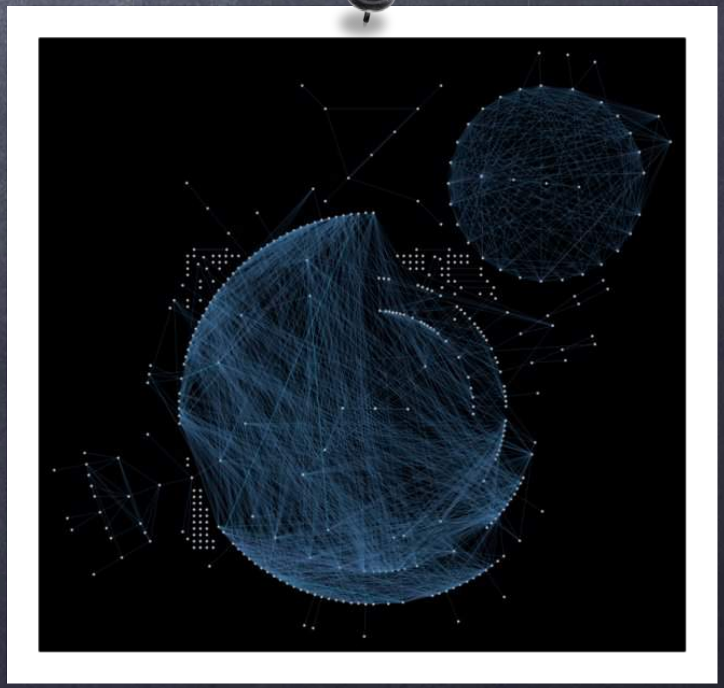
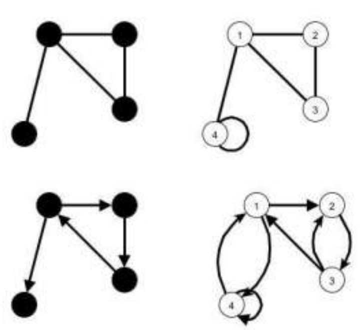
Random graph models for analysing real networks

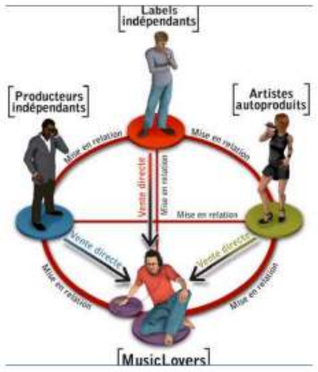
blerina sinaimeri



What is a network?

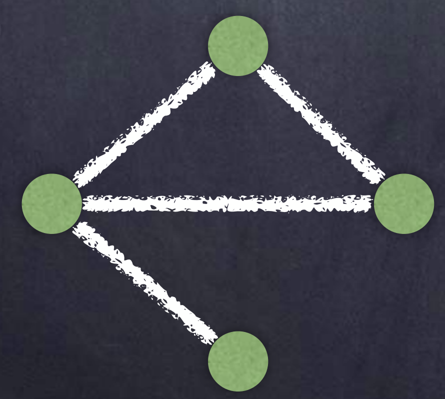
- A network is a collection of points (nodes, vertices) with their connections (edges) and where nodes/edges have names.
- Examples of real networks: internet (routers or computers connected between them), electric network, road maps, airline networks.
- Examples of virtual networks: World Wide Web, Facebook, Food-web in ecology.





Definitions

- A graph G=(V,E) where V is the set of nodes and E is the set of edges.
- |V|=n is the size of the graph.
- A networks is a graph where edges and or vertices have attributes (names).



Undirected graph

Directed graph

Basic concepts (I)

- A path in a graph is a sequence $v_1, v_2, \ldots v_k$ where no vertex is repeated and such that (v_i, v_{i+1}) is an edge in E.
- A cycle is a is a sequence $v_1, v_2, \ldots v_k, v_1$ where no vertex except v_1 is repeated and such that (v_i, v_{i+1}) is an edge in E.
- Two vertices u and v are connected in a graph if there exists a path between them.
- A graph is connected if every two vertices in V are connected.
- A graph can be decomposed in a unique way in a set of connected components. How many connected components can a graph have?

