Graph Theory

Topics

- Bridges of Königsberg & Euler’s Theorem
- Difference Between Network and Graph
- Degree, Average Degree and Degree Distribution
- Adjacency Matrix
- Paths, Distances, and Diameter
- Clustering
- Bipartite Networks
- Real Networks are Sparse

The Bridges of Königsberg & Euler's Theorem

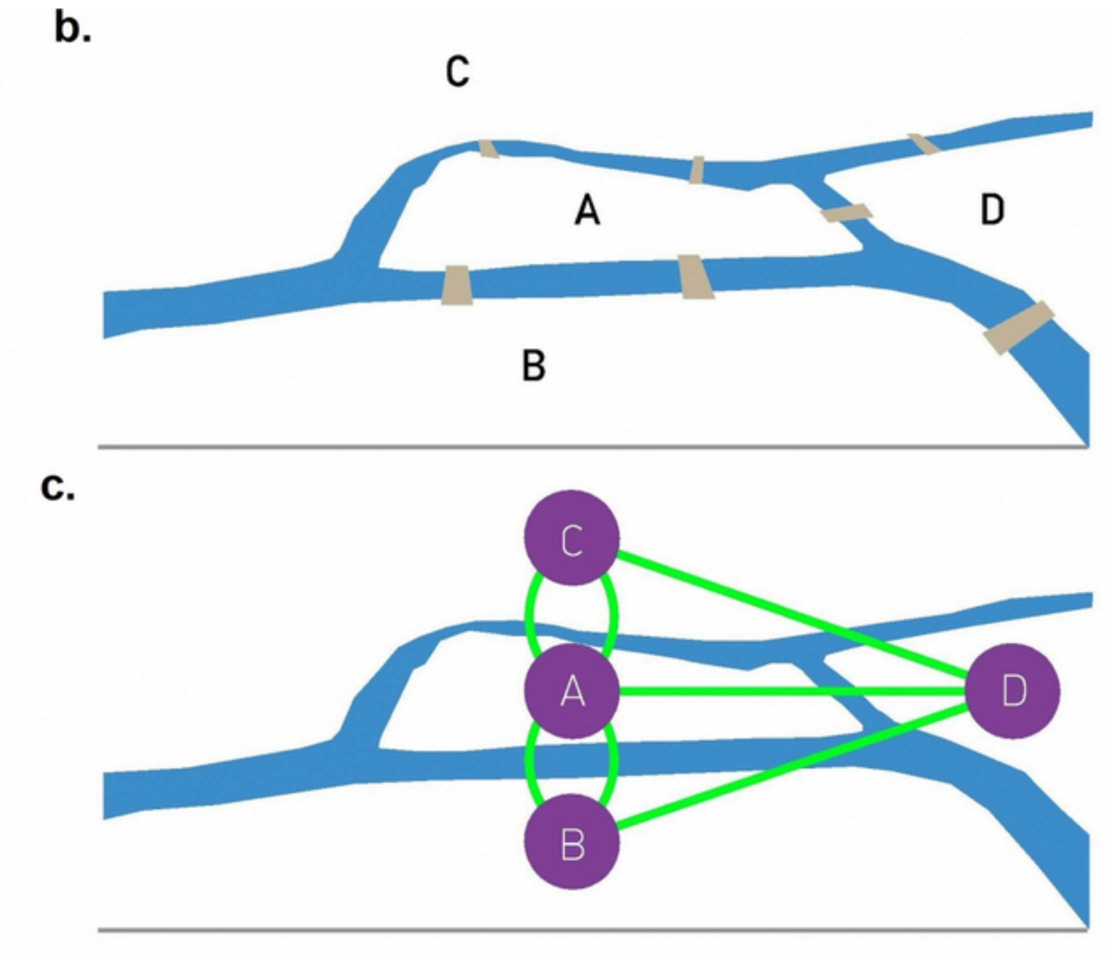
Seven bridges across the river Pregel (now in Russia).

Five of these bridges are connected to the mainland.

The remaining two crossed the two branches of the river.

The peculiar arrangement gave birth to the question: Can one walk across all seven bridges and never cross the same one twice (Euler Path)?

Problem remained unsolved until 1735, when Leonard Euler offered a rigorous mathematical proof that such path does not exist.



The Seven Bridges of Königsberg

Euler Path / Euler Circuit

Euler Path: Uses every edge of a graph exactly once, and also starts and ends at different vertices.

Euler Circuit: Uses every edge of a graph exactly once, and also starts and ends at the same vertex.

Number of Odd Vertices	Implication
0	There is at least one Euler Circuit
1	This is Impossible
2	There is no Euler Circuit but at least 1 Euler Path
More than 2	There are no Euler Circuits or Euler Paths

Implications

Euler showed that solutions to Graphs depend upon their inherent characteristics and not necessarily upon human wit.

Networks or Graphs?

Used interchangeably often.

They are different terms.

Different networks can have same underlying graph, i.e. mathematical representation.

