Introduction to Network Science

SICSS'23 - Istanbul

Onur Varol

@onurvarol
www.onurvarol.com
www.varollab.com

Sabancı Üniversitesi

VRLLab

ISTANBUL TECHNICAL UNIVERSITY
1773

EE

Physics

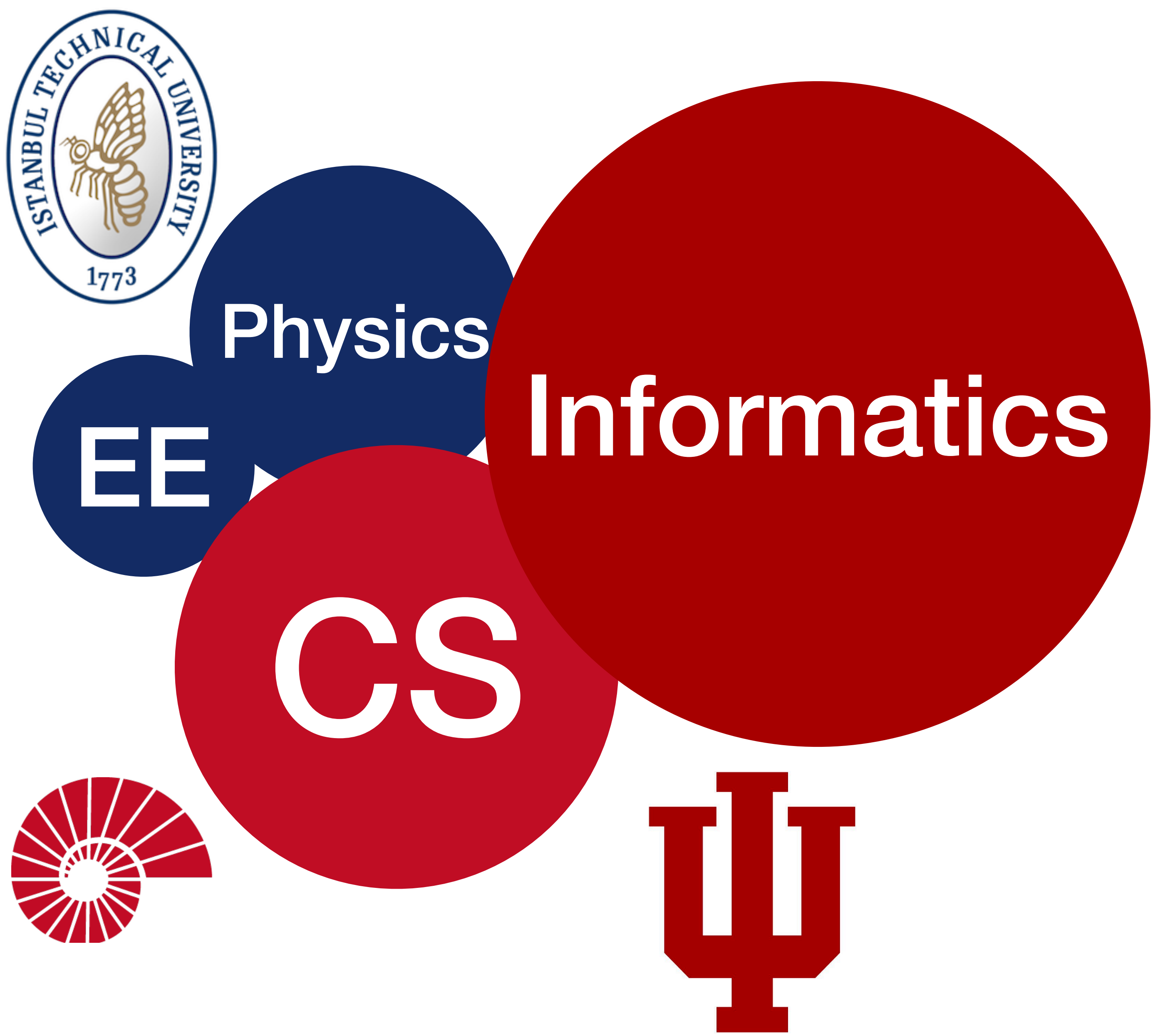
EE

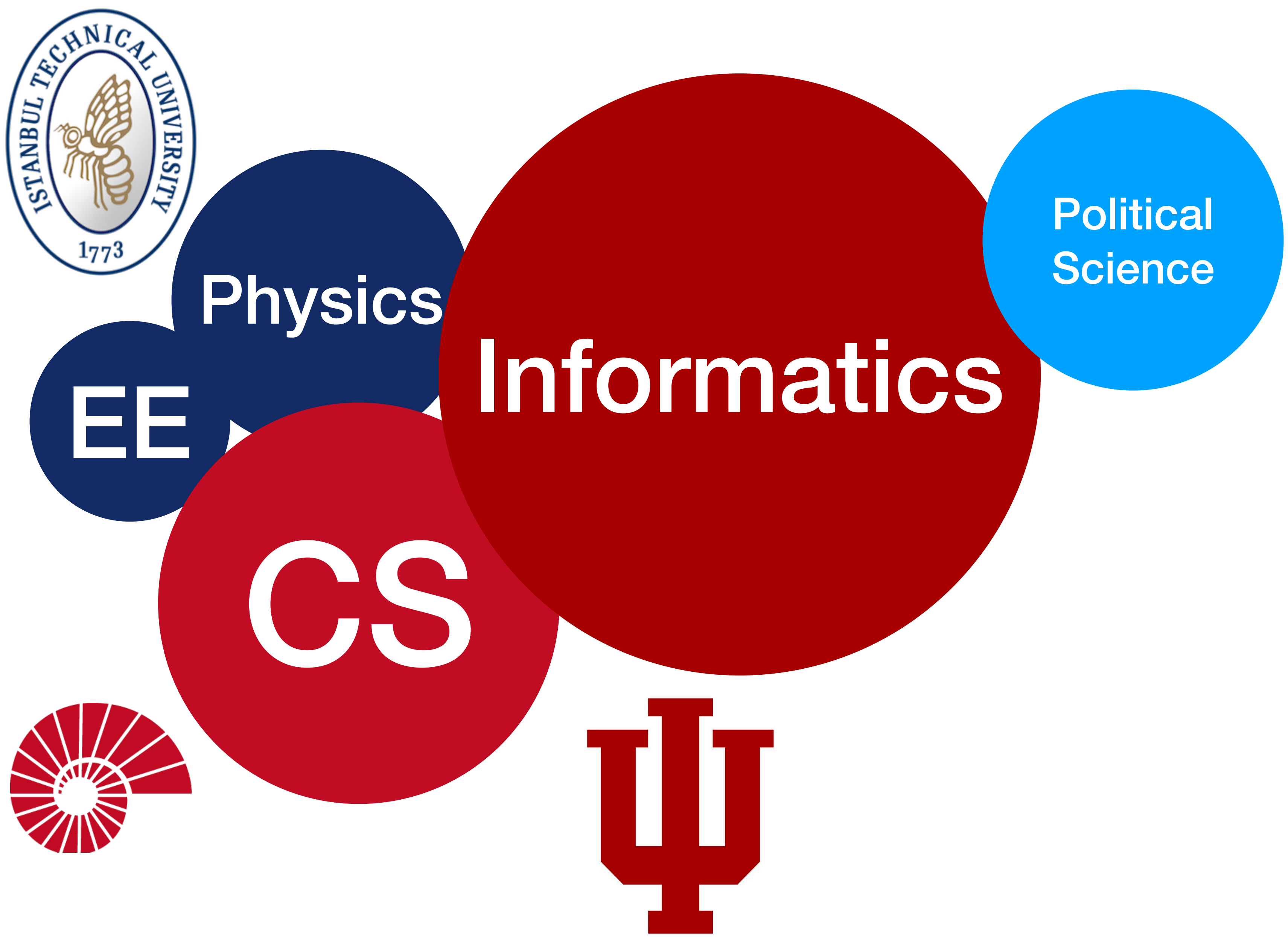
ISTANBUL TECHNICAL UNIVERSITY 1773

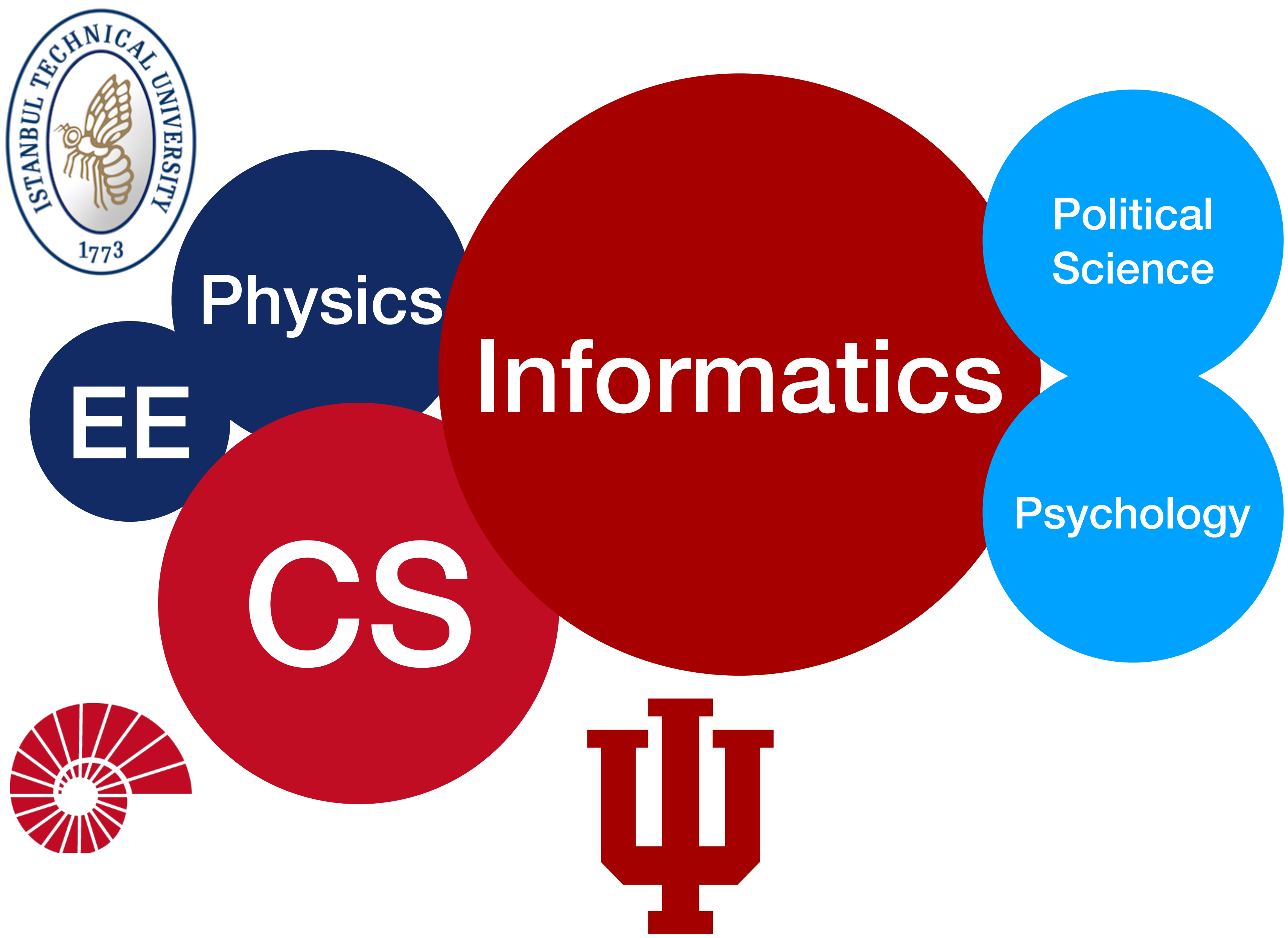
Physics