Basic concepts (II)

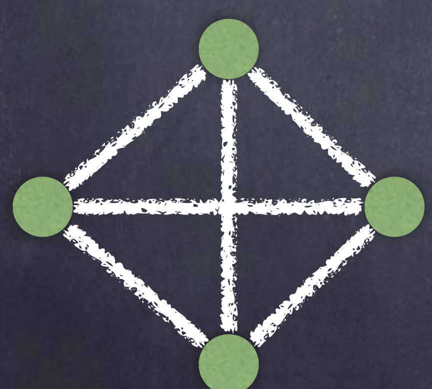
- The distance d(u,v) between two vertices u and v is the length of the shortest path between them.

- The degree of a vertex $d_G(v)$ is the number of vertices adjacent (connected) to it.

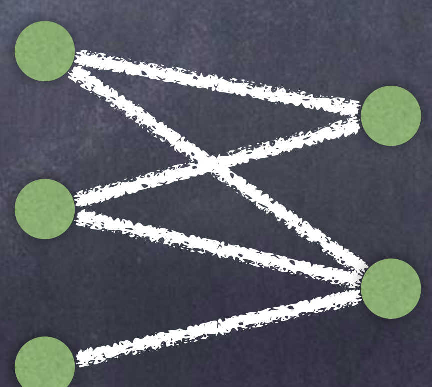
- A basic result in graph theory is the following: $\sum_{i \in V} d_i = 2|E|$

"Famous"graphs (I)

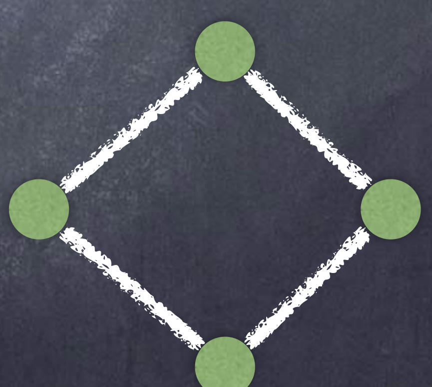
- The complete graph K_n
- The bipartite graph $G(V_1,V_2,E)$
- An odd (even) cycle C_{2k+1} (C_{2k})



Complete graph
(Clique)



Biparite graph

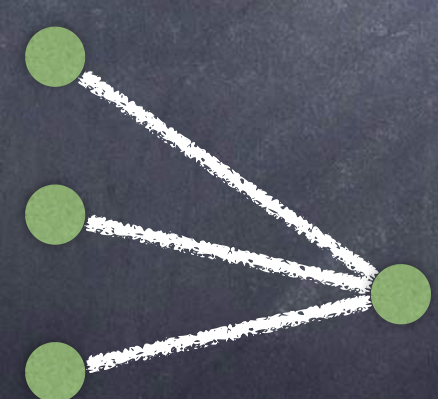


C_4

"Famous" graphs (II)

- The tree: is a connected graph without cycles
- The star S_k

A tree



The star S_3

Basic concepts (III)

- A subgraph of G=(V,E) is a graph G′=(V′, E′) with V′ and E′ subsets of V and E, respectively.
- A spanning subgraph of G is a subgraph that is connected and contains all the vertices of G.
- An induced subgraph of G is a subgraph G=(V′, E′) where E′contains all the edges incident to V′.

