Network Science (VU) (706.703)

Mathematics of Networks

Denis Helic

ISDS, TU Graz

October 5, 2017

Outline

Introduction

- Mathematics of networks: graph theory
- Graph theory is a huge field with many results
- We focus on results that are important for study of real-world networks
- The slides and course structure is based on Networks: An Introduction by Mark Newman
- More on graph theory in e.g. Graph Theory by Harary or Introduction to Graph Theory by West

Networks

- A *network* is a collection of nodes connected by links
- Internet: nodes are computers and links are cables
- WWW: nodes are Web pages and links are hyperlinks
- Citation network: nodes are articles and links are citations
- Social networks: nodes are people and links are friendships
- Food web: nodes are species and links are predations

Networks

- The number of nodes in a network is denoted by n and the number od links by m
- In most cases there is at most a single link between two nodes
- In rare cases there might be multiple links (*multilinks*) between two nodes
- Links that connect a node to itself are called *self-links*
- A network that has neither multilinks nor self-links is called *simple network*
- A network with multilinks is called *multinetwork*

Simple networks

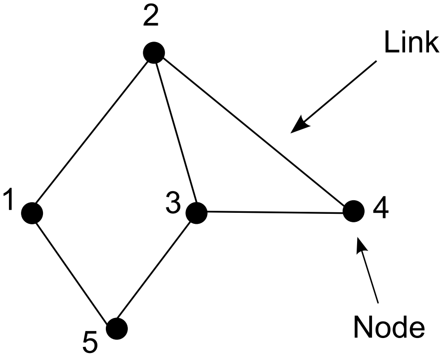


Figure: A simple graph

Multinetworks with self-links

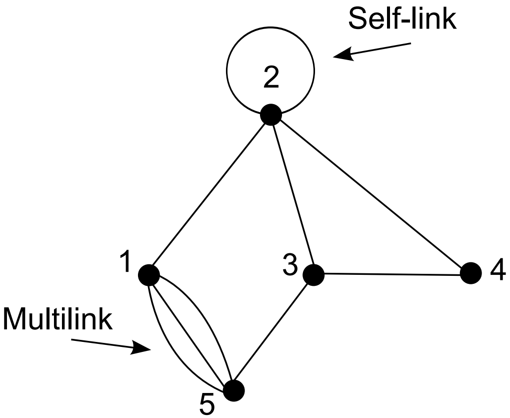
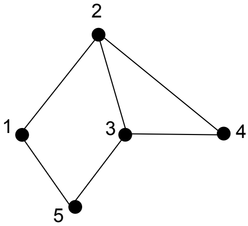


Figure: A simple graph with multilinks and self-links

Link lists

- There are number of ways to represent networks mathematically
- Consider a network with n nodes and let us label the nodes with integers $1 \ldots n$
- We denote a link between nodes i and j by (i, j)
- The complete network can be specified by n and list of links

The link list



$$(1,2),(1,5),(2,3),(2,4),(3,4),(3,5)$$

Link lists

- Link lists are typically used to store the network structure on computers
- SNAP library that we use in this course stores networks using link lists
- For mathematical purposes this representation is cumbersome
- We use the *adjacency matrix*

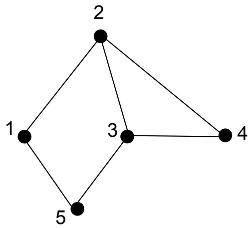
The adjacency matrix

Definition

The adjacency matrix $\mathbf{A}$ of a simple graph is the matrix with elements A_{ij} such that

$$A_{ij} = \begin{cases} 1 \text{ if there is a link between nodes } i \text{ and } j, \\ 0 \text{ otherwise.} \end{cases} \tag{1}$$

The adjacency matrix

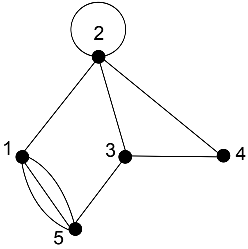


$$\mathbf{A} = \begin{pmatrix} 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \end{pmatrix} \tag{2}$$

The adjacency matrix

- For a network with no self-links the diagonal elements are all equal to zero
- The matrix is symmetric because if there is a link between i and j then there is also a link between j and i
- This holds for undirected links only
- We can use the adjacency matrix also for multinetworks and also for self-links
- E.g. for a triple link between i and j we set $A_{ij} = 3$
- For a self-link we set $A_{ii} = 2$ since each link has two ends

The adjacency matrix

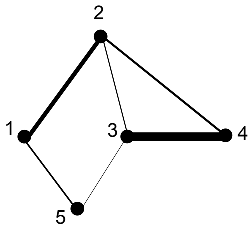


$$\mathbf{A} = \begin{pmatrix} 0 & 1 & 0 & 0 & 3 \\ 1 & 2 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 \\ 3 & 0 & 1 & 0 & 0 \end{pmatrix} \quad (3)$$

Weighted networks

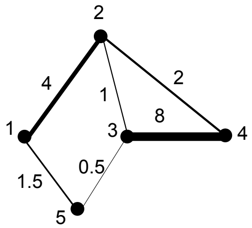
- Sometimes it is useful to represent links as having a strength or weight
- Internet: link weights might represent the data flow
- Social network: link value might represent the frequency of contact
- Information network: link value might represent the number of clicks on that link
- Weighted networks are also represented by the adjacency matrix