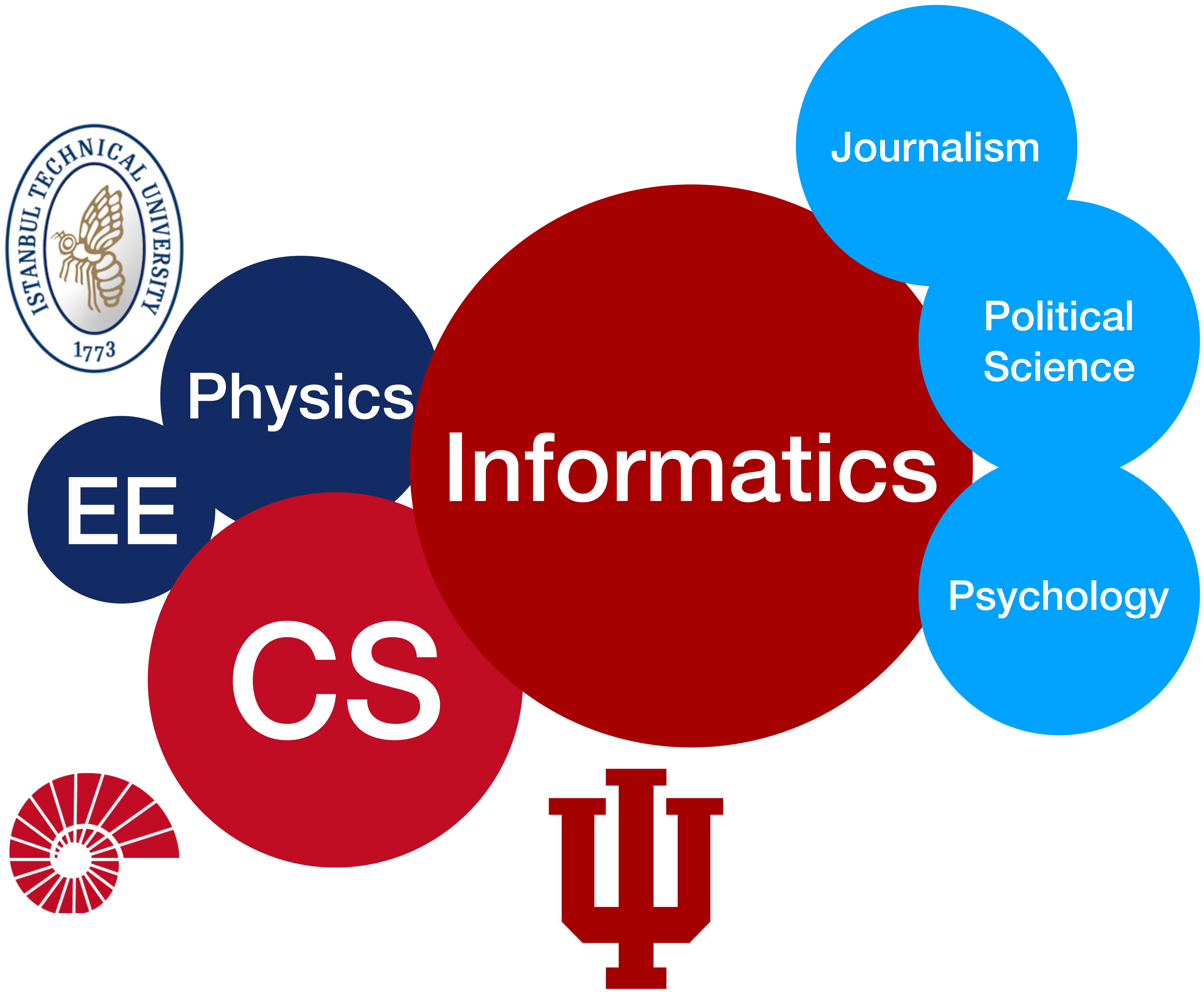
EE

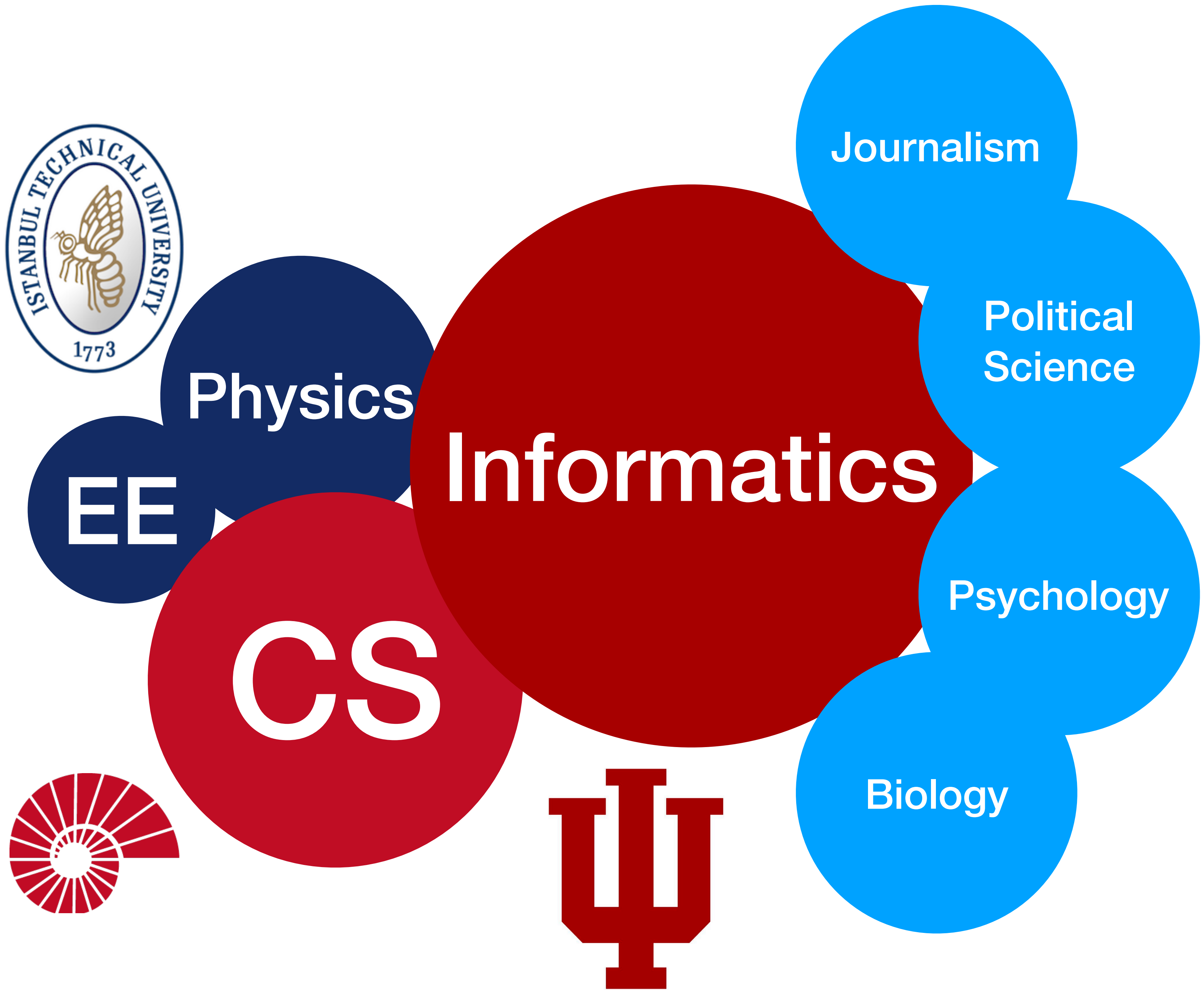
CS


ISTANBUL TECHNICAL UNIVERSITY
1773
Physics
EE
Informatics
CS

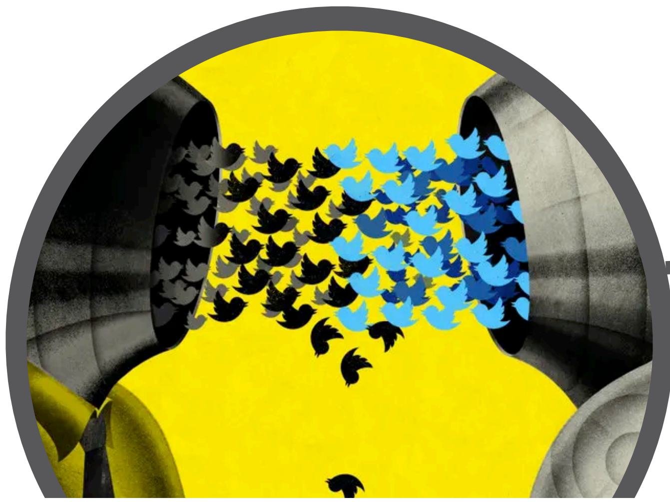

ISTANBUL TECHNICAL UNIVERSITY
1773
Physics
EE
Informatics
Political Science
CS


ISTANBUL TECHNICAL UNIVERSITY
1773
Physics
EE
Informatics
Political Science
CS
Psychology


ISTANBUL TECHNICAL UNIVERSITY
1773
Physics
EE
CS
Informatics
Journalism
Political Science
Psychology


ISTANBUL TECHNICAL UNIVERSITY
1773
Physics
EE
CS
Informatics
Journalism
Political Science
Psychology
Biology