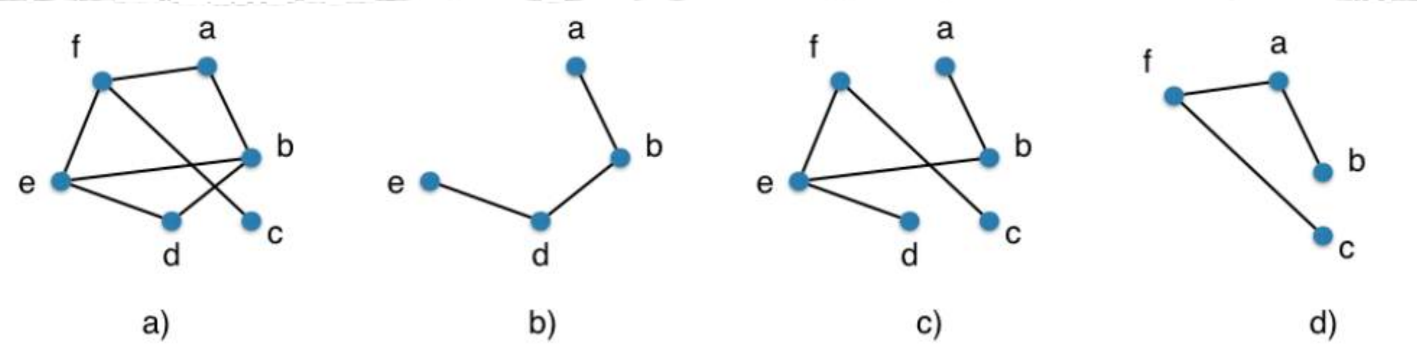
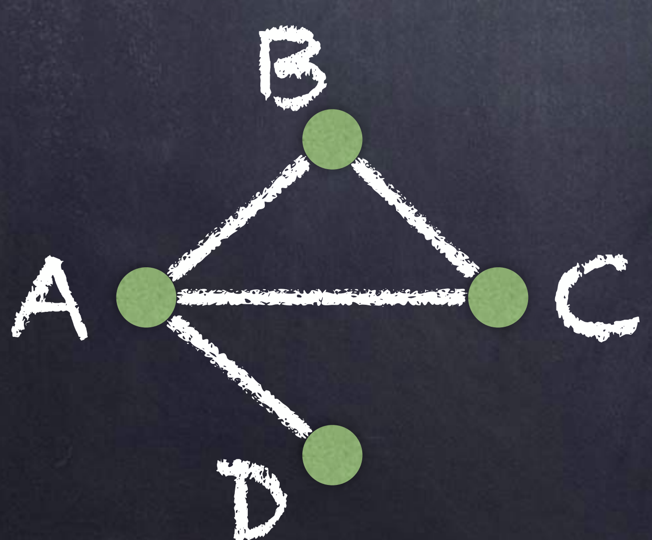


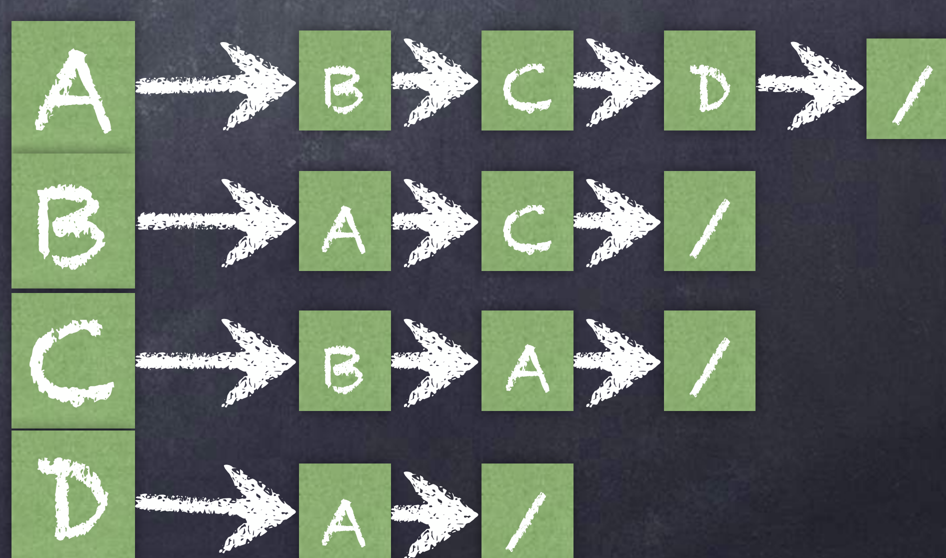
Figure 4: a) A graph *G*, b) A subgraph of *G*, c) A spanning subgraph of *G*, d) An induced subgraph of *G*.