Weighted networks



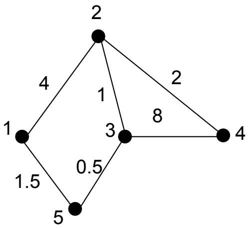
$$\mathbf{A} = \begin{pmatrix} 0 & 4 & 0 & 0 & 1.5 \\ 4 & 0 & 1 & 2 & 0 \\ 0 & 1 & 0 & 8 & 0.5 \\ 0 & 2 & 8 & 0 & 0 \\ 1.5 & 0 & 0.5 & 0 & 0 \end{pmatrix} \tag{4}$$

Weighted networks



$$\mathbf{A} = \begin{pmatrix} 0 & 4 & 0 & 0 & 1.5 \\ 4 & 0 & 1 & 2 & 0 \\ 0 & 1 & 0 & 8 & 0.5 \\ 0 & 2 & 8 & 0 & 0 \\ 1.5 & 0 & 0.5 & 0 & 0 \end{pmatrix} \tag{5}$$

Weighted networks



$$\mathbf{A} = \begin{pmatrix} 0 & 4 & 0 & 0 & 1.5 \\ 4 & 0 & 1 & 2 & 0 \\ 0 & 1 & 0 & 8 & 0.5 \\ 0 & 2 & 8 & 0 & 0 \\ 1.5 & 0 & 0.5 & 0 & 0 \end{pmatrix} \tag{6}$$

Directed networks

- In a *directed network* each link has a direction
- Each links points *from* one node *to* another
- Web: hyperlinks point from one page to another
- Citation networks: citations point from one article to another
- Directed networks are also represented by the adjacency matrix

Directed networks

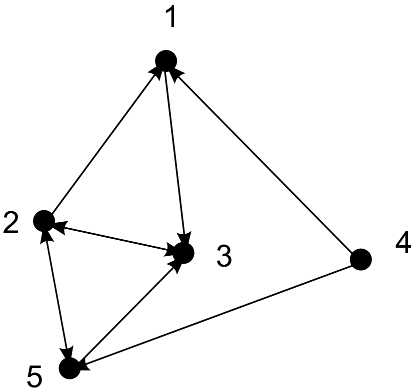


Figure: A directed network

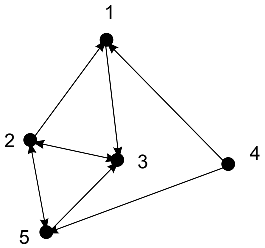
Directed networks

Definition

The adjacency matrix $\mathbf{A}$ of a directed networks is the matrix with elements A_{ij} such that

$$A_{ij} = \begin{cases} 1 \text{ if there is a link } \textit{from } j \textit{ to } i, \\ 0 \text{ otherwise.} \end{cases} \tag{7}$$

Directed networks



$$\mathbf{A} = \begin{pmatrix} 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \end{pmatrix} \tag{8}$$

Directed networks

- For the purpose of analysis it is sometimes useful to turn a directed network into a undirected one
- Some analytic techniques exist only for undirected networks
- One possibility is to ignore link directions completely
- We lose important information
- Better: cocitation and bibliographic coupling

Cocitation

- The *cocitation* of two nodes i and j in a directed network is the number of nodes that point to both i and j
- The number of papers that cite both i and j papers
- $A_{ik}A_{jk} = 1$ if i and j are both cited by k and zero otherwise

Cocitation

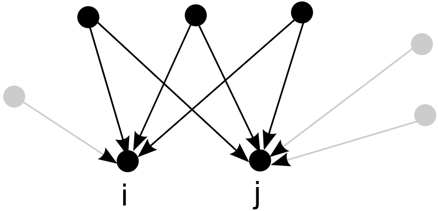


Figure: Cocitation: Nodes i and j are cited by three common papers, so their cocitation is 3.

Cocitation

Definition

The cocitation C_{ij} of i and j is

$$C_{ij} = \sum_{k=1}^{n} A_{ik}A_{jk} = \sum_{k=1}^{n} A_{ik}A_{kj}^{T} \tag{9}$$

$$\mathbf{C} = \mathbf{A}\mathbf{A}^{T} \tag{10}$$

Cocitation

- $\mathbf{C}$ is a $n \times n$ matrix
- It is symmetric since $\mathbf{C}^T = \left(\mathbf{A}\mathbf{A}^T\right)^T = \mathbf{A}\mathbf{A}^T = C$
- We define *cocitation network* in which there is a link if $C_{ij} > 0$ for $i \neq j$

Cocitation

- We can also make the cocitation network a weighted network with weights corresponding to C_{ij}
- Node pairs cited by more common papers have a stronger connection than those cited by fewer
- Higher cocitaiton is an indication that they deal with a similar topic
- The cocitation matrix is symmetric thus the cocitation network is undirected

Cocitation

The diagonal elements: total number of papers citing i

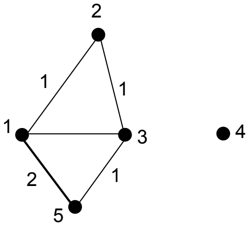
$$C_{ii} = \sum_{k=1}^{n} A_{ik}^2 = \sum_{k=1}^{n} A_{ik} \tag{11}$$

