Complex Networks mathematics of networks

2011.10.17

contents (1)

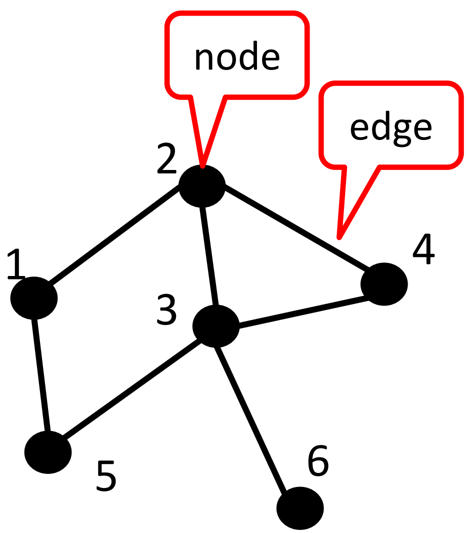
- networks and their representation
- the adjacency matrix
- weighted networks
- directed networks
- hypergraphs
- bipartite networks
- trees

contents (2)

- planar networks
- degree
- paths
- components
- independent paths, connectivity, and cut sets
- the graph laplacian
- random walks

networks and their representation

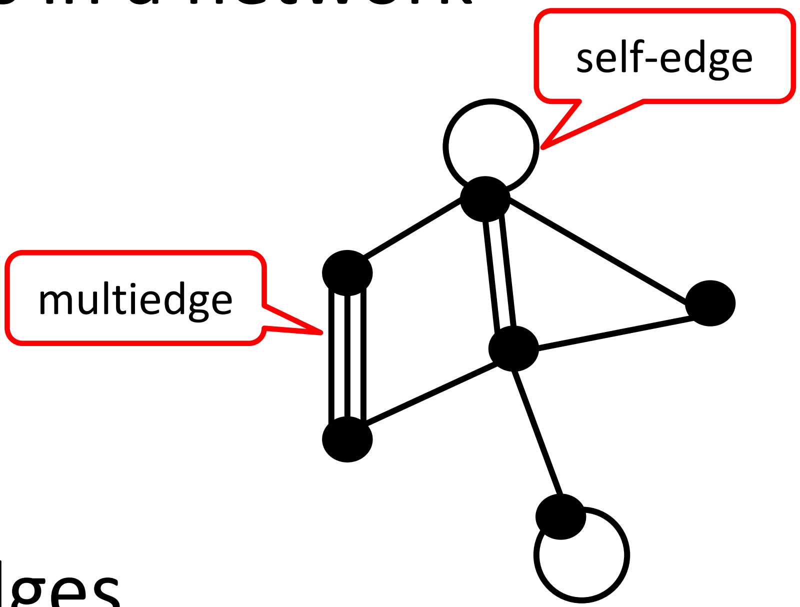
- a network (a graph) is a collection of vertices (nodes) joined by edges (links).



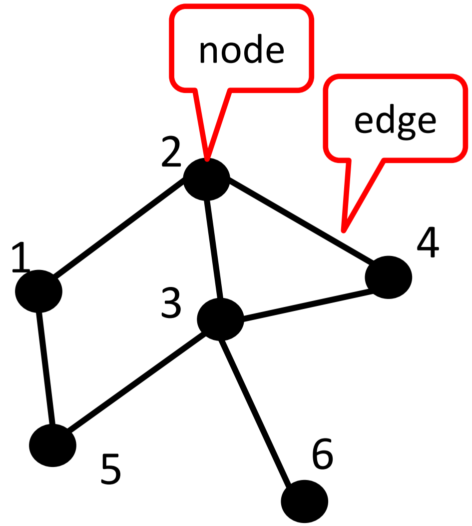
Network	Vertex	Edge
Internet	Computer or router	Cable or wireless data connection
World Wide Web	Web page	Hyperlink
Citation network	Article, patent, or legal case	Citation
Power grid	Generating station or substation	Transmission line
Friendship network	Person	Friendship
Metabolic network	Metabolite	Metabolic reaction
Neural network	Neuron	Synapse
Food web	Species	Predation

notations

- n: the number of vertices in a network
- m: the number of edges
- multiedge, self-edge
- multigraph: with multiedges



edge list & adjacency matrix



edge list

n=6

(1,2),(1,5),(2,3),(2,4),(3,4),(3,5),(3,6)

adjacency matrix

$$A_{ij} = \begin{cases} 1 & \text{if there is an edge between vertices i and j,} \\ 0 & \text{otherwise.} \end{cases}$$

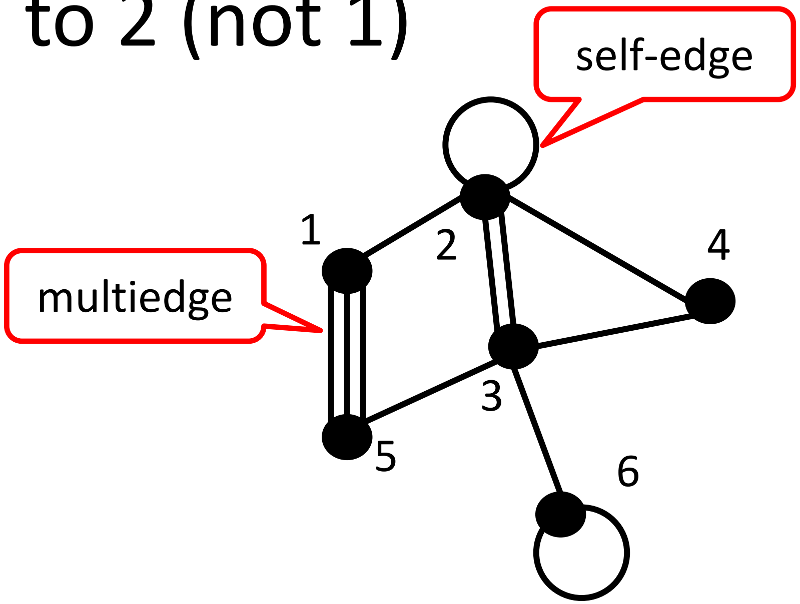
$$A = \begin{pmatrix} 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{pmatrix}$$

adjacency matrix

- no self-edge -> diagonal elements are all zero
- symmetric (for undirected networks)

- multiedge: setting A_{ij} equal to the multiplicity
- self-edge: setting A_{ij} equal to 2 (not 1)

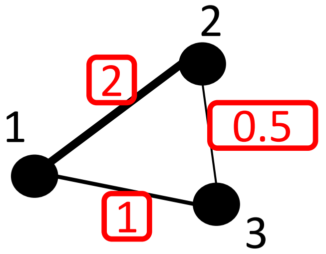
$$A = \begin{pmatrix} 0 & 1 & 0 & 0 & 3 & 0 \\ 1 & 2 & 2 & 1 & 0 & 0 \\ 0 & 2 & 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 0 \\ 3 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 2 \end{pmatrix}$$



weighted networks

- weights represent
  - the amount of data flowing/bandwidth (Internet)
  - total energy flow (food web)
  - frequency of contact (social network)

$$A = \begin{pmatrix} 0 & 2 & 1 \\ 2 & 0 & 0.5 \\ 1 & 0.5 & 0 \end{pmatrix}$$

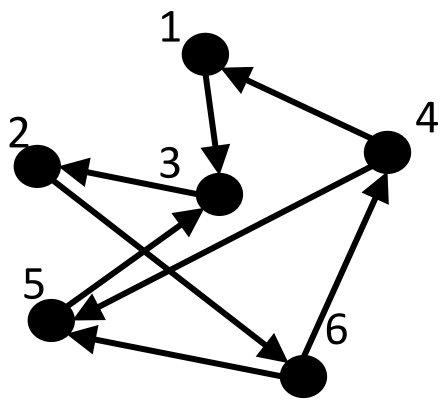


- weighted edge vs multiedge
  - switching between the two can be useful for analysis
- weights can be negative
  - animosity (social network)

directed network (digraph)

- each edge has a direction
  - hyperlink from one page to another (WWW)
- adjacency matrix is asymmetric