Representation of Graphs: Adjacency Matrix and Adjacency List

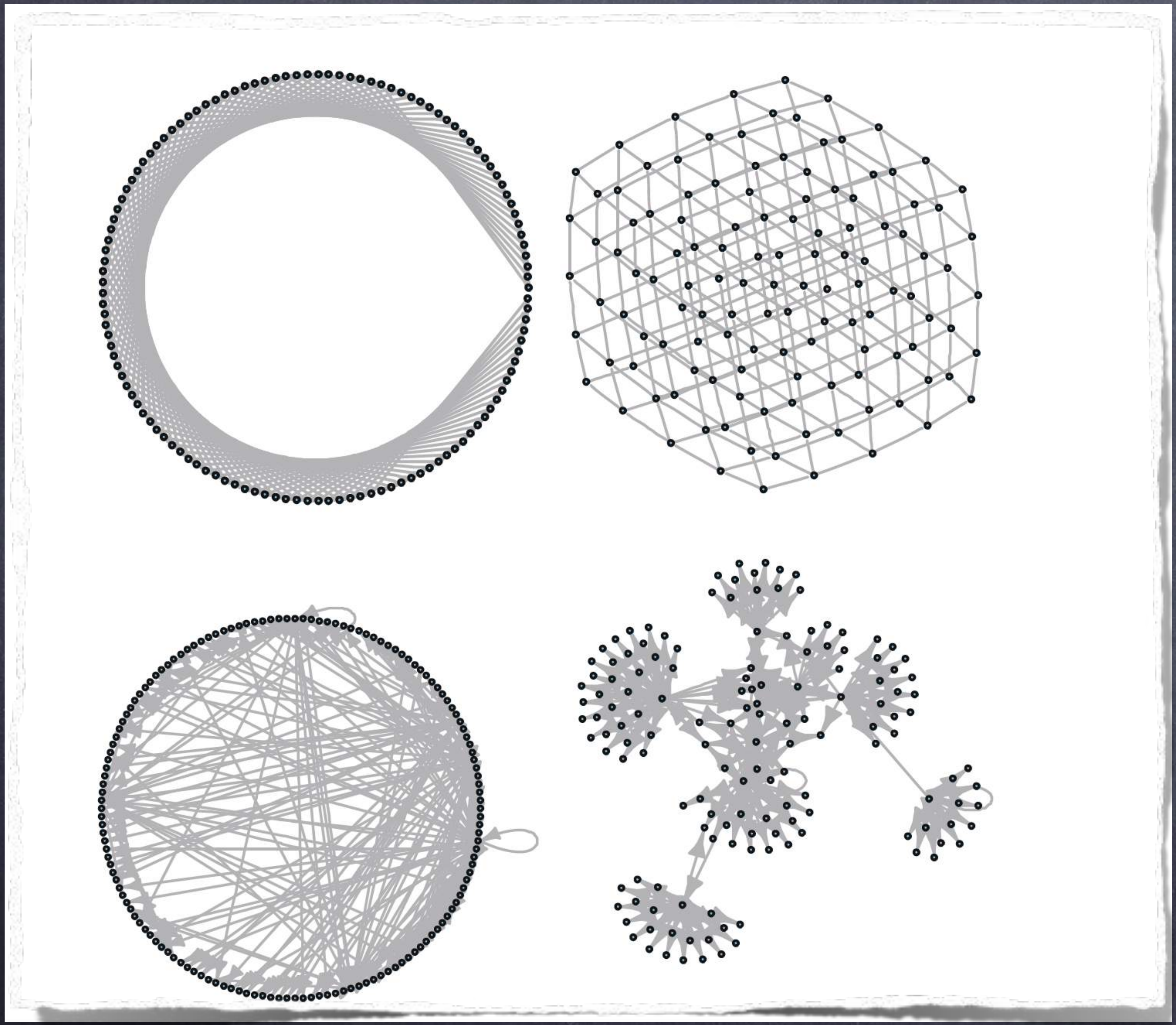
- Adjacency matrix: $M_{n \times n}$ such that $M[i,j]=1$ iff $\{i,j\}$ in E and $M[i,j]=0$ otherwise.
- Adjacency list: L_n such that $l[i]$ contains the list of adjacent vertices of i.



	A	B	C	D
A	0	1	1	1
B	1	0	1	0
C	1	1	0	0
D	1	0	0	0



Visual representation



Each row shows two different representation of the same network. (picture from Kolaczyk and Csardi (2014)).