Cocitation

$$\mathbf{A} = \begin{pmatrix} 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \end{pmatrix} \qquad (12)$$

$$\mathbf{C} = \begin{pmatrix} 2 & 0 & 1 & 0 & 2 \\ 0 & 2 & 1 & 0 & 0 \\ 1 & 1 & 3 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 2 & 0 & 1 & 0 & 2 \end{pmatrix} \qquad (13)$$

Cocitation



$$\mathbf{C} = \begin{pmatrix} 0 & 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 2 & 0 & 1 & 0 & 0 \end{pmatrix} \tag{14}$$

Bibiliographic coupling

- The *bibliographic coupling* of two nodes i and j in a directed network is the number of other nodes to which both i and j point
- The number of other papers that are cited by both i and j
- $A_{ki}A_{kj} = 1$ if i and j both cite k and zero otherwise

Bibliographic coupling

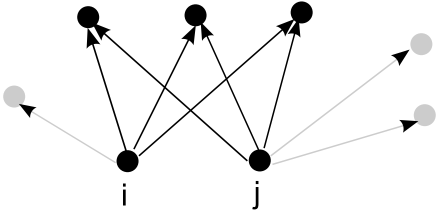


Figure: Bibliographic coupling: Nodes i and j cite three of the same papers, so their bibliographic coupling is 3.

Bibliographic coupling

Definition

The bibliographic coupling B_{ij} of i and j is

$$B_{ij} = \sum_{k=1}^{n} A_{ki} A_{kj} = \sum_{k=1}^{n} A_{ik}^{T} A_{kj} \tag{15}$$

$$\mathbf{B} = \mathbf{A}^{T}\mathbf{A} \tag{16}$$

Bibliographic coupling

- $\mathbf{B}$ is a $n \times n$ matrix
- It is symmetric since $\mathbf{B}^T = \left(\mathbf{A}^T\mathbf{A}\right)^T = \mathbf{A}^T\mathbf{A} = B$
- We define *bibliographic coupling network* in which there is a link if $B_{ij} > 0$ for $i \neq j$

Bibliographic coupling

- Again, we can make the bibliographic coupling network a weighted network with weights corresponding to B_{ij}
- Node pairs that cite both more common papers have a stronger connection than those citing fewer common papers
- Higher bibliographic coupling is an indication that they deal with a similar subject matter
- The bibliographic coupling matrix is symmetric thus the bibliographic coupling network is undirected

Bibliographic coupling

The diagonal elements: the number of papers i cites

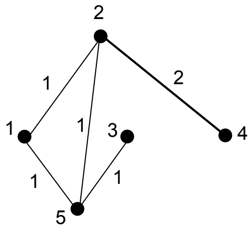
$$B_{ii} = \sum_{k=1}^{n} A_{ki}^2 = \sum_{k=1}^{n} A_{ki} \tag{17}$$

Bibliographic coupling

$$\mathbf{A} = \begin{pmatrix} 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \end{pmatrix} \quad (18)$$

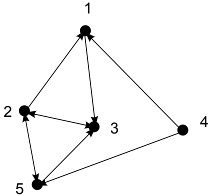
$$\mathbf{B} = \begin{pmatrix} 1 & 1 & 0 & 0 & 1 \\ 1 & 3 & 0 & 2 & 1 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 2 & 0 & 2 & 0 \\ 1 & 1 & 1 & 0 & 2 \end{pmatrix} \quad (19)$$

Bibliographic coupling

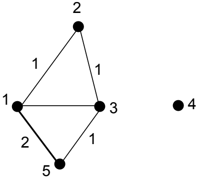


$$\mathbf{B} = \begin{pmatrix} 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 2 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 2 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \end{pmatrix} \tag{20}$$

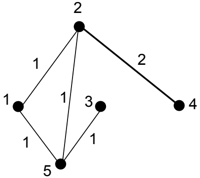
Cocitation/Bibliographic coupling



(a) A directed network



(b) Cocitation network



(c) Bibl. coupling network

Cocitation vs. bibliographic coupling

- Mathematically similar measures but give different results
- Strong cocitation: both nodes are pointed to by **many** of the same nodes
- Both nodes have to have **a lot** of incoming links in the first place
- Both papers have to be well cited: influential papers such as surveys, review articles, and so on

Cocitation vs. bibliographic coupling

- Strong bibliographic coupling: both papers cite **many** other papers
- They have large bibliographies
- The sizes of bibliographies vary less than the number of citations
- Bibliographic coupling is a more uniform indicator of paper similarity

Cocitation vs. bibliographic coupling

- Bibliographic coupling can be computed as soon as the paper is published
- Citation can be computed only after the paper has been cited
- Cocitation changes over the time
- That is the reason why bibliographic coupling is typically used as a similarity metric for papers in digital libraries
- This discussion points out the differences between incoming and outgoing links in a directed network (cf. PageRank, HITS, ...)