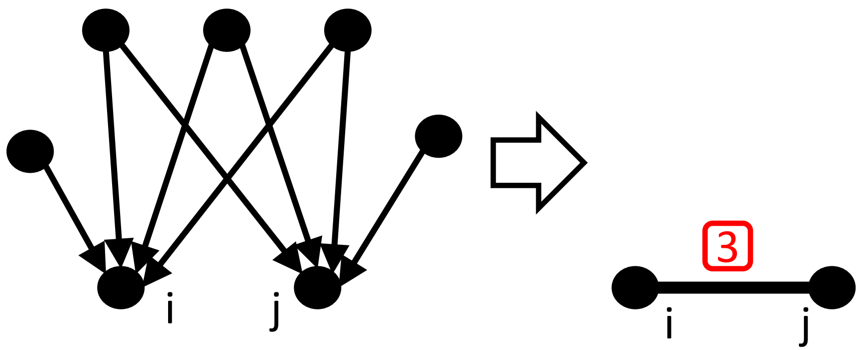
$$A_{ij} = \begin{cases} 1 & \text{if there is an edge from } j \text{ to } i, \\ 0 & \text{otherwise.} \end{cases}$$



$$A = \begin{array}{cc} & \begin{array}{cccccc} & & & j & & \end{array} \\ \begin{array}{c} \\ \\ i \\ \\ \\ \\ \end{array} & \begin{pmatrix} 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 \end{pmatrix} \end{array}$$

cocitation and bibliograhic coupling

- a directed network -> an undirected one
  - just ignoring the edge directions is easy, but it may lose valuable information
  - cocitation: # of vertices that have outgoing edges pointing to both i and j

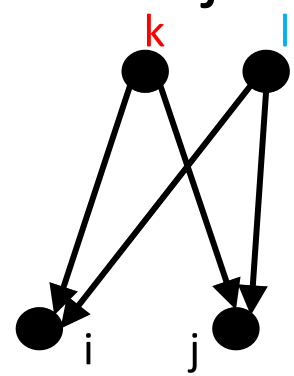


papers i & j are often co-cited
-> they are closely related

Adjacency matrix of cocitation

- C (cocitation matrix)
  - Cij : # of columns whose ith & jth elements are 1

$$A = \begin{array}{c} \\ \\ i \\ j \\ \\ \end{array} \begin{array}{c} \begin{array}{cccc} & k & l & \end{array} \\ \begin{pmatrix} .. & .. & .. & .. \\ .. & 1 & .. & .. \\ .. & 1 & .. & .. \\ .. & .. & .. & .. \end{pmatrix} \end{array}$$



$$C_{ij} = \sum_{k=1}^{n} A_{ik} A_{jk} = \sum_{k=1}^{n} A_{ik} A_{kj}^{T} \qquad i \neq j$$

A^T : transpose of A

$$C = \begin{array}{c} \\ \\ i \\ j \\ \\ \end{array} \begin{array}{c} \begin{array}{cccc} & k & l & \end{array} \\ \begin{pmatrix} .. & .. & .. & .. \\ .. & 1 & .. & .. \\ .. & 1 & .. & .. \\ .. & .. & .. & .. \end{pmatrix} \\ \boxed{A} \end{array} \begin{array}{c} \begin{array}{cccc} & i & j & \end{array} \\ \begin{pmatrix} .. & .. & .. & .. \\ .. & 1 & 1 & .. \\ .. & .. & .. & .. \\ .. & .. & .. & .. \end{pmatrix} \\ \boxed{A^T} \end{array} \begin{array}{c} \\ \\ k \\ l \\ \\ \end{array}$$

$$C = AA^T$$

C is symmetric because

$$C^T = (AA^T)^T = AA^T = C$$

more on citation matrix

- if all elements in A are zero or one,

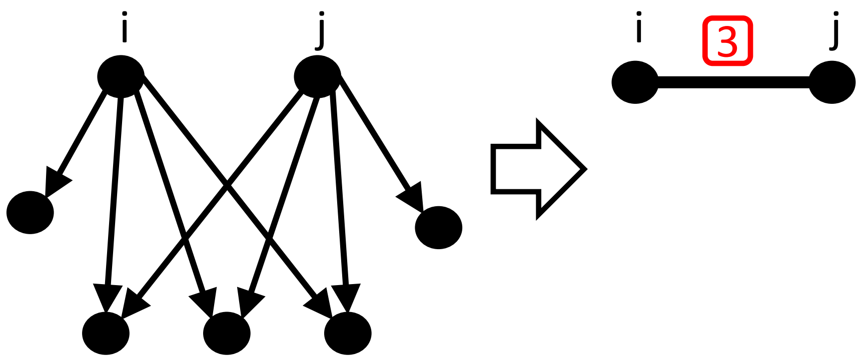
$$C_{ii} = \sum_{k=1}^{n} A_{ik}^2 = \sum_{k=1}^{n} A_{ik}$$ -> # of 1s in ith row

- we ignore these diagonal elements

$$C_{ij} = \begin{cases} \sum_{k=1}^{n} A_{ik} A_{kj}^T & i \neq j \\ 0 & i = j \end{cases}$$

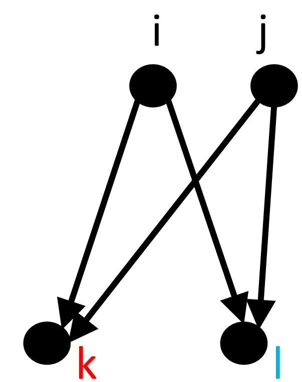
bibliographic coupling

- # of other vertices to which both point



i & j often cite the same papers
-> they are closely related

$$A = \begin{array}{c} \\ \\ k \\ l \\ \\ \end{array} \begin{array}{c} \begin{array}{cccc} & i & j & \end{array} \\ \begin{pmatrix} .. & .. & .. & .. \\ .. & 1 & 1 & .. \\ .. & .. & .. & .. \\ .. & .. & .. & .. \end{pmatrix} \end{array}$$



- B (bibliographic coupling)

$$B_{ij} = \sum_{k=1}^{n} A_{ki} A_{kj} = \sum_{k=1}^{n} A^T_{ik} A_{kj} \quad i \neq j$$

$$B = \begin{array}{c} \\ \\ i \\ j \\ \\ \end{array} \underbrace{\begin{array}{c} \begin{array}{cccc} & k & l & \end{array} \\ \begin{pmatrix} .. & .. & .. & .. \\ .. & 1 & .. & .. \\ .. & 1 & .. & .. \\ .. & .. & .. & .. \end{pmatrix} \end{array}}_{A^T} \underbrace{\begin{array}{c} \begin{array}{cccc} & i & j & \end{array} \\ \begin{pmatrix} .. & .. & .. & .. \\ .. & 1 & 1 & .. \\ .. & .. & .. & .. \\ .. & .. & .. & .. \end{pmatrix} \end{array}}_{A} \begin{array}{c} \\ \\ k \\ l \\ \\ \end{array}$$

$$B = A^T A$$

B is symmetric

$$B_{ij} = \begin{cases} \displaystyle\sum_{k=1}^{n} A^T_{ik} A_{kj} & i \neq j \\ 0 & i = j \end{cases}$$

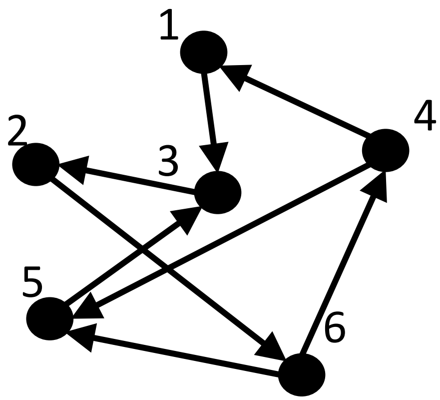
cocitation & bibliographic coupling

- mathematically similar, but practically different
- cocitation
  - is limited to influential papers
  - may change over time as the papers receive new citations
- bibliographic coupling
  - is more uniform indicator of similarity than cocitation
    - because the size of bibliography vary less than # of citations paper receive
  - can be computed as soon as a paper is published