VRLLab Focus Areas



Misinformation & Manipulation

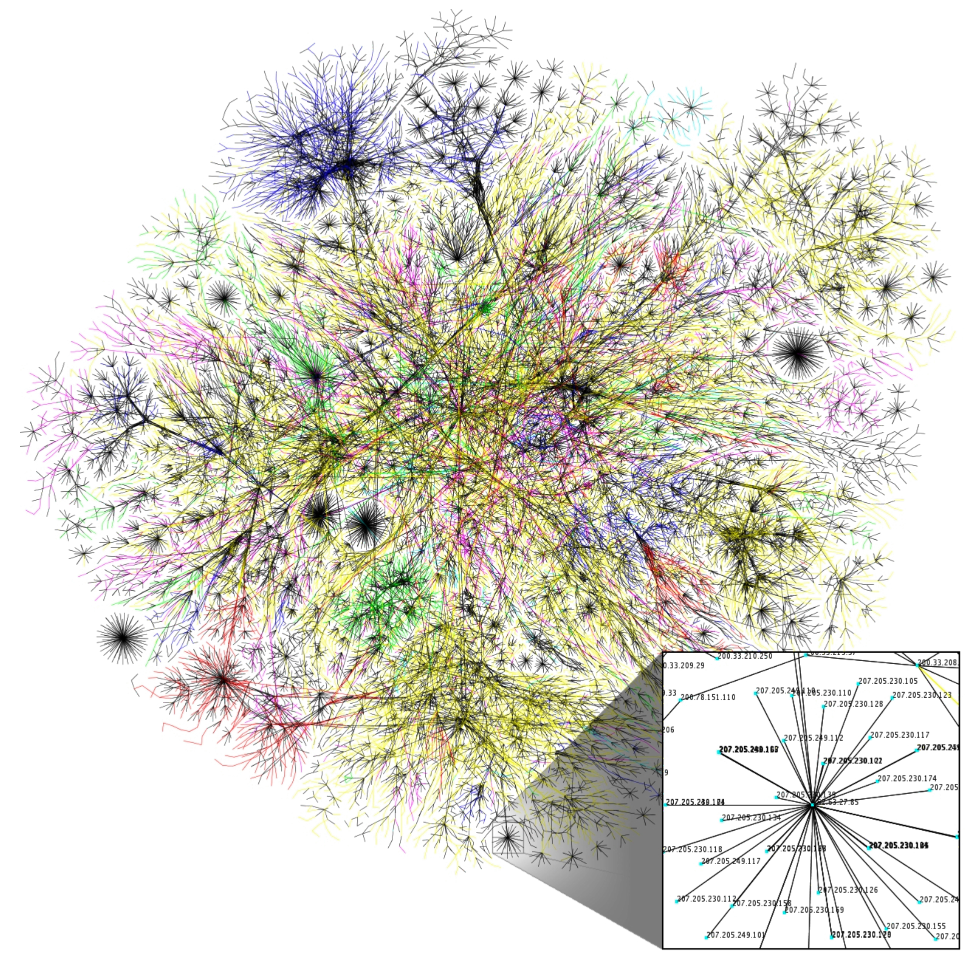
Science of Science

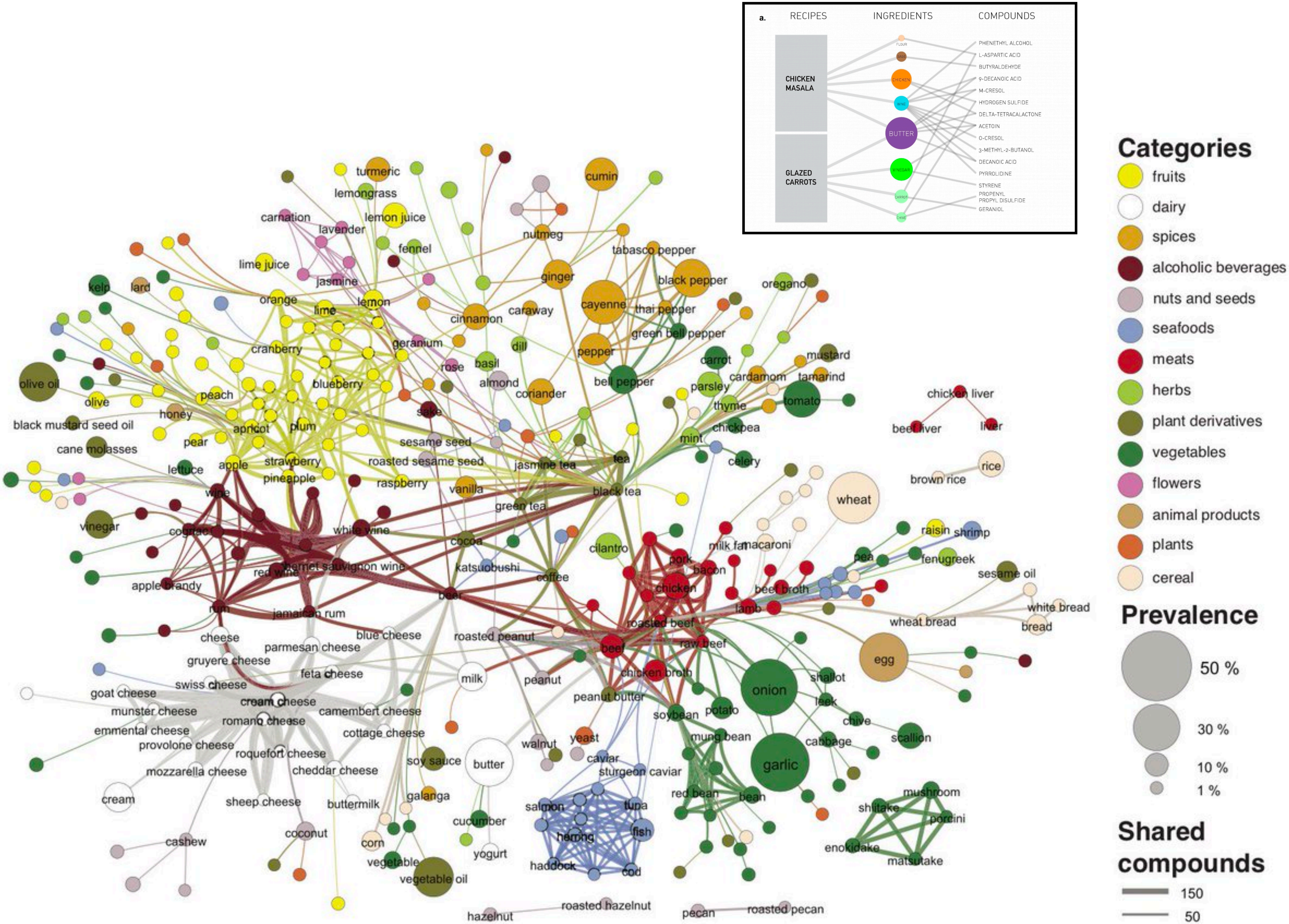
