sum_{i\leq N} k_i(\boldsymbol{A})$ is the average degree.

Tutorial 4

19. Prove the following identities for the density $\rho(\boldsymbol{A})$ of a graph with adjacency matrix $\boldsymbol{A}$:

$$\begin{aligned} &\textit{directed graphs}: && \rho(\boldsymbol{A}) = \overline{k}(\boldsymbol{A})/N \\ &\textit{nondirected graphs}: && \rho(\boldsymbol{A}) = \overline{k}(\boldsymbol{A})/(N{+}1) + \sum_i A_{ii}/N(N{+}1) \\ &\textit{simple nondirected graphs}: && \rho(\boldsymbol{A}) = \overline{k}(\boldsymbol{A})/(N{-}1) \end{aligned}$$

20. Calculate the diameter $d(\boldsymbol{A})$ and the degree distribution $p(k|\boldsymbol{A})$ for the middle and the right graph in question (i).

21. Consider the following degree distribution for an infinitely large non-directed graph: $p(k) = \mathrm{e}^{-q}q^k/k! \quad \forall k \in \mathbb{N}$. Calculate the average degree $\langle k\rangle = \sum_{k\geq 0} p(k)k$ and the degree variance $\sigma_k^2 = \langle k^2\rangle - \langle k\rangle^2$.

22. Consider the following degree distribution for an infinitely large non-directed graph: $p(k) = C\mathrm{e}^{-k} \quad \forall k \in \mathbb{N}$. Give a formula for the constant C. Calculate the average degree $\langle k\rangle = \sum_{k\geq 0} p(k)k$ and the degree variance $\sigma_k^2 = \langle k^2\rangle - \langle k\rangle^2$.

23. Consider the following degree distribution for an N-node non-directed graph: $p(k) = 0$ for $k = 0$ or $k > N$, and $p(k) = C_N k^{-\gamma}$ for $0 < k \leq N$. Calculate C_N. For which γ values is $p(k)$ normalisable for $N \to \infty$? Give formulas for $\langle k\rangle$ and the degree variance $\sigma_k^2 = \sum_{k\geq 0} p(k)k^2 - \langle k\rangle^2$. For which γ values is $\langle k\rangle$ finite in the limit $N \to \infty$? For what values of γ is the variance finite for $N \to \infty$? Give an estimate of the average and the variance for $\gamma = 2.5$ and $N = 10{,}000$, using the approximation $\sum_{k=1}^{N} k^{-\lambda} \approx \int_1^N \mathrm{d}k\ k^{-\lambda}$.

24. Calculate the degree distributions for the N-node graphs with the following adjacency matrices (check carefully whether they are directed or non-directed, and use the correct degree distribution definition in each case):
    (a) $A_{ij} = \delta_{i,j+1}$ for $j < N$, and $A_{iN} = 0$.
    (b) $A_{ij} = 1$ for all $i, j \in \{1, \ldots, N\}$
    (c) $A_{ij} = 0$ for all $i, j \in \{1, \ldots, N\}$
    (d) $A_{ij} = 1$ if either $i, j \in \{1, \ldots, N/2\}$ or $i, j \in \{N/2{+}1, \ldots, N\}$; $A_{ij} = 0$ otherwise
    (e) $A_{i1} = 1$ for all $i > 1$, $A_{ij} = 0$ for all other (i, j).

Tutorial 5

25. Show that the degree correlation ratio $\Pi(k, k'|\boldsymbol{A})$ of a 'regular' simple non-directed graph $\boldsymbol{A}$, i.e. one with $p(k|\boldsymbol{A}) = \delta_{k,k^\star}$ for some $k^\star \in \mathbb{N}$, is always equal to 1 for any (k, k').
26. Prove that $W_1(\vec{k}|\boldsymbol{A}) = p(\vec{k}|\boldsymbol{A})k^{\rm in}/\bar{k}(\boldsymbol{A})$ and that $W_2(\vec{k}'|\boldsymbol{A}) = p(\vec{k}'|\boldsymbol{A})k^{\rm out\prime}/\bar{k}(\boldsymbol{A})$.
27. Prove the following general bounds for the modularity: $-\frac{1}{2} \leq Q(\boldsymbol{A}) \leq \frac{1}{2}$.
28. Assign the following module labels to the nodes of the right graph in exercise 1: $x_1 = x_2 = 1$, $x_3 = x_4 = x_5 = 2$. Calculate the graph's modularity $Q(\boldsymbol{A})$. Next turn to graphs (b) and (d) in exercise 24. Assign the following module labels to the nodes: $x_i = 1$ for $i \leq N/2$ and $x_i = 2$ for $i > N/2$ (take N to be even). Calculate the modularity $Q(\boldsymbol{A})$ for both graphs.
29. Show how the total number of triangles $T(\boldsymbol{A})$ in a simple non-directed N-node graph can be calculated directly from the spectrum $\{\mu_1(\boldsymbol{A}), \ldots, \mu_N(\boldsymbol{A})\}$ of its adjacency matrix.
30. Calculate the adjacency matrix eigenvalue spectrum $\{\mu_1(\boldsymbol{A}), \ldots, \mu_N(\boldsymbol{A})\}$ of the middle graph in exercise (i). Use your result to calculate the average degree, and to prove that this graph has no closed paths of odd length.