Example with R

```
> a <- rbind(c(0,0,0,1,0,0),
+            c(0,0,1,0,0,0),
+            c(1,0,0,0,1,0),
+            c(0,0,0,0,0,1),
+            c(0,0,0,1,0,1),
+            c(0,1,0,0,0,0))
> a
     [,1] [,2] [,3] [,4] [,5] [,6]
[1,]    0    0    0    1    0    0
[2,]    0    0    1    0    0    0
[3,]    1    0    0    0    1    0
[4,]    0    0    0    0    0    1
[5,]    0    0    0    1    0    1
[6,]    0    1    0    0    0    0
```

definition of matrix A




```
> c <- a %*% t(a)
> diag(c) <- 0
> c
     [,1] [,2] [,3] [,4] [,5] [,6]
[1,]    0    0    0    0    1    0
[2,]    0    0    0    0    0    0
[3,]    0    0    0    0    0    0
[4,]    0    0    0    0    1    0
[5,]    1    0    0    1    0    0
[6,]    0    0    0    0    0    0
> b <- t(a) %*% a
> diag(b) <- 0
> b
     [,1] [,2] [,3] [,4] [,5] [,6]
[1,]    0    0    0    0    1    0
[2,]    0    0    0    0    0    0
[3,]    0    0    0    0    0    0
[4,]    0    0    0    0    0    1
[5,]    1    0    0    0    0    0
[6,]    0    0    0    1    0    0
>
```

$\boxed{C = AA^T}$

diagonal elements=0

1&5 are cocited by 4
4&5 are cocited by 6

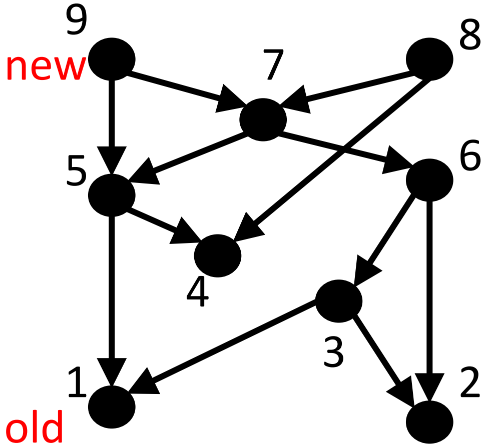
$\boxed{B = A^T A}$

diagonal elements=0

1&5 cocite 3
4&6 cocite 5

acyclic directed networks

- cycle : a closed loop (including self-edge)
- acyclic network (DAG) : without loop
- acyclic directed network
  - citation network : vertices are time-ordered



no upward edges -> no loop

proof of “acyclic -> no upward edges”

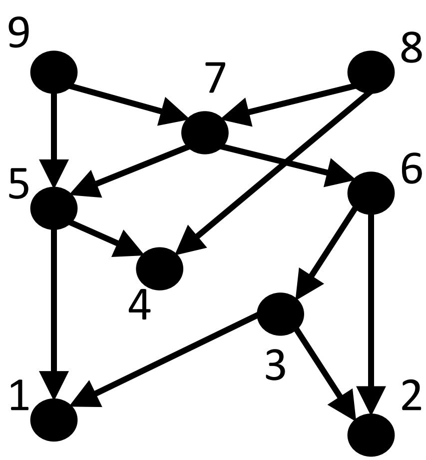
- an acyclic network of n vertices
- there must be at least one vertex that has no outgoing edges
  - a path across the network by following edges (at most n-1 times) will encounter a vertex with no outgoing edges
- then put the vertex at the bottom of the picture and remove the vertex and attached edges
- repeat the above process

cyclic or acyclic?

1. Find a vertex with no outgoing edges
2. If no such vertex exists, the network is <u>cyclic</u>. Otherwise, if such a vertex does exist, remove it and all its ingoing edges from the network.
3. If all vertices have been removed, the network is <u>acyclic</u>. Otherwise, go back to step 1

adjacency matrix of DAG is triangular

- vertices are numberd in the order they are removed in the previous algorithm
  - an edge from j to i only if j > i
  - no self-edge -> diagonal elements are 0



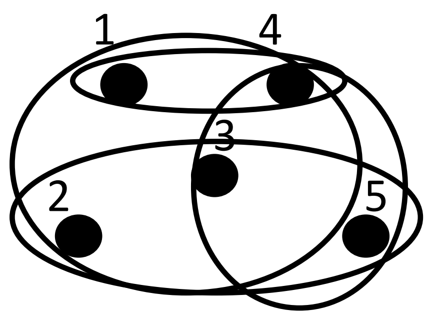
$$A = \begin{array}{cc} & \begin{array}{ccccccccc} & & & & j & & & & \end{array} \\ \begin{array}{c} \\ \\ \\ i \\ \\ \\ \\ \\ \\ \end{array} & \begin{pmatrix} 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \end{array}$$

acyclic <-> eigenvalues are zero

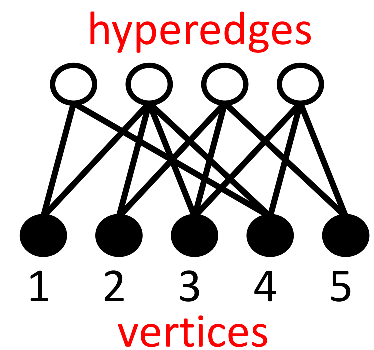
- ->
  - acyclic -> order the vertices described previously
  - adjacency matrix is strictly upper triangular
  - eigenvalues (diagonal elements) are all zero
- <-
  - prove contraposition
    - "cyclic -> at least one nonzero eivenvalue"
  - the total number Lr of cycles of length r is $L_r = \sum_{i=1}^{n} \kappa_i^r$
    - κ_i : ith eigenvalue
  - cyclic -> Lr > 0 -> at least one κ_i is greater than zero

hypergraphs

- Links sometimes join more than two vertices
  - families
  - actors in a film



{1,4}
{1,2,3,4}
{2,3,5}
{3,4,5}



Network	Vertex	Group	Section
Film actors	Actor	Cast of a film	3.5
Coauthorship	Author	Authors of an article	3.5
Boards of directors	Director	Board of a company	3.5
Social events	People	Participants at social event	3.1
Recommender system	People	Those who like a book, film, etc.	4.3.2
Keyword index	Keywords	Pages where words appear	4.3.3
Rail connections	Stations	Train routes	2.4
Metabolic reactions	Metabolites	Participants in a reaction	5.1.1

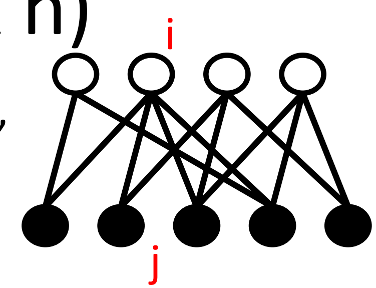
bipartite networks

- two kinds of vertices
  - original vertices and the groups to which they belong
- edges run only between vertices of unlike types
- incidence matrix B (g x n)