Detection of Social bots

Mental health & Well-being

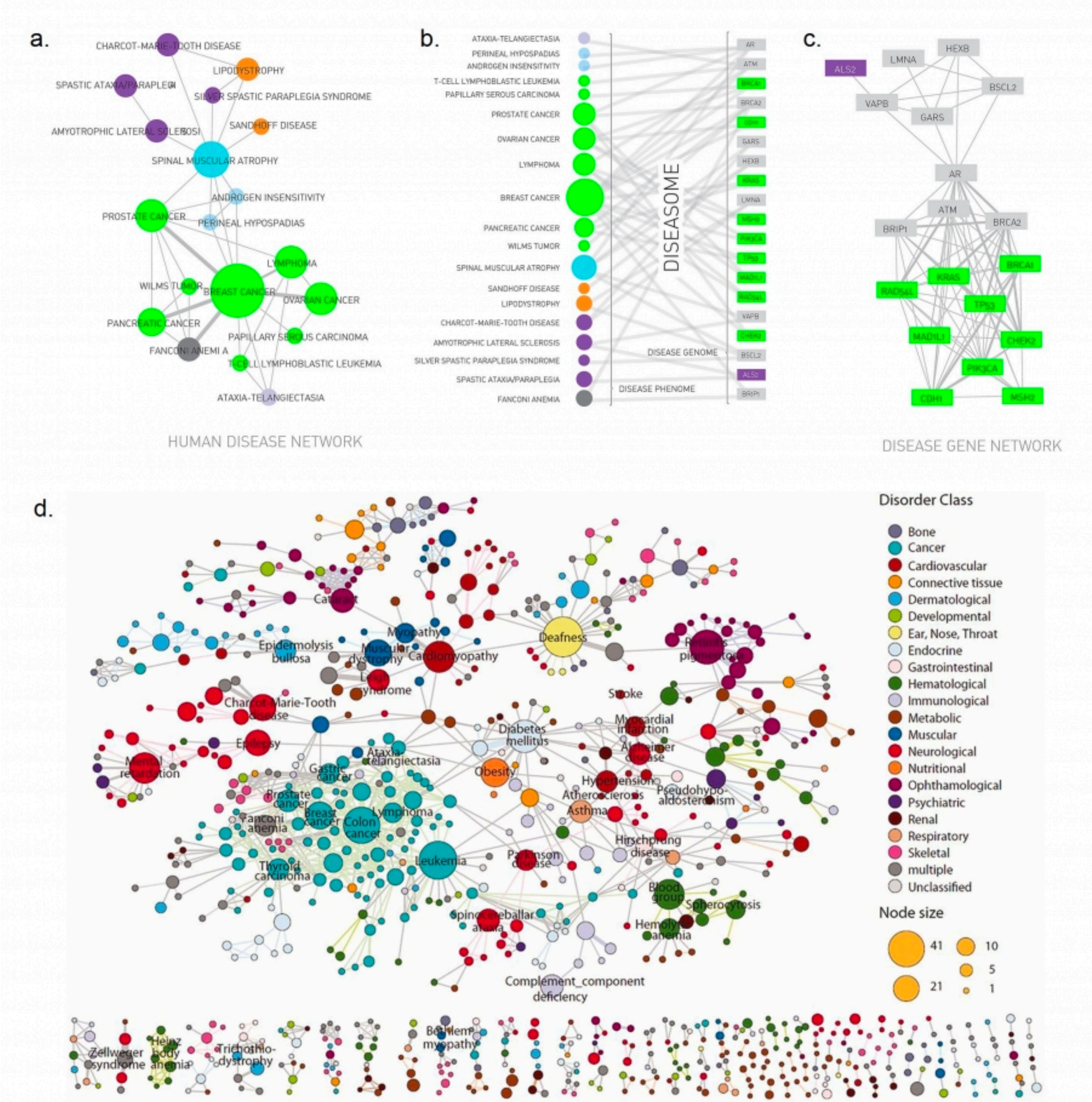
Complex networks are everywhere. Literally!