Network Science	Graph Theory
Network	Graph
Node	Vertex
Link	Edge

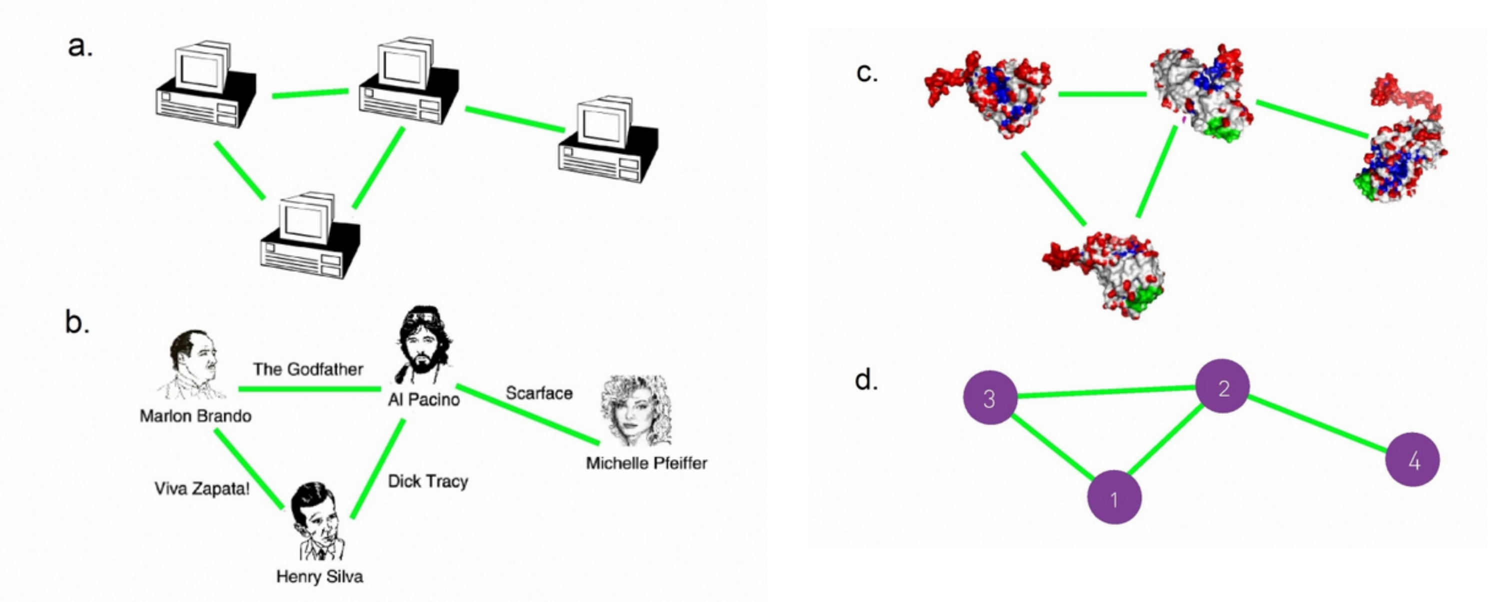
Different Networks Same Graph

Internet is a network of routers and cables.

WWW is a network of web pages and urls.

Several different networks can actually have the same underlying mathematical representation, i.e. graph.

In the figure here, networks **a**, **b**, and **c** have same underlying graph → **d**



Ten Reference Networks

Network	Nodes	Links	Directed / Undirected	N	L	‹K›
Internet	Routers	Internet connections	Undirected	192,244	609,066	6.34
WWW	Webpages	Links	Directed	325,729	1,497,134	4.60
Power Grid	Power plants, transformers	Cables	Undirected	4,941	6,594	2.67
Mobile-Phone Calls	Subscribers	Calls	[illegible]	36,595	91,826	2.51
Email	Email addresses	Emails	Directed	[illegible]	103,731	1.81
Science Collaboration	Scientists	Co-authorships	Undirected	23,1[illegible]	[illegible]	[illegible]
Actor Network	Actors	Co-acting	Undirected	702,388	29,397,908	83.71
Citation Network	Papers	Citations	Directed	449,673	4,689,479	10.43
E. Coli Metabolism	Metabolites	Chemical reactions	Directed	1,039	5,802	5.58
Protein Interactions	Proteins	Binding interactions	Undirected	2,018	2,930	2.90