# of groups

# of vertices

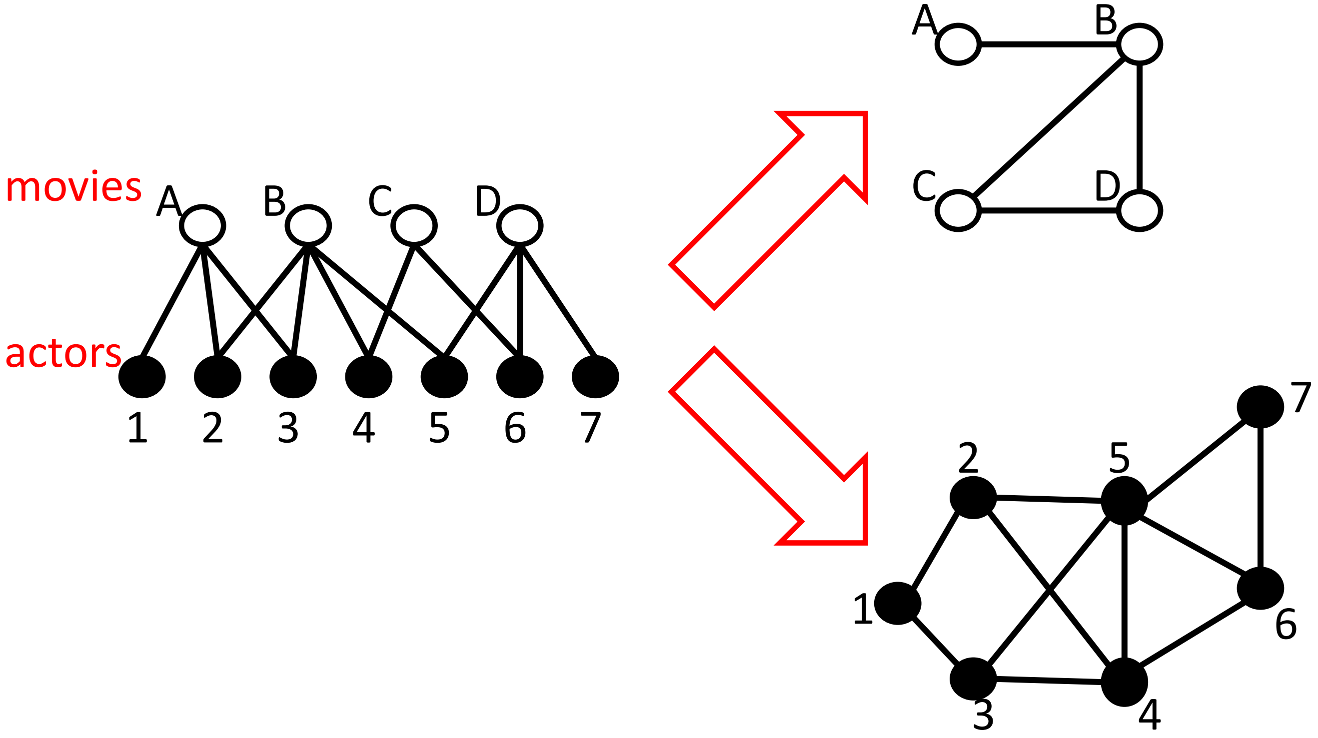
$$B_{ij} = \begin{cases} 1 & \text{if vertex j belongs to group i,} \\ 0 & \text{otherwise.} \end{cases}$$



$$\begin{array}{cc} & \text{vertices} \\ & j \\ \text{groups} & \\ i & B = \begin{pmatrix} 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 1 \end{pmatrix} \end{array}$$

one-mode projection

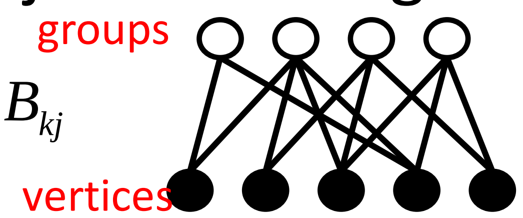
- bipartite -> unipartite
- discards a lot of the information



weighted projection

- projection onto (original) vertices
  - $B_{ki}B_{kj} = 1$ <-> i and j both belong to group k

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



n vertices

g groups

$$B = \begin{array}{cc} & \begin{array}{ccccc} & i & j & & \end{array} \\ \begin{array}{c} \\ \\ k \\ \\ \end{array} & \begin{pmatrix} 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 1 \end{pmatrix} \end{array}$$

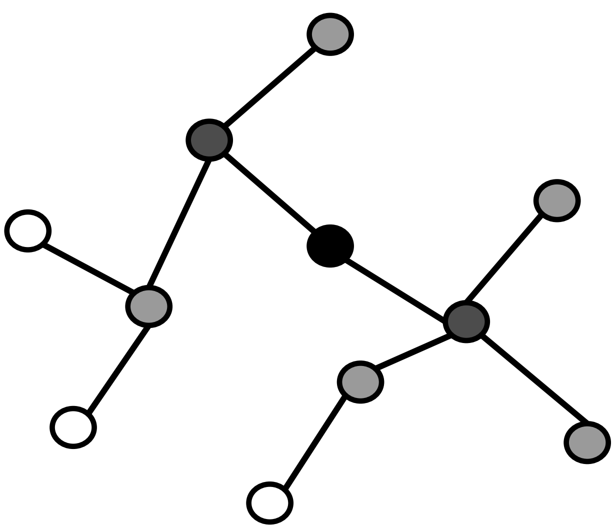
$$P = B^{T} B$$ n x n matrix

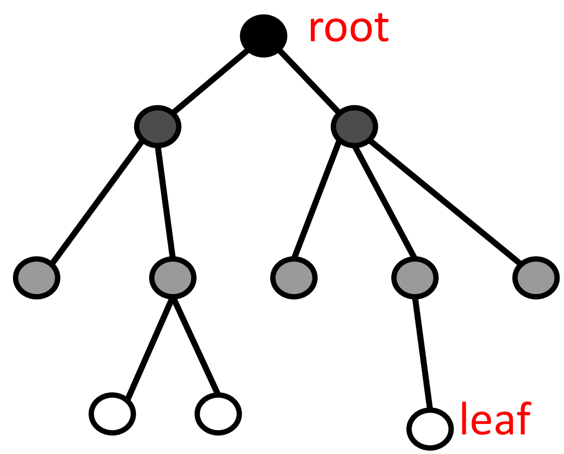
  - diagonal elements $P_{ii} = \sum_{k=1}^{g} B_{ki}^{2} = \sum_{k=1}^{g} B_{ki}$
    - # of groups to which vertex i belong
- projection onto groups

$$P' = BB^{T}$$ g x g matrix

trees

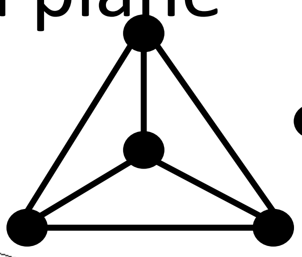
- connected, undirected network without any closed loop
- forest : collection of trees
- exactly one path between any pair of vertices
- (# of vertices) = (# of edges) + 1

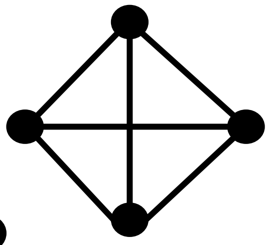


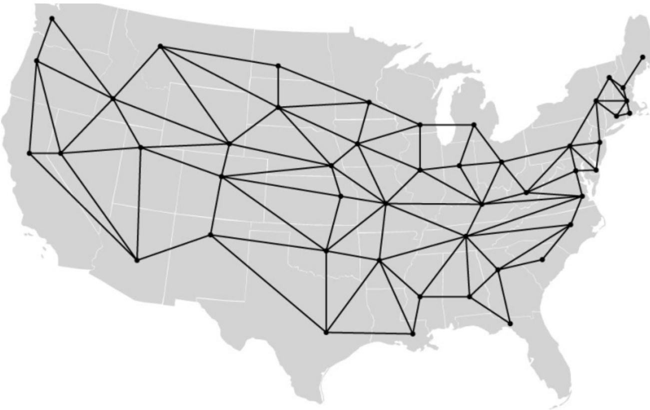


planar network

- a network that can be drawn on a plane without having any edges cross
- trees are planar
- examples
  - road network (without bridges)
  - shared borders between countries
    - four-color theorem



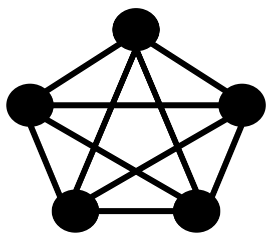




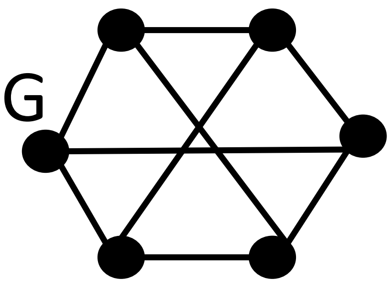
planar or not?

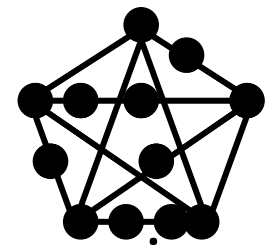
- Any network that contains a subset of vertices in the form of K_5 or UG is not planar.