Bipartite networks

- Another way to represent group memberships is by means of a *bipartite network*
- *Two-mode networks* in sociology
- In such networks we have two types of nodes
- One type represents the original nodes
- The other type represents the groups to which the original nodes belong (actors-movies, authors-papers, ...)
- The links can connect only nodes of different types

Bipartite networks

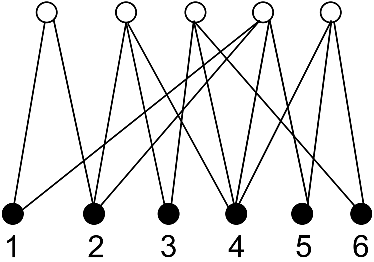


Figure: A bipartite network

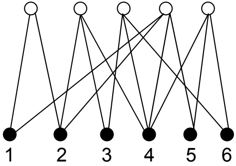
The incidence matrix

Definition

If n is the number of nodes and g is the number of groups, then the incidence matrix $\mathbf{B}$ is a $g \times n$ matrix with elements B_{ij} such that

$$B_{ij} = \begin{cases} 1 & \text{if node } j \text{ belongs to group } i, \\ 0 & \text{otherwise.} \end{cases} \tag{21}$$

The incidence matrix



$$\mathbf{B} = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 \\ 1 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 \end{pmatrix} \tag{22}$$

One-mode projections

- Sometimes we want to work with direct connections between nodes of the same type
- We infer such connections from the bipartite network by creating a *one-mode projection*
- E.g. for the actor-movie network we create a one-mode projection onto actors
- Two actors are connected if they appeared in a movie together
- In the projection on the movies, two movies are connected if they share a common actor

One-mode projections

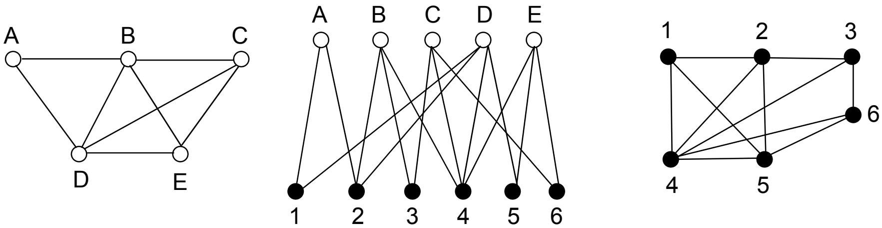


Figure: One-mode projections of a bipartite network

One-mode projections

- One-mode projections constructed in this way are useful but a lot of information is lost
- E.g. if actors are connected that means that they acted together in a movie but we do not know in how many movies
- We can capture this information by making the one-mode projections weighted
- Mathematically, we can write the projection in the terms of the incidence matrix
- $B_{ki}B_{kj} = 1$ iff i and j belong to the same group k

Projection on nodes

Definition

The total number P_{ij} of groups to which both i and j belong is

$$P_{ij} = \sum_{k=1}^{g} B_{ki} B_{kj} = \sum_{k=1}^{g} B_{ik}^{T} B_{kj} \tag{23}$$

$$\mathbf{P} = \mathbf{B}^T \mathbf{B} \tag{24}$$

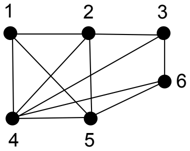
Projection on nodes

The diagonal elements: the number of groups to which i belongs

$$P_{ii} = \sum_{k=1}^{g} B_{ki}^2 = \sum_{k=1}^{g} B_{ki} \tag{25}$$

- $\mathbf{P}$ is similar to the bibliographic coupling matrix. We can turn it into the adjacency matrix of a weighted network by setting the diagonal elements to zero

Projection on nodes



$$\mathbf{P} = \begin{pmatrix} 2 & 2 & 0 & 1 & 1 & 0 \\ 2 & 3 & 1 & 2 & 1 & 0 \\ 0 & 1 & 2 & 2 & 0 & 1 \\ 1 & 2 & 2 & 4 & 2 & 2 \\ 1 & 1 & 0 & 2 & 2 & 1 \\ 0 & 0 & 1 & 2 & 1 & 2 \end{pmatrix} \tag{26}$$

Projection on groups

Definition

The number P'_{ij} of common members of groups i and j is

$$P'_{ij} = \sum_{k=1}^{n} B_{ik} B_{jk} = \sum_{k=1}^{n} B_{ik} B^T_{kj} \tag{27}$$

$$\mathbf{P}' = \mathbf{B}\mathbf{B}^T \tag{28}$$

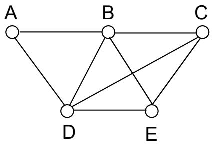
Projection on groups

The diagonal elements: the number of members in group i

$$P'_{ii} = \sum_{k=1}^{n} B_{ik}^2 = \sum_{k=1}^{n} B_{ik} \tag{29}$$

- $\mathbf{P}$ is similar to the cocitation matrix. We can turn it into the adjacency matrix of a weighted network by setting the diagonal elements to zero

Projection on groups



$$\mathbf{P}' = \begin{pmatrix} 2 & 1 & 0 & 2 & 0 \\ 1 & 3 & 2 & 2 & 1 \\ 0 & 2 & 3 & 1 & 2 \\ 2 & 2 & 1 & 4 & 2 \\ 0 & 1 & 2 & 2 & 3 \end{pmatrix} \tag{30}$$