Table 2.1

Canonical Network Maps

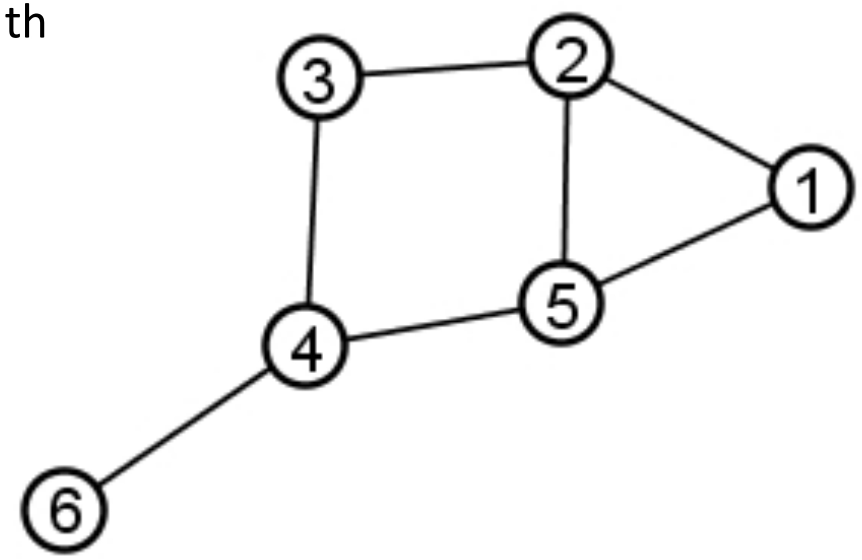
The basic characteristics of ten networks used throughout this book to illustrate the tools of network science. The table lists the nature of their nodes and links, indicating if links are directed or undirected, the number of nodes (N) and links (L), and the average degree for each network. For directed networks the average degree shown is the average in- or out-degrees $\langle k \rangle = \langle k_{in} \rangle = \langle k_{out} \rangle$ (see Equation (2.5)).

Routers are the nodes for Internet

Webpages are the nodes for WWW

Degree

Degree – we denote with k_i the degree of the i^{th} node in the network



- k_1 = Degree for node 1 = 2
- k_2 = Degree for node 2 = 3
- K_3 = Degree for node 3 = 2
- k_4 = Degree for node 4 = 3
- k_5 = Degree for node 5 = 3
- k_6 = Degree for node 6 = 1

Adding all the degrees we get 2+3+2+3+3+1 = 14

- However, we can see that there are only 7 total links and not 14
- Total number of links = ½ total number of degrees for all the nodes

Thus, for undirected graph total number of links (L) = $\frac{1}{2}\sum_{i=1}^{N} k_i$

Average Degree

Average degree <k> = total number of degrees for all the nodes / total number of nodes = 2 x total number of links / total number of nodes:

$$\langle k \rangle = \frac{1}{N}\sum_{i=1}^{N} k_i = \frac{2L}{N}$$

For directed graphs:

$$k_i = (k_i^{in} + k_i^{out})$$

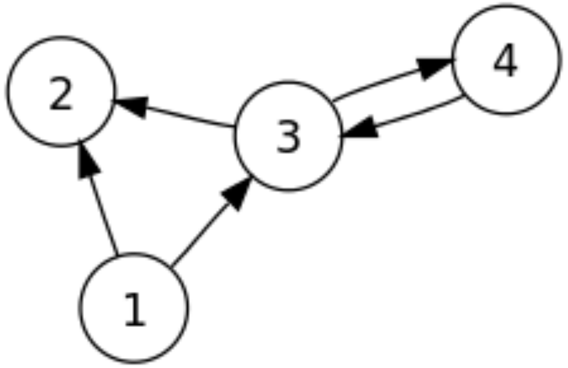
Total number of links for directed graphs:

$$L = \sum_{i=1}^{N} k_i^{in} = \sum_{i=1}^{N} k_i^{out}$$

Average degree of directed network is:

$$\langle k^{in} \rangle = \frac{1}{N}\sum_{i=1}^{N} k_i^{in} = \langle k^{out} \rangle = \frac{1}{N}\sum_{i=1}^{N} k_i^{out} = \frac{L}{N}$$

Directed Graph



Node 1 has two out-degrees only; In other words, $k_1^{in} = 0, and\ k_1^{out} = 2$

Node 2 has two in-degrees only; In other words, $k_2^{in} = 2, and\ k_2^{out} = 0$

Node 3 has two in-degrees and two out-degrees; In other words, $k_3^{in} = 2$, and $k_3^{out} = 2$

Node 4 has one in-degree and one out-degree; In other words, $k_4^{in} = 1, and\ k_4^{out} = 1$