Summary statistics of a network

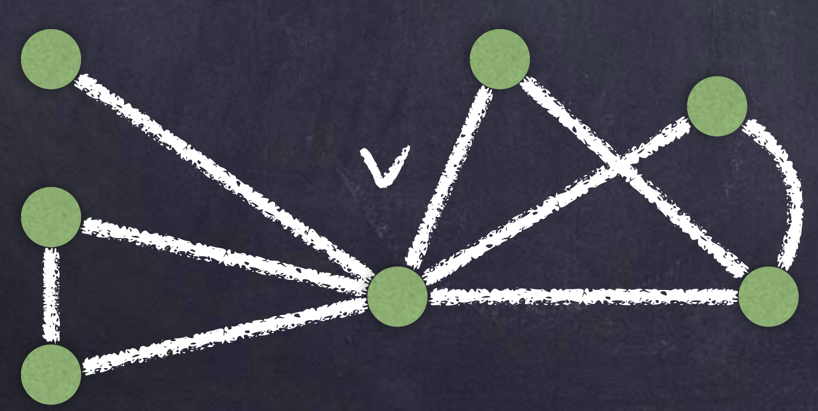
- A small network can be visualised directly by its graph but a large one can be more difficult.
- Thus, it is necessary to define a set of summary statistics to describe and compare networks.
  - Clustering
  - Diameter and average path length
  - Centrality
  - Degree distributions
  - ...

Individual clustering coefficient

- The clustering coefficient measures the degree of clustering of a typical node's neighborhood. It is defined as the likelihood that any two nodes with a common neighbor are themselves connected. The individual clustering coefficient for a vertex v is given by:

$$CC(G,v) = \frac{\#\text{triangles that contain } v}{(1/2)d(v)[d(v)-1]}$$

Example



d(v)=6 and v participates in 3 triangles. Hence CC(G,v)=3/[0.5*6*5]=1/5

Overall clustering coefficient

- In a social network measures the extent to which my friends are friends with one another.

$$CC(G) = (1/|V|) \sum_{v} CC(G,v)$$

Overall clustering coefficient

- In a social network measures the extent to which my friends are friends with one another.

$$CC(G)= (1/|V|) \sum_{v} CC(G,v)$$

- What is the maximum and minimum value for CC(G)?

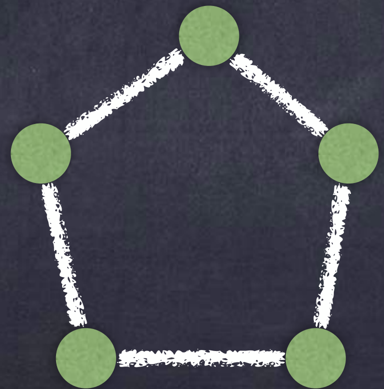
Overall clustering coefficient

- In a social network measures the extent to which my friends are friends with one another.

$$CC(G) = (1/|V|) \sum_{v} CC(G,v)$$

- What is the maximum and minimum value for CC(G)?