Degree

- The *degree* of a node is the number of links connected to it
- We denote the degree of node i by k_i

The degree in terms of the adjacency matrix (undirected networks)

$$k_i = \sum_{j=1}^{n} A_{ij} \tag{31}$$

Degree

- Every link has two ends, hence there are $2m$ link ends in an undirected network
- The number of link ends is equal to the sum of the degrees of all the nodes

The degrees and the number of links

$$2m = \sum_{i=1}^{n} k_i \tag{32}$$

$$m = \frac{1}{2}\sum_{i=1}^{n} k_i = \frac{1}{2}\sum_{ij} A_{ij} \tag{33}$$

Mean degree

The mean degree c in an undirected graph

$$c = \frac{1}{n}\sum_{i=1}^{n} k_i \tag{34}$$

$$c = \frac{2m}{n} \tag{35}$$

Network density

- The maximum number of links in a simple network is equal to the number of possible combinations of node pairs: $\binom{n}{2} = \frac{1}{2}n(n-1)$

Density is the fraction of links that actually exist

$$\rho = \frac{m}{\binom{n}{2}} = \frac{2m}{n(n-1)} = \frac{c}{n-1} \tag{36}$$

Network density

- The density lies in the range $0 \le \rho \le 1$
- What is the behavior of ρ as $n \to \infty$
- If ρ tends to a constant as $n \to \infty$ the network is said to be *dense*. The fraction of non-zero elements in the adjacency matrix remains constant as the network gets larger.

Network density

- If $\rho \to 0$ as $n \to \infty$ the network is said to be *sparse*. The fraction of non-zero elements in the adjacency matrix also tends to zero.
- In particular, a network is *sparse* if the mean degree c tends to constant as n becomes larger.
- Almost all empirical networks we are interested in are sparse: the Web, Wikipedia, social networks, ...
- This has some important consequences when we design network algorithms

Degree

- In directed networks we have *in-degree* and *out-degree*
- In-degree is the number of ingoing links and out-degree is the number of outgoing links

The degree in directed networks

$$k_i^{in} = \sum_{j=1}^{n} A_{ij} \tag{37}$$

$$k_j^{out} = \sum_{i=1}^{n} A_{ij} \tag{38}$$

Mean degree

The mean degree c in a directed graph

$$m = \sum_{i=1}^{n} k_i^{in} = \sum_{j=1}^{n} k_j^{out} = \sum_{ij} A_{ij} \tag{39}$$

$$c_{in} = \frac{1}{n} \sum_{i=1}^{n} {k_i}^{in} = \frac{1}{n} \sum_{j=1}^{n} {k_j}^{out} = c_{out} \tag{40}$$

$$c = \frac{m}{n} \tag{41}$$

Paths

- A *path* in a network is a sequence of nodes such that that each consecutive pair of nodes is connected by a link
- A path is a route between two nodes across a network
- In directed networks each link is traversed in the link direction
- A path can intersect itself, e.g. a node can be visited more than once, or a link can be traversed more than once
- If the path does not intersect itself it is called a *self-avoiding path*
- The *length* of a path is the number of links traversed along that path

Paths

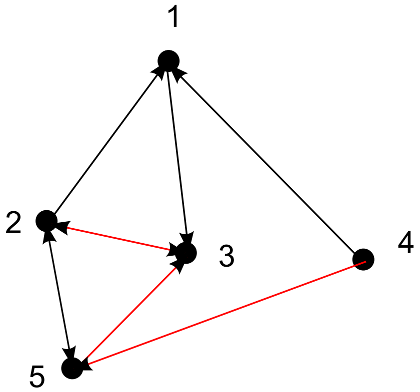


Figure: A path of length three in a network

Number of paths

- A_{ij} is 1 if there is a link from j to i, and 0 otherwise
- $A_{ik}A_{kj}$ is 1 if there is a path of length 2 from j to i via k

The total number $N_{ij}^{(2)}$ of paths of length 2 from j to i

$$N_{ij}^{(2)} = \sum_{k=1}^{n} A_{ik}A_{kj} = \left[\mathbf{A}^2\right]_{ij} \tag{42}$$

- $[...]_{ij}$ denotes the ijth element of the matrix

Number of paths

- $A_{ik}A_{kl}A_{lj}$ is 1 if there is a path of length 3 from j to i via l and k

The total number $N_{ij}^{(3)}$ of paths of length 3 from j to i

$$N_{ij}^{(3)} = \sum_{k,l=1}^{n} A_{ik}A_{kl}A_{lj} = \left[\mathbf{A}^3\right]_{ij} \tag{43}$$

Number of paths

- WE can generalize to the paths of arbitrary length r

The total number $N_{ij}^{(r)}$ of paths of length r from j to i

$$N_{ij}^{(r)} = [\mathbf{A}^r]_{ij} \tag{44}$$

Number of cycles

- Paths that start and end at i are cycles in a network
- The number of cycles of length r is $[\mathbf{A}^r]_{ii}$