Total number of in-degree links = 0 + 2 + 2 + 1 = 5; and

Total number of out-degree links = 2 + 0 + 2 + 1 = 5

Also, total number of links L = k^{in} = k^{out} = 5

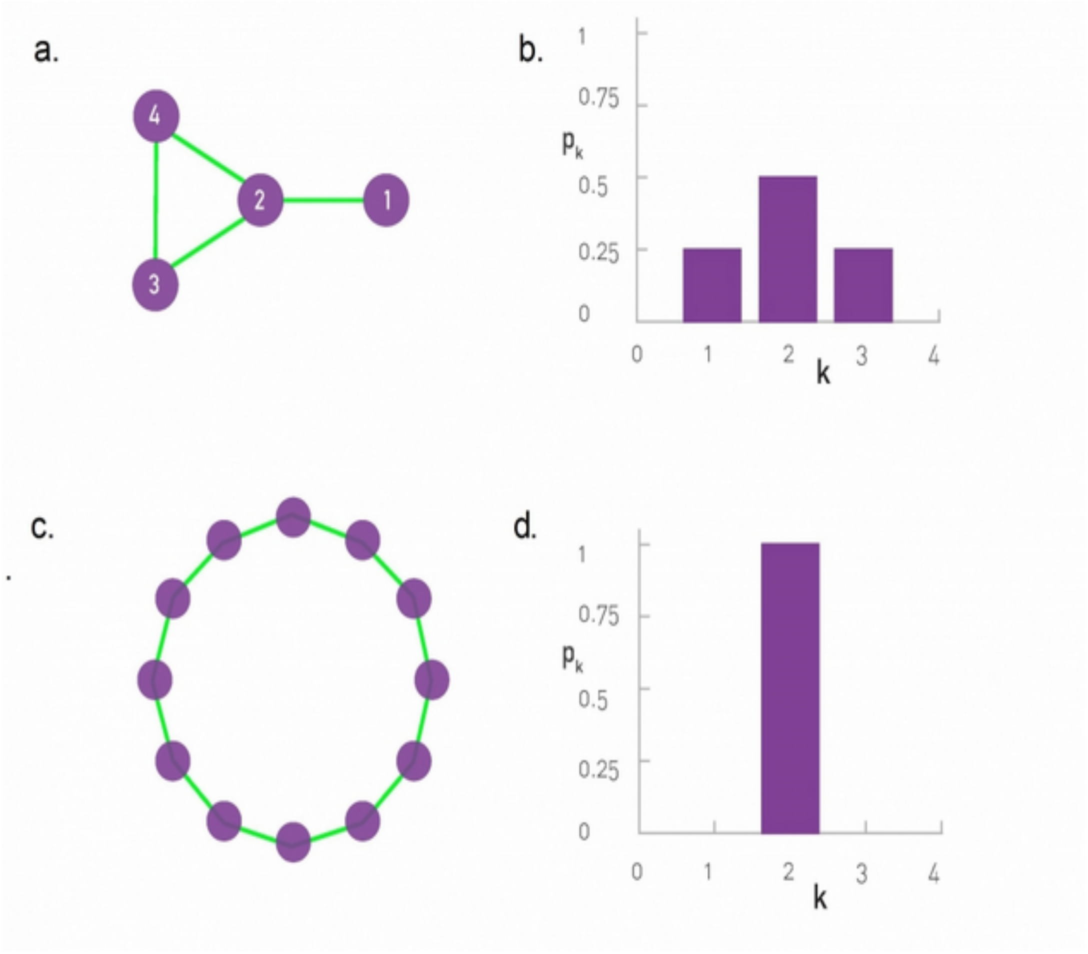
Average in-degree = $\langle k^{in} \rangle$ = 5/4 = 1.25 = Average out-degree = $\langle k^{out} \rangle$ = 5/4=1.25

Degree Distribution

Degree distribution of a network is shown with number of links per node on the x-axis, and the corresponding probability for those links on the y-axis.

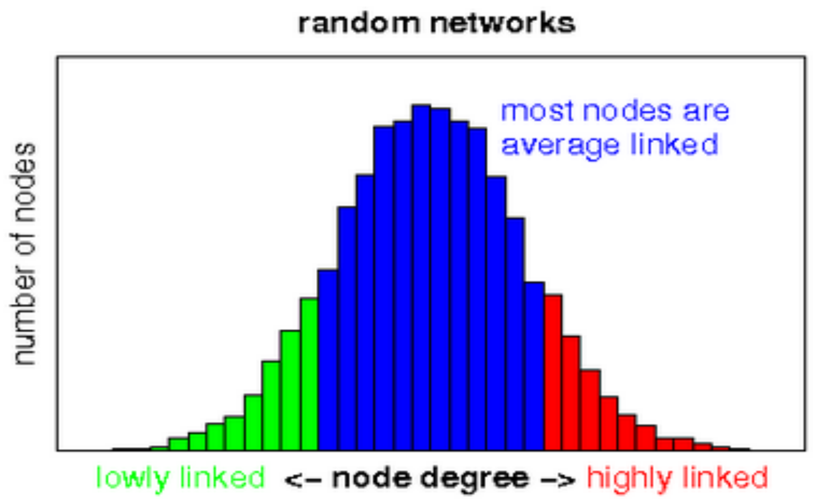
In figure **a**, probability of having one link per node is 1 (out of total 4 links), probability of having two links per node is 3 (out of total 4 links), and probability of having three links per node is 1 (out of total 4 links).

In figure **c**, probability of having number of links different from 2 per node is 0 (every node has exactly 2 links).