K_5



UG



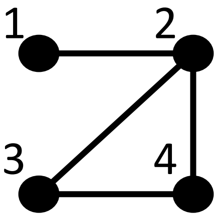


expansion of K_5

- Any expansion of K_5 or UG is not planar.
- Kuratowski's theorem
  – Every non-planar network contains at least one subgraph that is an expansion of K_5 or UG.

degree

- k_i : the degree of vertex i $\quad k_i = \sum_{j=1}^{n} A_{ij}$
  - \# of edges connected to it



$$\begin{array}{rc} & \\ k_2 & \\ A = & \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{pmatrix} \end{array}$$

m

- (sum of all degrees) = 2 x (# of edges)

$$\sum_{i=1}^{n} k_i = \sum_{i=1}^{n} \sum_{j=1}^{n} A_{ij} = 2m$$

- c : mean degree

$$c = \frac{1}{n} \sum_{i=1}^{n} k_i = \frac{2m}{n}$$

- maximum possible number of edges

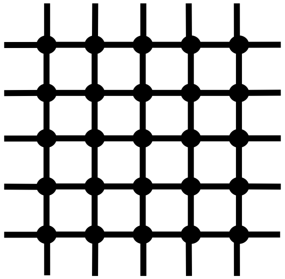
$$ {}_nC_2 = \binom{n}{2} = \frac{1}{2} n(n-1)$$

density

- density (or connectance) $\rho = \frac{m}{\binom{n}{2}} = \frac{2m}{n(n-1)} = \frac{c}{n-1} \approx \frac{c}{n}$
  $0 \leq \rho \leq 1$

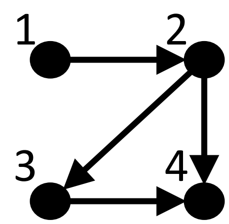
  networks is sufficiently large

- dense : $\rho \to const$ as $n \to \infty$
- sparse : $\rho \to 0$ as $n \to \infty$
- almost all of the networks we consider are sparse (except food webs)
  – important for developing algorithms and models
- k-regular : all vertices have degree k



degrees in directed networks

- in-degree $k_i^{in} = \sum_{j=1}^{n} A_{ij}$
- out-degree $k_i^{out} = \sum_{i=1}^{n} A_{ij}$



$$
A = \begin{array}{cc} & \begin{array}{cccc} & k_2^{out} & & \end{array} \\ \begin{array}{c} \\ k_2^{in} \\ \\ \\ \end{array} & \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \end{pmatrix} \end{array}
$$

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

- mean in-degree c_{in}
- mean out-degree c_{out}

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

- -> we will just denote both by c $c = \frac{m}{n}$

path

- a route across the network that runs from vertex to vertex along the edges of the network
- self-avoiding path : a path that does not intersect itself
- length : # of edges traversed along the path
- # of paths of a given length r
  - $A_{ik}A_{kj} = 1$ if there is a path j -> k -> i
  - # of paths of length 2 from j to i : $N_{ij}^{(2)} = \sum_{k=1}^{n} A_{ik}A_{kj} = \left[A^2\right]_{ij}$
  - # of paths of length r from j to i : $N_{ij}^{(r)} = \left[A^r\right]_{ij}$

cycles

- paths of length r that start and end at the same vertex $L_r = \sum_{i=1}^{n} \left[A^r\right]_{ii} = TrA^r$ 1->2->3->1 and 2->3->1->2 are distinct
  - counting each loop only once is not easy
- L_r in terms of eigenvalues of A (undirected)
- diagonal matrix of eigenvalues

$$A = UKU^T$$

orthogonal matrix of eigenvectors

undirected graph -> A is symmetric
-> A is diagonalizable

$$A^r = \left(UKU^T\right)^r = UK^rU^T \qquad \because UU^T = U^TU = E$$

$$L_r = Tr(UK^rU^T) = Tr(U^TUK^r) = TrK^r = \sum_i \kappa_i^r$$

$$\because Tr(AB) = Tr(BA)$$

κ_i : ith eigenvalue of A

cycles of directed graphs

- $L_r = \sum_i \kappa_i^r$ is true also for directed graphs
  - although A cannot be diagonalized
- proof
  - Every real matrix can be written in the form
    - Schur decomposition

$$A = Q\underbrace{T}_{\text{upper triangular matrix}}\underbrace{Q^T}_{\text{orthogonal matrix}}$$

  - Eigenvalues of T are the same as those of A

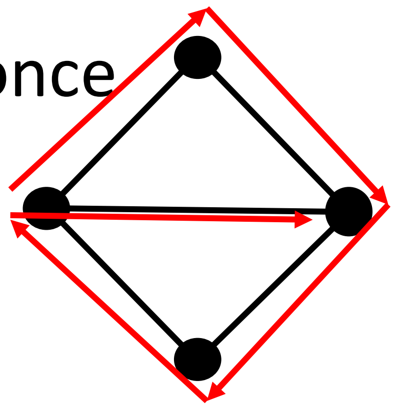
$$L_r = TrA^r = Tr(QT^rQ^T) = Tr(Q^TQT^r) = TrT^r = \sum_i \kappa_i^r$$

geodesic path (shortest path)

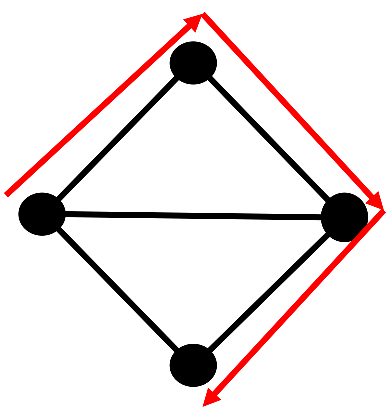
- geodesic distance between vertices i and j
  - smallest value of r such that $\left[A^r\right]_{ij} > 0$
- self-avoiding: no loop
- diameter : the longest geodesic path between any pair of vertices in the network

Eulerian and Hamiltonian path

- Eulerian path
  - a path that traverses each edge exactly once

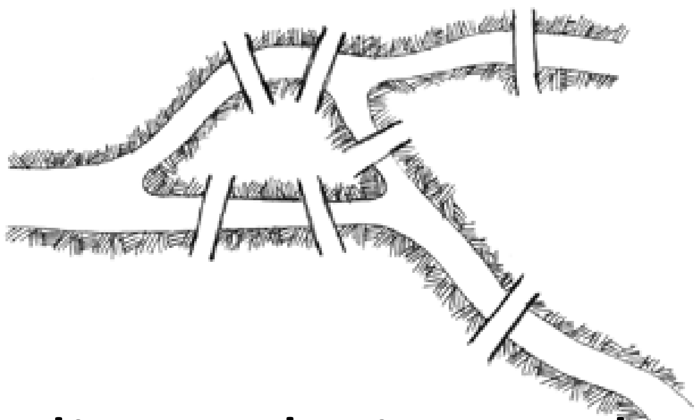
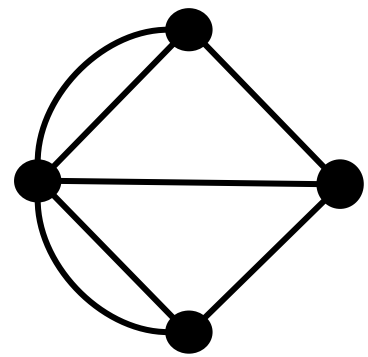


- Hamiltonian path
  - a path that visit each vertex exactly once
  - self-avoiding



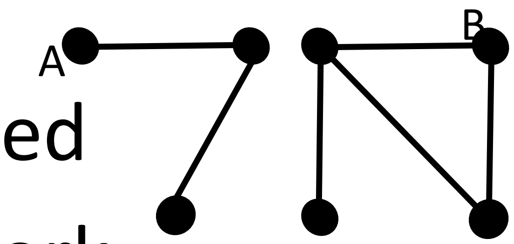
Königsberg bridge problem

- Does there exist any walking route that crosses all seven bridges exactly once each?