Tutorial 6

31. Calculate the Laplacian matrix $\boldsymbol{L}$ of the middle graph in exercise 1, and its eigenvalue spectrum. Hint: write $\boldsymbol{L} = \mathbb{1} + \boldsymbol{B}$ and first find the eigenvalues of $\boldsymbol{B}$, where $\mathbb{1}$ is the unity matrix. Use your result to prove that this graph has only one connected component.
32. Use the results of the previous exercise to solve the dynamical equations describing a diffusion process on the middle graph in exercise 1, that starts with $z_i(t) = z_0\delta_{i2}$ (i.e. diffusion from the central node $i = 2$). Verify that your result makes sense for $t = 0$ and in the limit $t \to \infty$. Verify that the quantity $\sum_i z_i(t)$ is conserved over time.
33. Show that for regular N-node graphs, i.e. those for which all N degrees $k_i(\boldsymbol{A})$ are identical, one can express the Laplacian eigenvalue spectrum in terms of the adjacency matrix eigenvalue spectrum. Give the mathematical relation between the two spectra in explicit form.
34. Consider the N-node graph with the following adjacency matrix entries, with $N > 2$: $A_{ij} = \delta_{i,j+1 \ mod \ N} + \delta_{i,j-1 \ mod \ N}$. Calculate the adjacency matrix spectrum $\varrho(\mu|\boldsymbol{A})$ and the Laplacian spectra and $\varrho_{\mathrm{Lap}}(\mu|\boldsymbol{A})$. Hints: use the result of the previous exercise, and try Fourier modes $x_k = \mathrm{e}^{\mathrm{i}\omega k}$ as an ansatz for the eigenvectors. Confirm that the smallest eigenvalue of the Laplacian is zero, and use the spectrum to determine the number of connected components in the graph.

Tutorial 7

35. For the Erdös-Rènyi model we know that $\langle \bar{k}(\boldsymbol{A})\rangle = p^\star(N-1)$. Calculate $\langle \bar{k}^2(\boldsymbol{A})\rangle$. Calculate the variance $\sigma^2_{\bar{k}} = \langle \bar{k}^2(\boldsymbol{A})\rangle - \langle \bar{k}(\boldsymbol{A})\rangle^2$ in the finite connectivity regime, and express it in terms of $\langle k\rangle$ for $N \to \infty$. What can you conclude from the result?
36. Show for the following degree distribution that $\sum_{k\geq 0} p(k) = 1$ and $\sum_{k\geq 0} p(k)k = q$, without using its generating function:
$$p(k) = (\frac{q}{1+q})^k/(1+q) \tag{158}$$
37. Calculate the generating function $G(x)$ for the following degree distribution, with $\alpha \in [0,1]$ and $q_1, q_2 \in \mathbb{N}$:
$$p(k) = \alpha\delta_{k,q_1} + (1-\alpha)\mathrm{e}^{-q_2} q_2^k/k!$$
38. Confirm that the three generating functions for regular, Poissonnian, and exponential random graphs all obey: $G(0) = p(0)$, $G(1) = 1$, and $\lim_{x\to 1} x\frac{\mathrm{d}}{\mathrm{d}x}G(x) = \langle k\rangle$. Calculate expressions for $\langle k^2\rangle$ from the three generating functions.
39. Prove that $\langle k^2\rangle \geq \langle k\rangle$ for all graphs, with equality if and only if $p(k) = 0$ for all $k > 1$.
40. Construct a large 2-regular graph, i.e. one with $p(k) = \delta_{k,2}$ and large N, that does not have a giant component. Prove your claim.

Tutorial 8

41. Show that the distribution $p(\boldsymbol{A})$ which maximises the Shannon entropy for the set of simple nondirected graphs, subject to the soft average degree constraint $\sum_{\boldsymbol{A}\in\mathcal{G}} p(\boldsymbol{A})\bar{k}(\boldsymbol{A}) = \langle k\rangle$ and subject to normalisation, is the Erdös-Rényi ensemble.
42. Show that the probabilities in the ensemble of simple nondirected graphs with soft-constrained degree sequences can, for large N and finite degrees $\{k_i\}$, be written as
$$p(\boldsymbol{A}|\boldsymbol{k}) = \prod_{i<j}\Big[\big(\frac{k_ik_j}{N\langle k\rangle} + \mathcal{O}(\frac{1}{N^2})\big)\delta_{A_{ij},1} + \big(1 - \frac{k_ik_j}{N\langle k\rangle} + \mathcal{O}(\frac{1}{N^2})\big)\delta_{A_{ij},0}\Big]$$
43. Calculate the degree distribution of small-world graphs built by superimposing a Poissonnian graph with average degree q on a one-dimensional periodic ring, for $N\to\infty$. You may assume that there are no common entries in the adjacency matrices of both.
44. Show that the distribution $p(k=0) = 0$ and $p(k>0) = 4/k(k+1)(k+2)$ solves the preferential attachment equation $\frac{1}{2}(k-1)p(k-1) - \frac{1}{2}(k+2)p(k) + \delta_{k1} = 0$.
45. Show that the previous distribution obeys $\sum_{k\geq 0} p(k) = 1$ and $\langle k\rangle = \sum_{k\geq 0} kp(k) = 2$. Hint: find constants a, b, c such that $1/k(k+1)(k+2) = a/k + b/(k+1) + c/(k+2)\ \forall k > 0$.
46. Calculate the leading orders in N of the number of *directed* N-node graphs with average degree $\bar{k}$, in the finite connectivity regime where $\bar{k}N\in\mathbb{N}$ and $\bar{k}\ll N$.