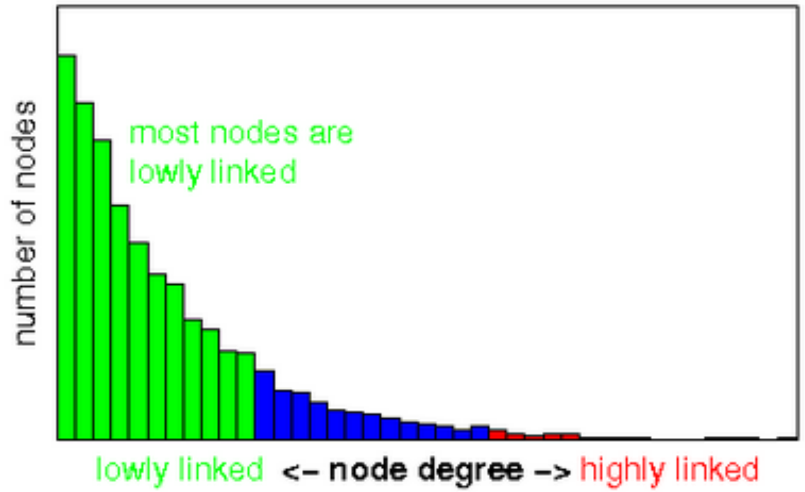


Degree Distribution of a Random Network

Random networks have some nodes that have fewer connections, and some nodes that are highly connected. Most of the links are average linked.



Internet, on the other hand, has only very few webpages (nodes) that are highly popular and well linked such as Google, and Facebook, and remaining webpages (nodes) have very few connections (links).

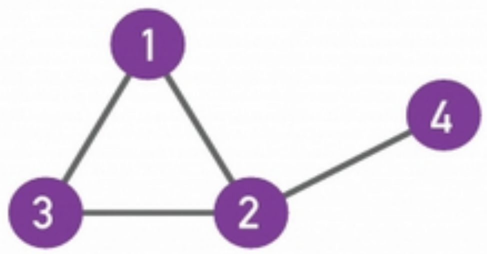


Images source: http://www.network-science.org/powerlaw_scalefree_node_degree_distribution.html

a. **Adjacency matrix**

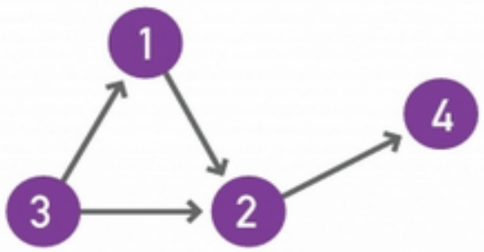
$$A_{ij} = \begin{matrix} A_{11} & A_{12} & A_{13} & A_{14} \\ A_{21} & A_{22} & A_{23} & A_{24} \\ A_{31} & A_{32} & A_{33} & A_{34} \\ A_{41} & A_{42} & A_{43} & A_{44} \end{matrix}$$

b. **Undirected network**



$$A_{ij} = \begin{matrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{matrix}$$

c. **Directed network**



$$A_{ij} = \begin{matrix} 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{matrix}$$

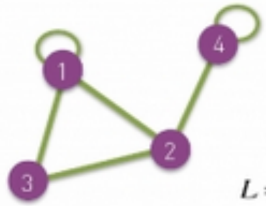
a. Undirected

$$A_{ij} = \begin{pmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$$

$A_{ii} = 0 \qquad A_{ij} = A_{ji}$

$L = \frac{1}{2}\sum_{i,j=1}^{N} A_{ij} \qquad <k> = \frac{2L}{N}$

b. Self-loops

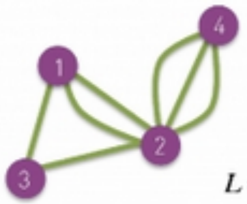


$$A_{ij} = \begin{pmatrix} 1 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix}$$

$\exists i, A_{ii} \neq 0 \qquad A_{ij} = A_{ji}$

$L = \frac{1}{2}\sum_{i,j=1, i\neq j}^{N} A_{ij} + \sum_{i=1}^{N} A_{ii} \qquad ?$

c. Multigraph (undirected)

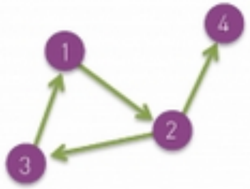


$$A_{ij} = \begin{pmatrix} 0 & 2 & 1 & 0 \\ 2 & 0 & 1 & 3 \\ 1 & 1 & 0 & 0 \\ 0 & 3 & 0 & 0 \end{pmatrix}$$

$A_{ii} = 0 \qquad A_{ij} = A_{ji}$

$L = \frac{1}{2}\sum_{i,j=1}^{N} A_{ij} \qquad <k> = \frac{2L}{N}$

d. Directed

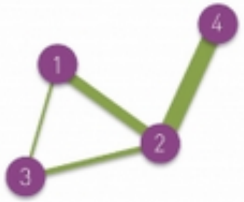


$$A_{ij} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

$A_{ij} \neq A_{ji}$

$L = \sum_{i,j=1}^{N} A_{ij} \qquad <k> = \frac{L}{N}$

e. Weighted (undirected)



$$A_{ij} = \begin{pmatrix} 0 & 2 & 0.5 & 0 \\ 2 & 0 & 1 & 4 \\ 0.5 & 1 & 0 & 0 \\ 0 & 4 & 0 & 0 \end{pmatrix}$$