The total number L_r of cycles of length r in a network

$$L_r = \sum_{i=1}^{n} [\mathbf{A}^r]_{ii} = \mathrm{Tr}\mathbf{A}^r \tag{45}$$

- Tr is a trace of a matrix, i.e. the sum of elements on the main diagonal

Number of cycles

- We can express the last equation in terms of the eigenvalues of the adjacency matrix
- For undirected graphs the adjacency matrix is symmetric
- The adjacency matrix has n real eigenvalues
- The eigenvectors have real elements
- The adjacency matrix can be written in form $\mathbf{A} = \mathbf{U}\mathbf{K}\mathbf{U}^T$
- $\mathbf{U}$ is the orthogonal matrix of eigenvectors and $\mathbf{K}$ is the diagonal matrix of eigenvalues

Number of cycles

- Then $\mathbf{A}^r = \left(\mathbf{U}\mathbf{K}\mathbf{U}^T\right)^r = \mathbf{U}\mathbf{K}^r\mathbf{U}^T$
- Since $\mathbf{U}\mathbf{U}^T = \mathbf{I}$ because $\mathbf{U}^T = \mathbf{U}^{-1}$

The total number L_r of cycles of length r in a network

$$L_r = \mathrm{Tr}\left(\mathbf{U}\mathbf{K}^r\mathbf{U}^T\right) = \mathrm{Tr}\left(\mathbf{U}\mathbf{U}^T\mathbf{K}^r\right) = \mathrm{Tr}\mathbf{K}^r = \sum_i \kappa_i^r \tag{46}$$

Number of cycles

- The last follows since trace of a matrix is invariant under cyclic permutations
- κ_i is the ith eigenvalue of the adjacency matrix
- Same equation holds for directed networks, although the proof is a bit more complicated
- Although some eigenvalues might be complex they always come in complex-conjugate pairs: $det(\kappa\mathbf{I} - \mathbf{A})$
- Each term is complemented by another that is its complex conjugate and thus the sum is always real

Geodesic paths

- A *geodesic path* or a *shortest path* is a path between two nodes such that no shorter path exists
- It is possible that there is no shortest path between two nodes if they are not connected
- By convention we say that the distance between those two nodes is infinite

Geodesic paths

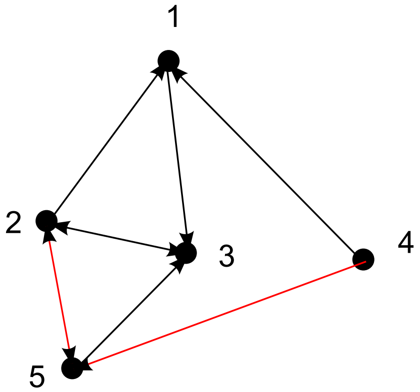


Figure: A geodesic (shortest) path of length two between two nodes

Geodesic paths

- Geodesic paths are self-avoiding paths
- There may be more than one geodesic path in a network
- Teh *diameter* of a network is a length of the longest shortest path in that network

Components

- Sometimes there is no path between two nodes
- A network might be divided into two or more node subgroups with no connection between the groups
- If there exist a node pair with no path between them the network is *disconnected*
- If there is a path from every node to every other node then the network is *connected*
- The subgroups in a network are called *components*
- A single node with no links is also a component of size 1 and a connected network has a single component

Components

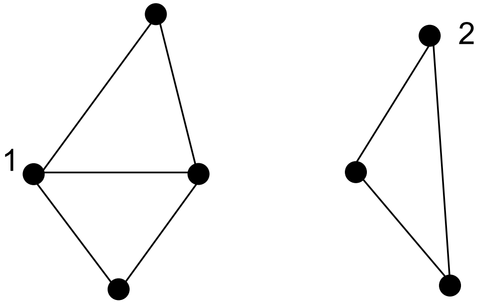


Figure: A network with two components

Components

- With a proper labeling we can write the adjacency matrix in the following form

$$\mathbf{A} = \begin{pmatrix} \boxed{} & 0 & \cdots \\ 0 & \boxed{} & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix} \tag{47}$$

Components in directed networks

- Now we take into account the direction of links
- E.g. each hyperlink on the Web is a directed link
- If we ignore directions we have the undirected case and speak about *weakly connected components*
- Sometimes, we have a directed path from A to B, but no such path from B to A

Components in directed networks

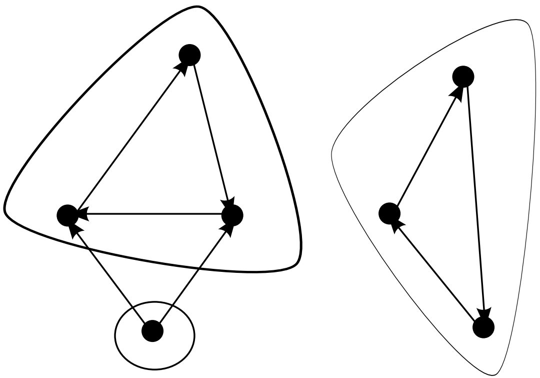


Figure: Components in a directed network

Components in directed networks

- If both paths exist then A and B are *strongly connected*
- Subsets of nodes that are strongly connected are called *strongly connected components*
- A single node with constitutes a strongly connected component of size 1
- Every node in a strongly connected component must belong to at least one cycle
- Every strongly connected component in a directed acyclic networks has only a single node