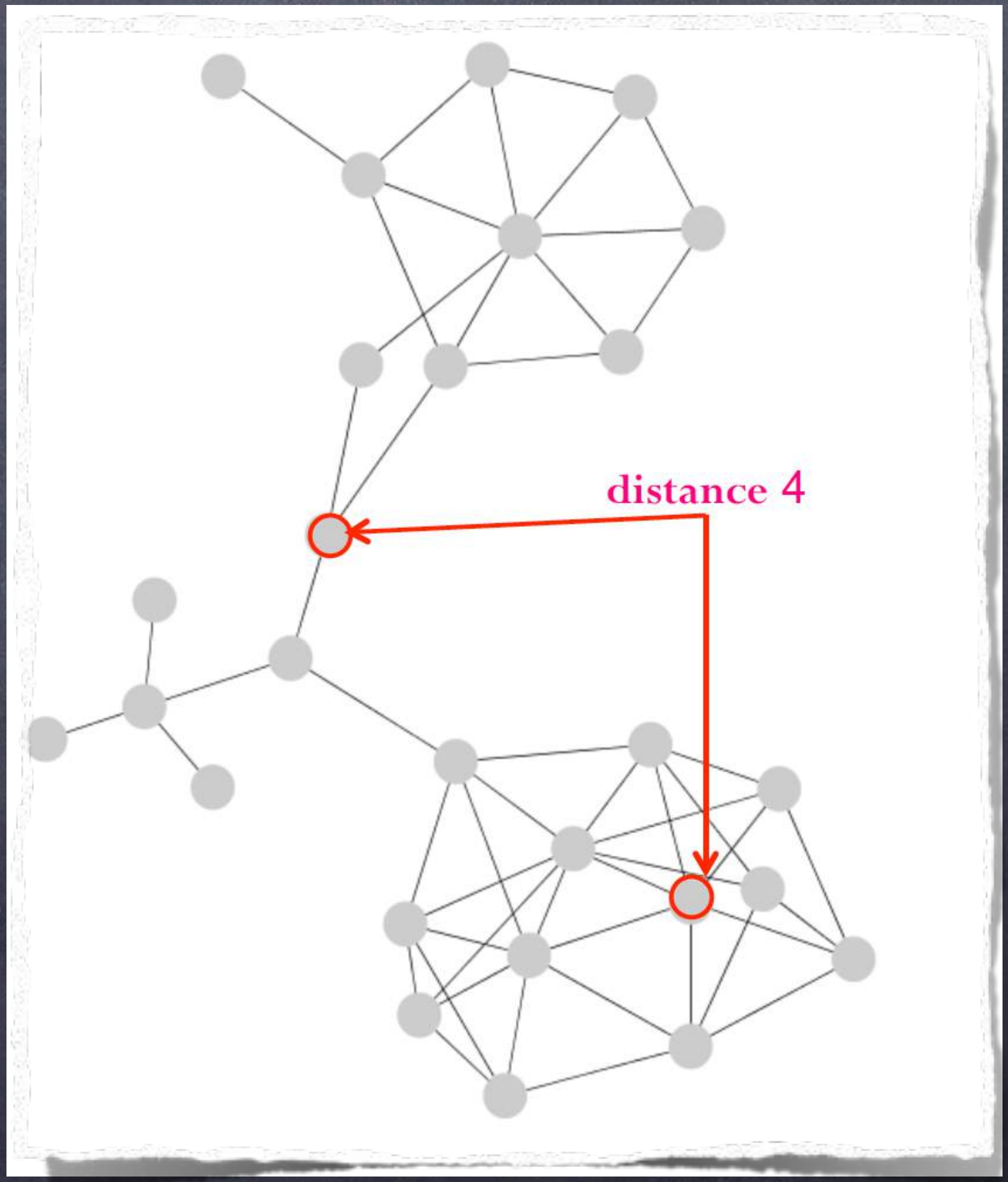
Max nr of triangles
CC(G)=1



No triangles
CC(G)=0

Diameter and average path length

- The diameter of a graph is the maximum length of any shortest paths between two vertices.
  - $\text{diameter}(G) = \max_{u,v} d(u,v)$
- The average path length is the average distance between any two nodes in the network:
  - $\text{averg\_path}(G) = \sum_{u,v} d(u,v) \,/\, n(n-1)$



Diameter and average path length?

- How are the diameter and average path length related?
    - ????

Diameter and average path length?

- How are the diameter and average path length related?
    - Average path length is bounded from above by the diameter; in some cases, it can be much shorter than the diameter.

Centrality

- There are different concepts of centrality all trying to capture the importance of a vertex position in the network.
    - Degree centrality of a vertex v: dc(v)=d(v)/(n-1)
    - Betweeness of a vertex, b(v), which is the number of shortest paths betweens pair of vertices which pass through v. This is an indication of a vertex's centrality in the network.

Degree distribution

- The degree distribution is a description of relative frequencies of vertices that have different degrees d.
  - For a given graph G: P(G) is the fraction of vertices of degree d for d=0,...,n-1
  - For a random graph G: P(G) is a probability distribution.

Random graph models

- A random graph model is a collection (finite or countable) $\mathcal{G}$ of graphs and a probability distribution $\mathcal{P}$ over the collection $\mathcal{G}$.

A Simple Random Graph Model
Erdös-Renyi model

- The Erdös-Renyi model is denoted G(n,p) and is one of the simplest random models.
- All graphs on n vertices.
- Every edge is formed with probability $p \in (0, 1)$ independently of every other edge.

A Simple Random Graph Model
Erdös-Renyi model

- What is the expected degree of a vertex v?
    - E[d(v)] ??

A Simple Random Graph Model
Erdös-Renyi model

- What is the expected degree of a vertex v?
  - $E[d(v)] = (n-1)p$
- What is the expected number of edges?
  - $E[|E|] = ??$

A Simple Random Graph Model
Erdös-Renyi model

- What is the expected degree of a vertex v?
  - $E[d(v)] = (n-1)p$
- What is the expected number of edges?
  - $E[|E|] = p\, n(n-1)/2$