$A_{ii} = 0 \qquad A_{ij} = A_{ji}$

$<k> = \frac{2L}{N}$

f. Complete Graph (undirected)

$$A_{ij} = \begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix}$$

$A_{ii} = 0 \qquad A_{i\neq j} = 1$

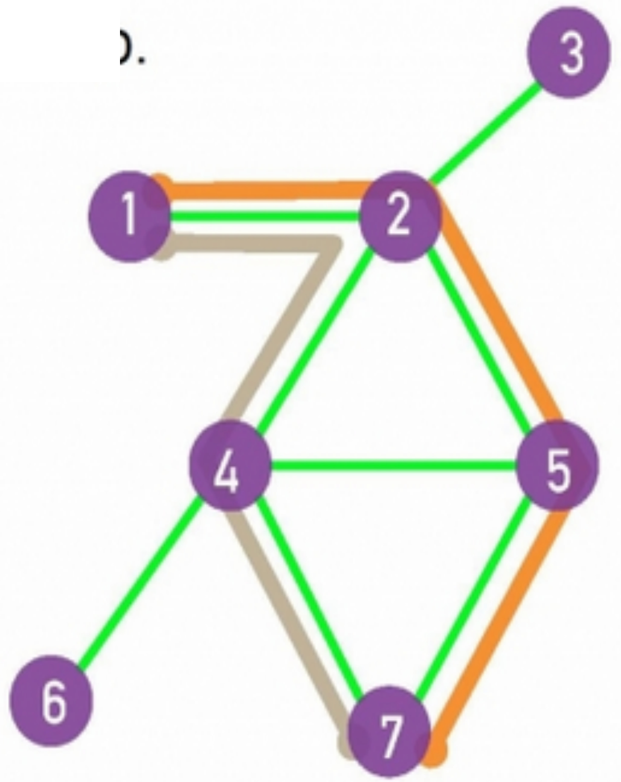
$L = L_{max} = \frac{N(N-1)}{2} \qquad <k> = N-1$

Paths, Distance and Diameter

Path between nodes 1 to 7 is given as 1->2->5->7, hence path length = 3

Distance is the shorted paths between two nodes. For instance, distance between nodes 1 and 5 i.e. d_{15} is equal to 2

Network diameter is the largest distance in the network, d_{max} = 3 here (between nodes 1 and 7, 3 and 6, 3 and 7, etc.)



Clustering Coefficient

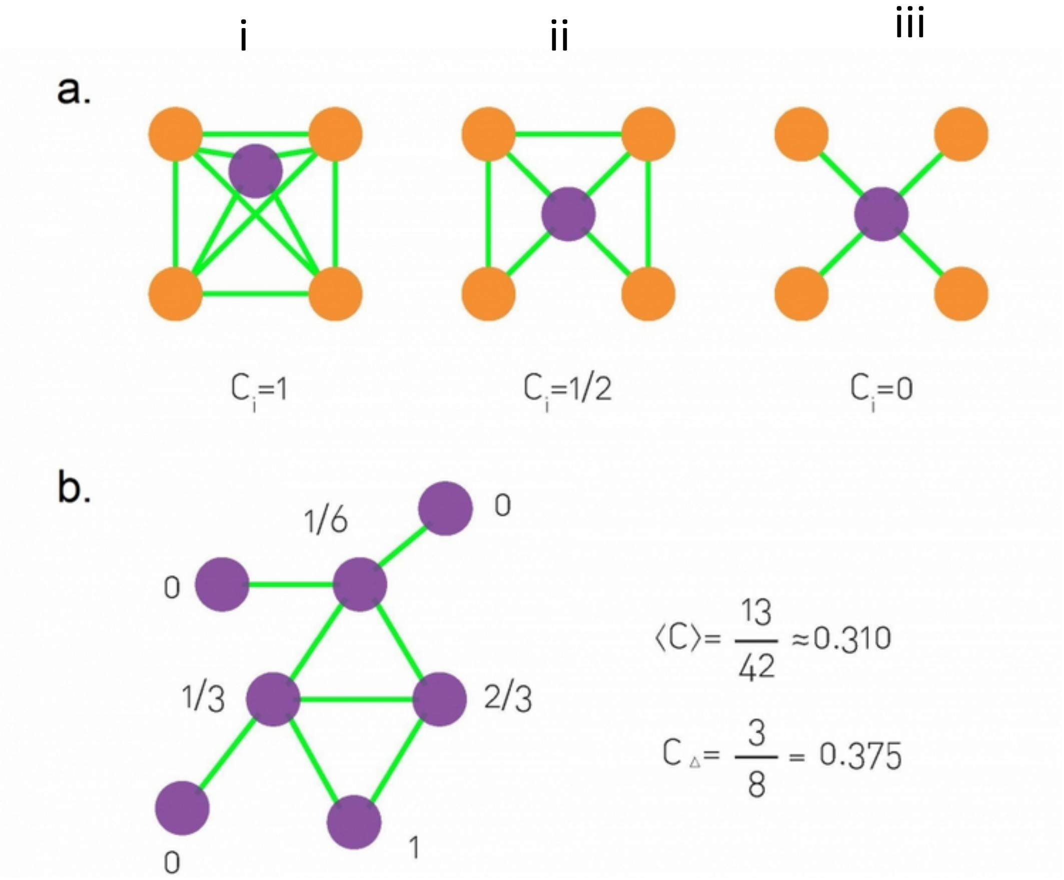
Captures the degree to which the neighbors of a given node link to each other. For a node k_i, the local clustering coefficient is given as:

$$C_i = \frac{2L_i}{k_i(k_i - 1)}$$

C_i = 0 if none of the neighbors of node I link to each other.