Components in directed networks

- Sometimes we are interested in other kinds of components (e.g. which Web pages can I reach from a given Web page)
- *Out-component* is the set of nodes that reachable via directed paths from a specified node A, and including A itself
- Links from external nodes (such that are not in an out-component) only point inward towards the members of the component

Components in directed networks

- Out-component is a property of the network structure and a starting node
- Out-components of all members of a strongly connected component are identical (since all members of a strongly connected component are mutually reachable)
- Thus, out-components belong to strongly connected components

Components in directed networks

- Similarly, *in-component* is the set of nodes (including A) from which via directed paths a specified node A can be reached
- Links to external nodes (such that are not in an in-component) only point outward from the members of the component
- In-component is a property of the network structure and a starting node

Components in directed networks

- In-components of all members of a strongly connected component are identical (since all members of a strongly connected component are mutually reachable)
- Therefore, in-components belong to strongly connected components
- A strongly connected component is the intersection of its in- and out-components

Components in directed networks

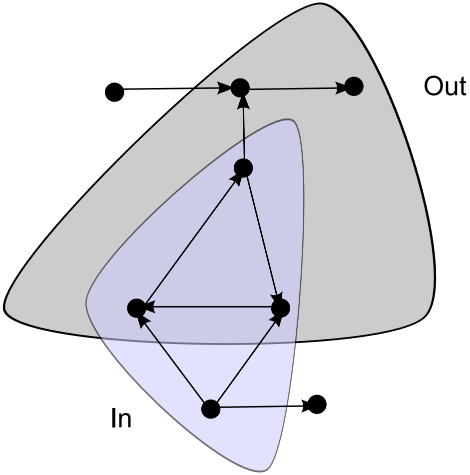


Figure: In- and out-components in a directed network

The graph Laplacian

- The adjacency matrix captures the whole structure of a network
- There is another matrix, closely related to the adjacency matrix
- However, it differs in some important aspects which can provide some additional information about the network structure
- This is the graph Laplacian

The graph Laplacian

Definition

The degree matrix $\mathbf{D}$ of a simple undirected graph is the diagonal matrix with the node degrees along its diagonal:

$$\mathbf{D} = \begin{pmatrix} k_1 & 0 & 0 & \dots \\ 0 & k_2 & 0 & \dots \\ 0 & 0 & k_3 & \dots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix} \tag{48}$$

The graph Laplacian

Definition

The graph Laplacian $\mathbf{L}$ of a simple undirected graph is defined as:

$$\mathbf{L} = \mathbf{D} - \mathbf{A} \tag{49}$$

The graph Laplacian

Definition

The graph Laplacian $\mathbf{L}$ of a simple undirected graph is the matrix with elements L_{ij} such that

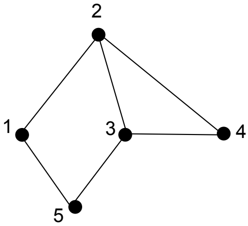
$$L_{ij} = \begin{cases} k_i \text{ if } i = j \\ -1 \text{ if there is a link between nodes } i \text{ and } j \text{ and } i \neq j \\ 0 \text{ otherwise.} \end{cases} \tag{50}$$

The graph Laplacian

- Alternatively, we can write
- δ_{ij} is the Kronecker delta, which is 1 for $i = j$ and 0 otherwise

$$L_{ij} = \delta_{ij} k_i - A_{ij} \tag{51}$$

The graph Laplacian



$$\mathbf{A} = \begin{pmatrix} 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \end{pmatrix} \tag{52}$$

The graph Laplacian

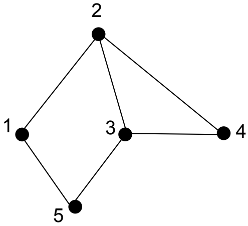


Figure: $D = diag(sum(A))$

$$\mathbf{D} = \begin{pmatrix} 2 & 0 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 & 0 \\ 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 2 \end{pmatrix} \tag{53}$$

The graph Laplacian

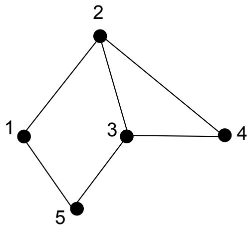


Figure: $L = D - A$

$$\mathbf{L} = \begin{pmatrix} 2 & -1 & 0 & 0 & -1 \\ -1 & 3 & -1 & -1 & 0 \\ 0 & -1 & 3 & -1 & -1 \\ 0 & -1 & -1 & 2 & 0 \\ -1 & 0 & -1 & 0 & 2 \end{pmatrix} \tag{54}$$

Eigenvalues of the graph Laplacian

- The eigenvalues of the graph Laplacian are its most interesting property
- The Laplacian is a symmetric matrix → it has real eigenvalues
- We can even show that all of its eigenvalues are non-negative
- Also, we can show that its smallest eigenvalue $\lambda_1 = 0$