- -> finding Eulerian path on the right network
  - at most two vertices with odd degree
  - all four vertices have odd degree -> no solution

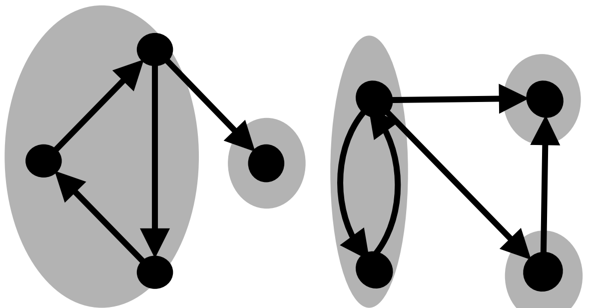
components



- no path from A to B -> disconnected
- component : subgroups in a network
- block diagonal matrix

$$A = \begin{pmatrix} \square & 0 & \cdots \\ 0 & \square & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix}$$

- components in directed networks
  - two (undirected network)
  - five (directed network)



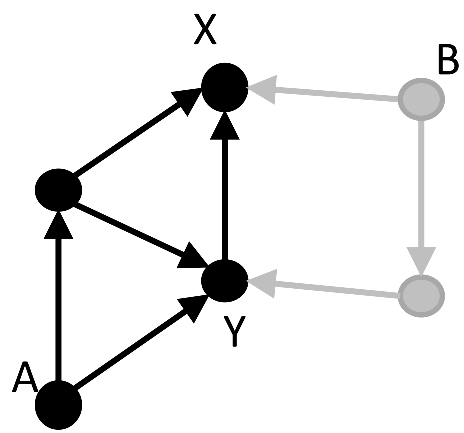
strongly connected components (SCC)

- A and B are connected if and only if there exists both A->B and B->A
- SCC is a maximal subset of vertices such that there is a directed path in both directions between every pair in the subset
- each vertex belongs to exactly one SCC
- every SCC with more than one vertex must contain at least one cycle

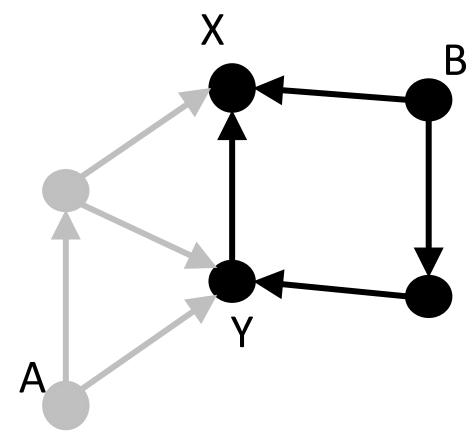
out-component in a directed network

- the set of vertices that are reachable via directed paths starting at a specific vertex A
- depends on network and starting vertex

out-component of vertex A

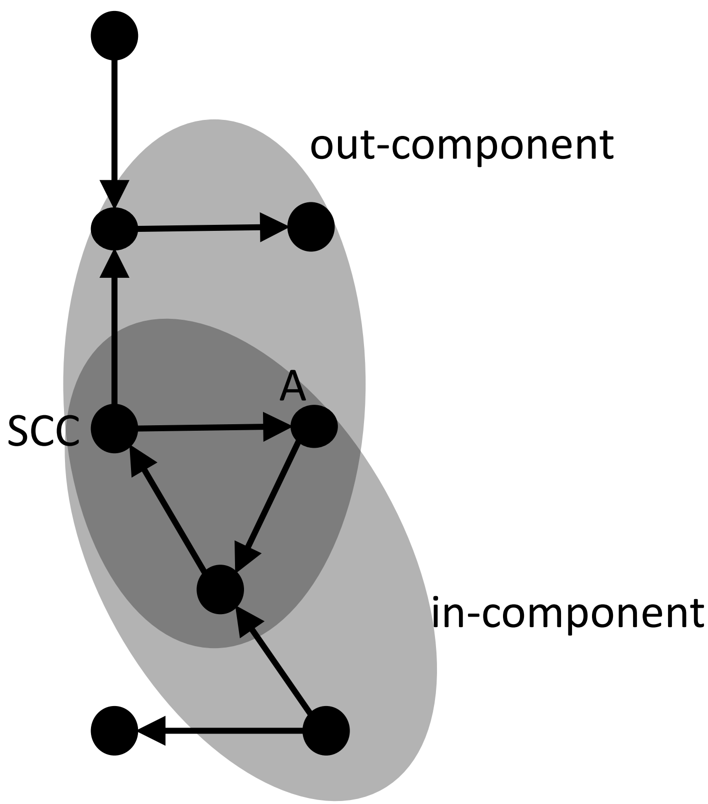


out-component of vertex B



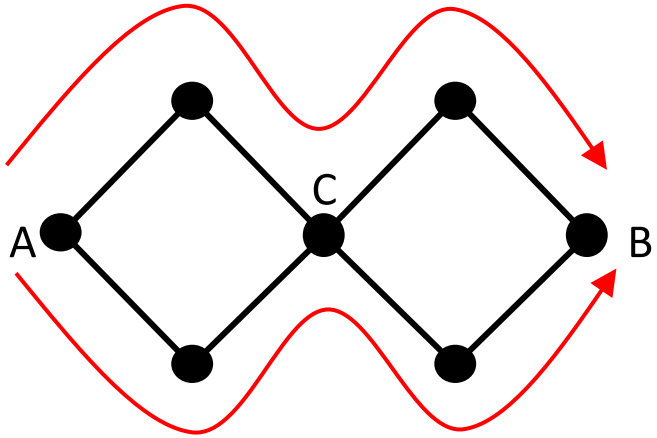
in-component & out-component

- in-component :reachable to vertex A
- out-component : reachable from vertex A
- SCC: intersection of in and out



independent paths

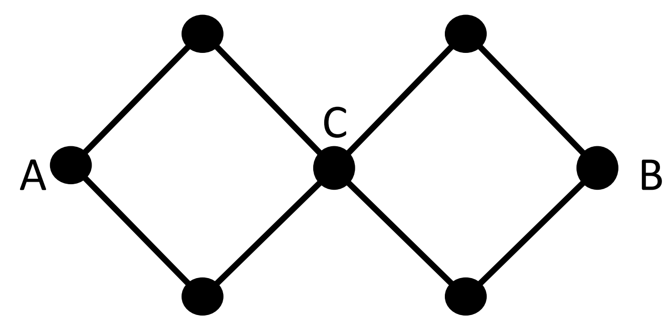
- edge-independent path share no edges
- vertex-independent path share no vertices (except starting and ending vertices)
- vertex-independent -> edge-independent
  - but the reverse is not true



2 edge-independent paths
1 vertex-independent path

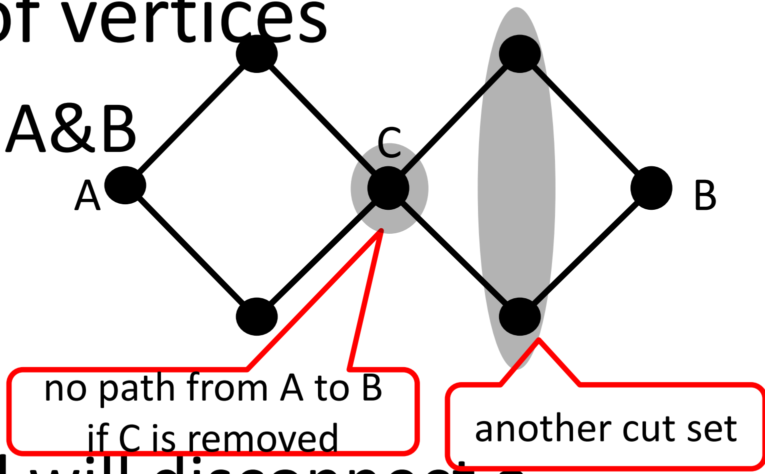
more on independent paths

- There can be only a finite number of independent paths between any two vertices in a finite network
- connectivity : # of independent paths between a pair of vertices
  - A and B have edge connectivity 2 but vertex connectivity 1
  - strength of connection
    - discovering communities
    - finding bottlenecks



cut set

- a set of vertices whose removal will disconnect a specified pair of vertices
  - C forms a cut set of size 1 for A&B