C_i = 1 if the neighbors of node I form a complete graph, i.e. they all link to each other.

In figure a(ii), purple node has 4 links i.e. K4 = 4, and its 4 neighbors have 3 links among themselves i.e. L4 = 3, thus C_i = (2 x 3) / (4 x (4-1)) = 6/12=1/2



Bipartite Networks

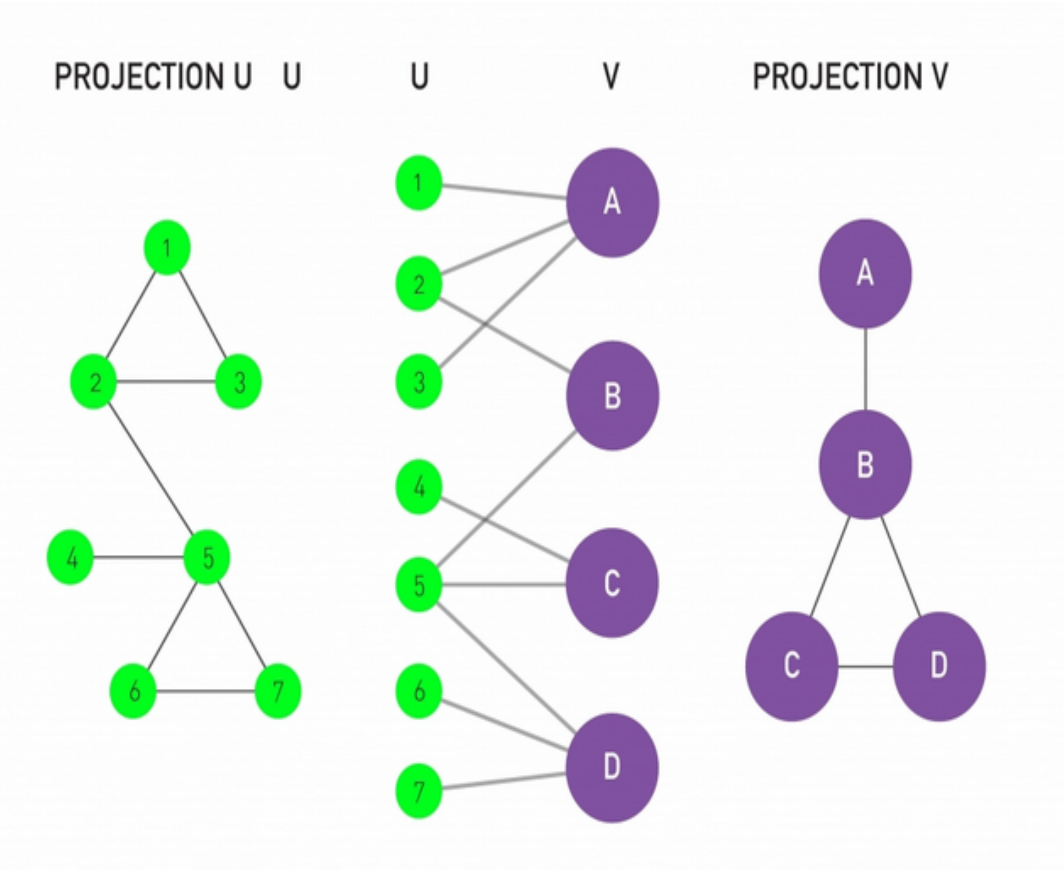
Network of Actors (U) and Movies (V).

Several actors act in a movie.