The link incidence matrix

Definition

The link incidence matrix $\mathbf{B}$ of a simple undirected graph with n nodes and m links is an $m \times n$ matrix with elements B_{ij} such that

$$B_{ij} = \begin{cases} 1 \text{ if end 1 of link } i \text{ is attached to node } j \\ -1 \text{ if end 2 of link } i \text{ is attached to node } j \\ 0 \text{ otherwise.} \end{cases} \tag{55}$$

- We designate for each link one end as end 1 and other as end 2
- Each row of the link incidence matrix has exactly one 1 and one -1

The link incidence matrix

- What is the value of $B_{ki}B_{kj}$ for $i \neq j$

The link incidence matrix

- What is the value of $B_{ki}B_{kj}$ for $i \neq j$
- If link k connects i and j then the product has value -1, otherwise it is 0
- What is the value of $\sum_k B_{ki}B_{kj}$ for $i \neq j$

The link incidence matrix

- What is the value of $B_{ki}B_{kj}$ for $i \neq j$
- If link k connects i and j then the product has value -1, otherwise it is 0
- What is the value of $\sum_k B_{ki}B_{kj}$ for $i \neq j$
- In a simple graph there is at most one link connecting i and j
- If there is a link between i and j the sum is -1, otherwise it is 0

The link incidence matrix

- What is the value of B_{ki}^2 for $i = j$

The link incidence matrix

- What is the value of B_{ki}^2 for $i = j$
- If link k connects to i the product has value 1, otherwise it is 0
- What is the value of $\sum_k B_{ki}^2$ for $i = j$

The link incidence matrix

- What is the value of B_{ki}^2 for $i = j$
- If link k connects to i the product has value 1, otherwise it is 0
- What is the value of $\sum_k B_{ki}^2$ for $i = j$
- It is equal to the degree k_i of node i

The link incidence matrix

- Thus, $\sum_k B_{ki} B_{kj} = L_{ij}$
- The diagonal elements L_{ii} are equal to the degrees k_i
- The off-diagonal elements are -1 if there is a link between i and j

$$\mathbf{L} = \mathbf{B}^T \mathbf{B} \tag{56}$$

Eigenvalues of the graph Laplacian

- Let $\mathbf{v}_i$ be an eigenvector of $\mathbf{L}$ with eigenvalue λ_i, then $\mathbf{L}\mathbf{v}_i = \lambda_i \mathbf{v}_i$

$$\mathbf{v}_i^T \mathbf{B}^T \mathbf{B} \mathbf{v}_i = \mathbf{v}_i^T \mathbf{L} \mathbf{v}_i = \lambda_i \mathbf{v}_i^T \mathbf{v}_i = \lambda_i \tag{57}$$

- We assume that $\mathbf{v}_i$ is normalized, so that its scalar product with itself is 1

Eigenvalues of the graph Laplacian

- Any eigenvalue λ_i is equal to the scalar product of $(\mathbf{Bv}_i)$ with itself
- $(\mathbf{v}_i^T\mathbf{B})(\mathbf{Bv}_i)$
- $(\mathbf{Bv}_i)$ is a vector with real elements
- The product is the sum of the squares of real elements
- $\lambda_i \geq 0$, for all i
- In fact, the Laplacian always has at least one zero eigenvalue

Eigenvalues of the graph Laplacian

$$\mathbf{L}\begin{pmatrix}1\\ \vdots\\ 1\\ \vdots\end{pmatrix} = \begin{pmatrix}\sum_j(\delta_{1j}k_{1j} - A_{1j})\\ \vdots\\ \sum_j(\delta_{ij}k_{ij} - A_{ij})\\ \vdots\end{pmatrix} = \begin{pmatrix}k_1 - \sum_j A_{1j}\\ \vdots\\ k_i - \sum_j A_{ij}\\ \vdots\end{pmatrix} = \begin{pmatrix}k_1 - k_1\\ \vdots\\ k_i - k_i\\ \vdots\end{pmatrix} = \begin{pmatrix}0\\ \vdots\\ 0\\ \vdots\end{pmatrix} = 0\begin{pmatrix}1\\ \vdots\\ 1\\ \vdots\end{pmatrix} \tag{58}$$

Eigenvalues of the graph Laplacian

- The vector $\mathbf{1}$ is always an eigenvector of $\mathbf{L}$ with eigenvalue 0
- There are no negative eigenvalues, thus this is the lowest eigenvalue
- Convention: $\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n$
- We always have $\lambda_1 = 0$

Componenents and the algebraic connectivity

- Suppose we have a network with c different components
- The components have sizes $n_1, n_2, \dots, n_c$

$$\mathbf{L} = \begin{pmatrix} \boxed{} & 0 & \dots \\ 0 & \boxed{} & \dots \\ \vdots & \vdots & \ddots \end{pmatrix} \tag{59}$$

Componenents and the algebraic connectivity

$$\mathbf{v} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 0 \\ 0 \\ \vdots \end{pmatrix} \tag{60}$$

- We have n_1 ones and this is an eigenvector with eigenvalue 0
- We have c such eigenvectors

Componenents and the algebraic connectivity

- In a network with c components c eigenvalues are equal to 0
- The second eigenvalue λ_2 of the graph Laplacian is non-zero iff the network is connected
- The second eigenvalue of the Laplacian is called *algebraic connectivity*
- It is a measure of how connected is a network, i.e. how difficult is to divide that network