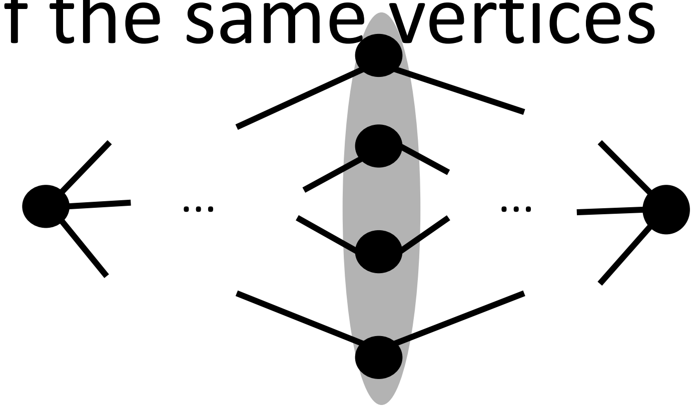


- edge cut set
  - a set of edges whose removal will disconnect a specified pair of vertices
- minimum cut set : the smallest cut set

Menger's theorem

- If there is no cut set of size less than n between a given pair of vertices, then there are at least n independent paths between the same vertices
  - this theorem applies both to edges and to vertices
- The size of the minimum vertex cut set that disconnects a given pair of vertices is equal to the vertex connectivity of the same vertices

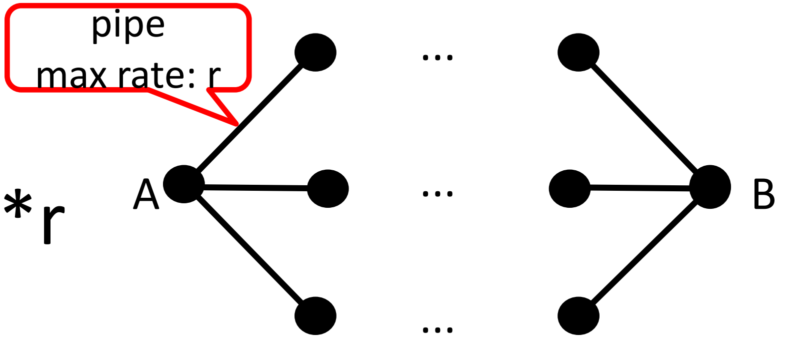
min cut set		independent paths
n	->	n or more
n or more	<-	n



maximum flow

- a network of water pipes
- the max rate from A to B =

(# of edge-independent pahts)*r



- proof
  - n independent paths -> at least n*r of flows (lower bound)
  - a cut set of n edges -> at most n*r of flows (upper bound)
  - the max rate is exactly n*r
- max-flow/min-cut theorem
  - individual pipes can have different capacities

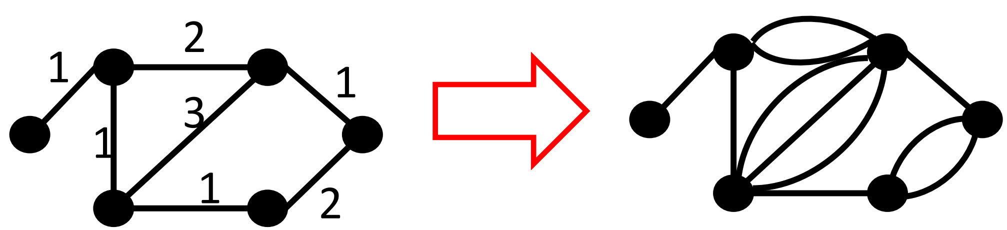
these three are numerically equal

- the edge connectivity of a pair of vertices
  - the number of edge-independent paths
- the size of the minimum edge cut set
  - the number of edges that must be removed to disconnect them
- the maximum flow between the vertices

these are equal for directed network as well

max-flows on weighted networks

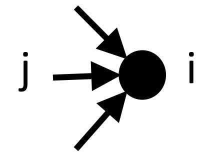
- max-flows/min-cut theorem can be extended to weighted networks
  - the maximum flow between a given pair of vertices in a network is equal to the sum of the weights on the edges of the minimum edge cut set that separate the same two vertices
- proof
  - transform weighted edges to multiedges



diffusion process on networks

- spreading (ideas/diseases/...) on networks
- ψ_i:some commodity or substance at vertex i
- $C(\psi_i-\psi_j)$: flow from i to j (C:constant)

$$\frac{d\psi_i}{dt} = C\sum_j A_{ij}(\psi_j - \psi_i)$$



degree of i

$$\frac{d\psi_i}{dt} = C\sum_j A_{ij}\psi_j - C\psi_i\sum_j A_{ij} = C\sum_j A_{ij}\psi_j - C\psi_i k_i$$

$$= C\sum_j (A_{ij} - \delta_{ij}k_i)\psi_j$$

1 if i=j
0 otherwise

$$\frac{d\psi}{dt} = C(A-D)\psi \qquad D = \begin{pmatrix} k_1 & 0 & 0 & \cdots \\ 0 & k_2 & 0 & \cdots \\ 0 & 0 & k_3 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

graph Laplacian (1)

$$\frac{d\psi}{dt} = C(A-D)\psi$$

graph Laplacian

$$L = D - A$$

$$\Rightarrow \quad \frac{d\psi}{dt} + CL\psi = 0$$

similar to diffusion equation

- graph Laplacian is for
  - random walk
  - resistor networks
  - graph partitioning
  - network connectivity

graph Laplacian (2)

$$L = D - A$$

$$L_{ij} = \begin{cases} k_i & \text{if i = j,} \\ -1 & \text{if i} \neq \text{j and there is an edge (i,j),} \\ 0 & \text{otherwise} \end{cases}$$

degree

edge

$$D = \begin{pmatrix} k_1 & 0 & -1 & \cdots \\ 0 & k_2 & 0 & \cdots \\ -1 & 0 & k_3 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

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

- ψ as linear combination of eigenvectors of L

$$\psi(t) = \sum_i a_i(t) v_i$$

v_i:eigenvectors of L $\quad Lv_i = \lambda_i v_i$

eigenvectors of a symmetric matrix are orthogonal

$$\frac{d\psi}{dt} + CL\psi = 0 \Rightarrow \sum_i \left(\frac{da_i}{dt} + C\lambda_i a_i \right) v_i = 0$$

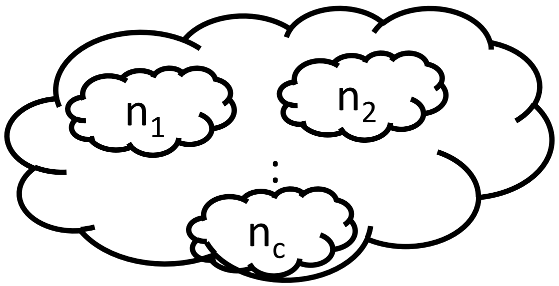
$$\frac{da_i}{dt} + C\lambda_i a_i = 0 \qquad a_i(t) = a_i(0) e^{-C\lambda_i t}$$

eigenvalues of graph Laplacian

- Laplacian is symmetric, so it has real eigenvalues
- they are also non-negative

components and connectivity

- Laplacian -> block diagonal



Laplacian of each component

$$L = \begin{pmatrix} \square & 0 & \cdots \\ 0 & \square & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix}$$

$$v = \begin{pmatrix} 1 \\ \vdots \\ 0 \\ \vdots \end{pmatrix} \begin{array}{l} \left.\begin{matrix} \\ \\ \end{matrix}\right\} n_1 \text{ ones} \\ \left.\begin{matrix} \\ \\ \end{matrix}\right\} \text{zeros} \end{array}$$

is an eigenvector of L with eigenvalue zero

$$Lv = 0v$$

- (# of zero eigenvalues) = (# of components)

->the second eigenvalue of graph Laplacian λ_2 is non-zero if and only if the network is connected