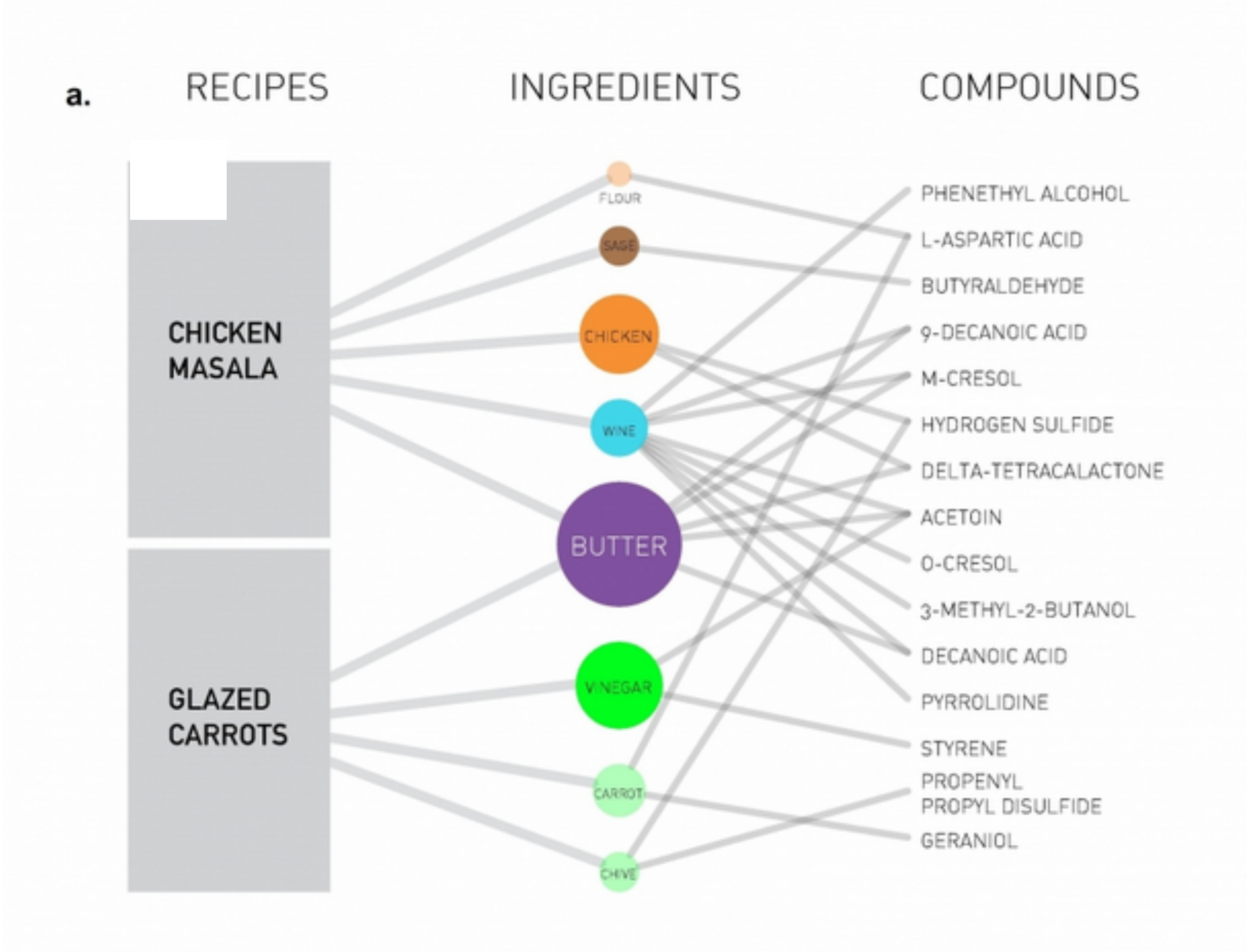
One actor acts in several movies.

Here, projection U, shows network of actors who have worked together.

Here, projection V, shows network of movies that are connected through same actor(s).



Tripartite Networks



Real Networks are Sparse

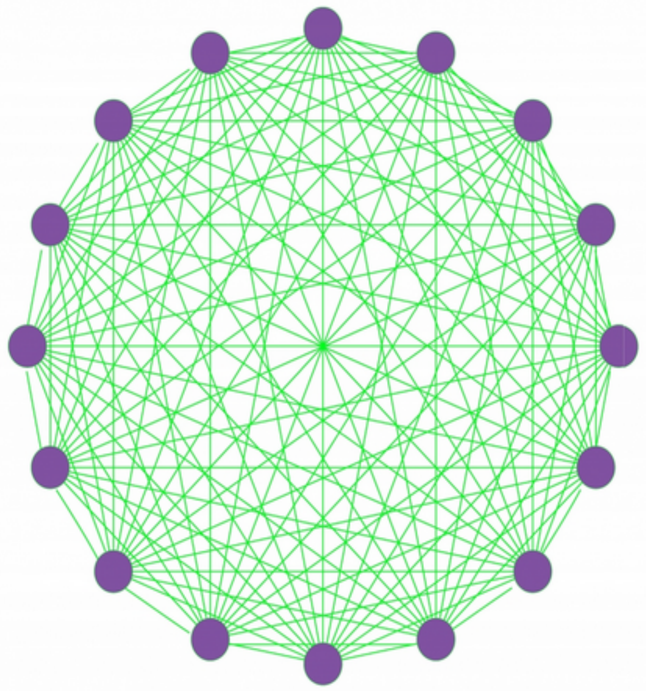
Real networks are not densely linked.

Complete graph (where every node is connected to every other node) has total N/2 links.

$L_{max} = C_2^N = N(N-1)/2$

Network with 16 nodes will have total C_2^N links i.e. 16! / (16-2)!2! = 16 X 15 / 2 = 8 X 15 = 120 links.

Number of links in Real network is << L_{max}



Questions? Thoughts? Visit the course's online discussion forum.