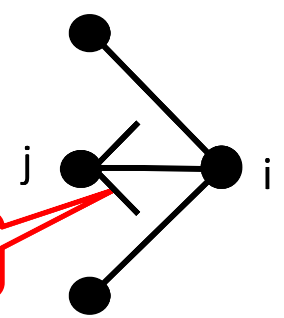
random walk

- a path across a network created by taking repeated random steps
  - used for sampling and ranking
- $p_i(t)$:probability that the walk is at vertex i at time t

$$p_i(t) = \sum_j \frac{A_{ij}}{k_j} p_j(t-1)$$

vector with element p_i

$$\mathbf{p}(t) = \mathbf{A}\mathbf{D}^{-1}\mathbf{p}(t-1)$$



degree:k_i

$$\mathbf{D}^{-1} = \begin{pmatrix} 1/k_1 & 0 & 0 & \cdots \\ 0 & 1/k_2 & 0 & \cdots \\ 0 & 0 & 1/k_3 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

$$\mathbf{D}^{1/2} = \begin{pmatrix} \sqrt{k_1} & 0 & 0 & \cdots \\ 0 & \sqrt{k_2} & 0 & \cdots \\ 0 & 0 & \sqrt{k_3} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

reduced adjacency matrix

symmetric

$$\mathbf{p}(t) = \mathbf{A}\mathbf{D}^{-1}\mathbf{p}(t-1)$$

$$\mathbf{D}^{-1/2}\mathbf{p}(t) = [\mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2}][\mathbf{D}^{-1/2}\mathbf{p}(t-1)]$$

$$\mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2} = \begin{cases} 1/\sqrt{k_i k_j} & \mathbf{A}_{ij} = 1 \\ 0 & \mathit{otherwise} \end{cases}$$

- when $t \to \infty$

$$\mathbf{p} = \mathbf{A}\mathbf{D}^{-1}\mathbf{p}$$

$$(\mathbf{I} - \mathbf{A}\mathbf{D}^{-1})\mathbf{p} = (\mathbf{D} - \mathbf{A})\mathbf{D}^{-1}\mathbf{p} = \mathbf{L}\mathbf{D}^{-1}\mathbf{p} = 0$$

->D^{-1}p is an eigenvector of the Laplacian with eigenvalue 0

- connected network ->only one eigenvector (with eigenvalue 0) whose components are all equal

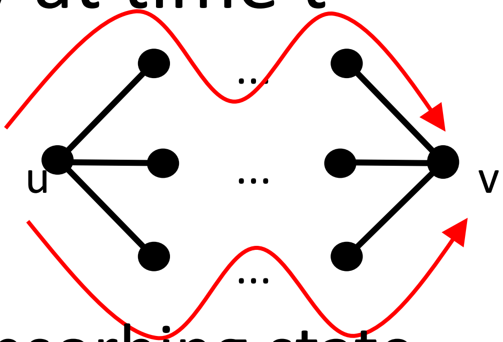
$$\mathbf{D}^{-1}\mathbf{p} = a\mathbf{1} \qquad p_i = ak_i$$

$$\mathbf{p} = a\mathbf{D}\mathbf{1} \qquad p_i = \frac{k_i}{\sum_j k_j} = \frac{k_i}{2m}$$

-> probability is proportional to the degree of the vertex

random walk with absorbing state(1)

- first passage time:# of steps from u to v
- $p_v(t)$: probability that a walk is at v at time t
- $p_v(t) - p_v(t-1)$: prob. that a walk has first passage time exactly t



v:absorbing state
(never go out from v)

- mean first passage time

$$\tau = \sum_{t=0}^{\infty} t[p_v(t) - p_v(t-1)]$$

- trick for calculating $p_v(t)$ is in the next slides

random walk with absorbing state(2)

if i≠v, no term in A_{vj} in the sum

- $A_{iv}=0 \because$ v is absorbing state

$$p_i(t)=\sum_j \frac{A_{ij}}{k_j}p_j(t-1)=\sum_{j(\neq v)}\frac{A_{ij}}{k_j}p_j(t-1) \qquad A_{iv}=0$$

$$\mathbf{p}'(t)=\mathbf{A}'\mathbf{D}'^{-1}\mathbf{p}'(t-1)$$

p'(t):p with vth element removed

A',D': A and D with vth row and column removed

$$\mathbf{p}'(t)=[\mathbf{A}'\mathbf{D}'^{-1}]^t\mathbf{p}'(0)$$

$$p_v(t)=1-\sum_{i(\neq v)}p_i(t)=1-\mathbf{1}^T\mathbf{p}'(t) \qquad \mathbf{1}=(1,1,1,\dots)$$

$$\tau=\sum_{t=0}^{\infty}t[p_v(t)-p_v(t-1)]=\sum_{t=0}^{\infty}t\mathbf{1}^T[\mathbf{p}'(t-1)-\mathbf{p}'(t)]=\mathbf{1}^T[\mathbf{I}-\mathbf{A}'\mathbf{D}'^{-1}]^{-1}\mathbf{p}'(0)$$

$$\because \sum_{t=0}^{\infty}t(\mathbf{M}^{t-1}-\mathbf{M}^t)=[\mathbf{I}-\mathbf{M}]^{-1}$$

random walk with absorbing state(3)

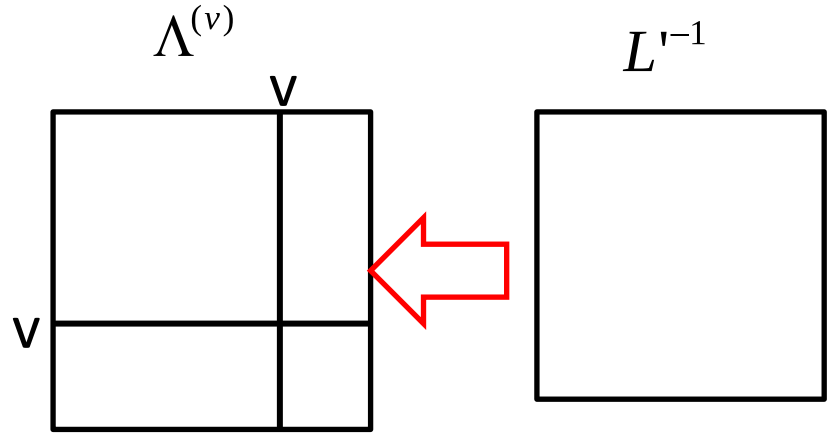
$$[\mathbf{I}-\mathbf{A}'\mathbf{D}'^{-1}]^{-1} = \mathbf{D}'[\mathbf{D}'-\mathbf{A}']^{-1} = \mathbf{D}'\mathbf{L}'^{-1} \qquad \because [\mathbf{AB}]^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$$

L' : graph Laplacian with the vth row and column removed (vth reduced Laplacian)

$$\tau = \mathbf{1}^T[\mathbf{I}-\mathbf{A}'\mathbf{D}'^{-1}]^{-1}\mathbf{p}'(0) = \mathbf{1}\cdot\mathbf{D}'\mathbf{L}'^{-1}\cdot\mathbf{p}'(0)$$

- L' can have inverse
- $\Lambda^{(v)}$: equal to L'^{-1} with a vth row and column reintroduced

$$\Lambda^{(v)}_{ij} = \begin{cases} 0 & \text{if } i = v \text{ or } j = v \\ [L'^{-1}]_{ij} & \text{if } i < v \text{ and } j < v \\ [L'^{-1}]_{i-1,j} & \text{if } i > v \text{ and } j < v \\ [L'^{-1}]_{i,j-1} & \text{if } i < v \text{ and } j > v \\ [L'^{-1}]_{i-1,j-1} & \text{if } i > v \text{ and } j > v \end{cases}$$



random walk with absorbing state(4)

$$\tau = \mathbf{1} \cdot \mathbf{D}'\mathbf{L}'^{-1} \cdot \mathbf{p}'(0)$$

a walk starting at vertex u at time 0

$$\mathbf{p}'(0) = (0,0,\ldots,\underset{u}{1},0,\ldots,0)$$

$$\therefore \tau = \sum_i k_i \Lambda_{iu}^{(v)}$$

- mean first passage time to v is the sum over other starting vertices