- Technological networks
  - The Internet
  - Powergrid
- Information networks
  - WWW
  - Wikipedia
- Academic networks
  - Collaborations
  - Citation networks
- Social networks
- Biological networks

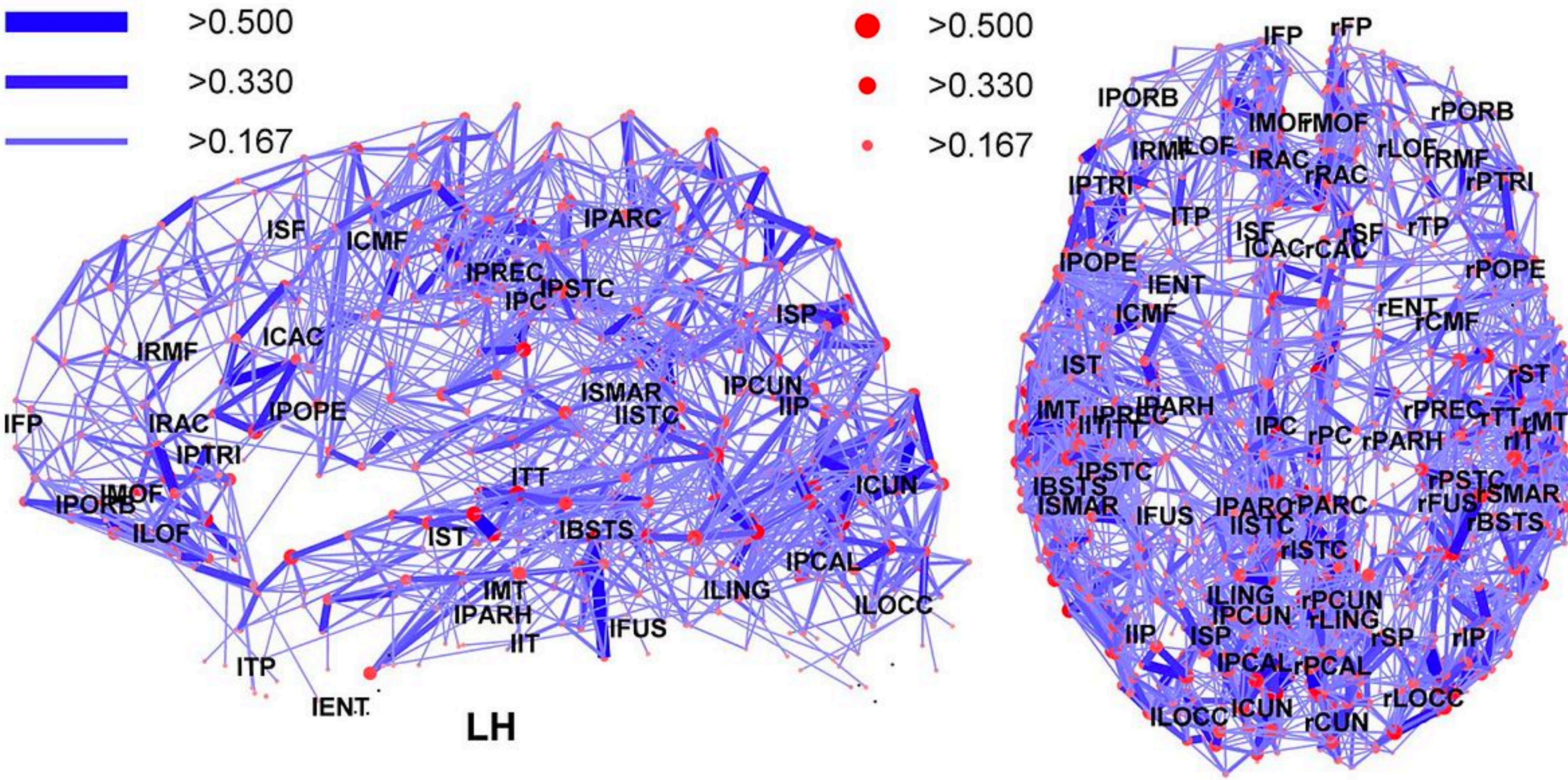




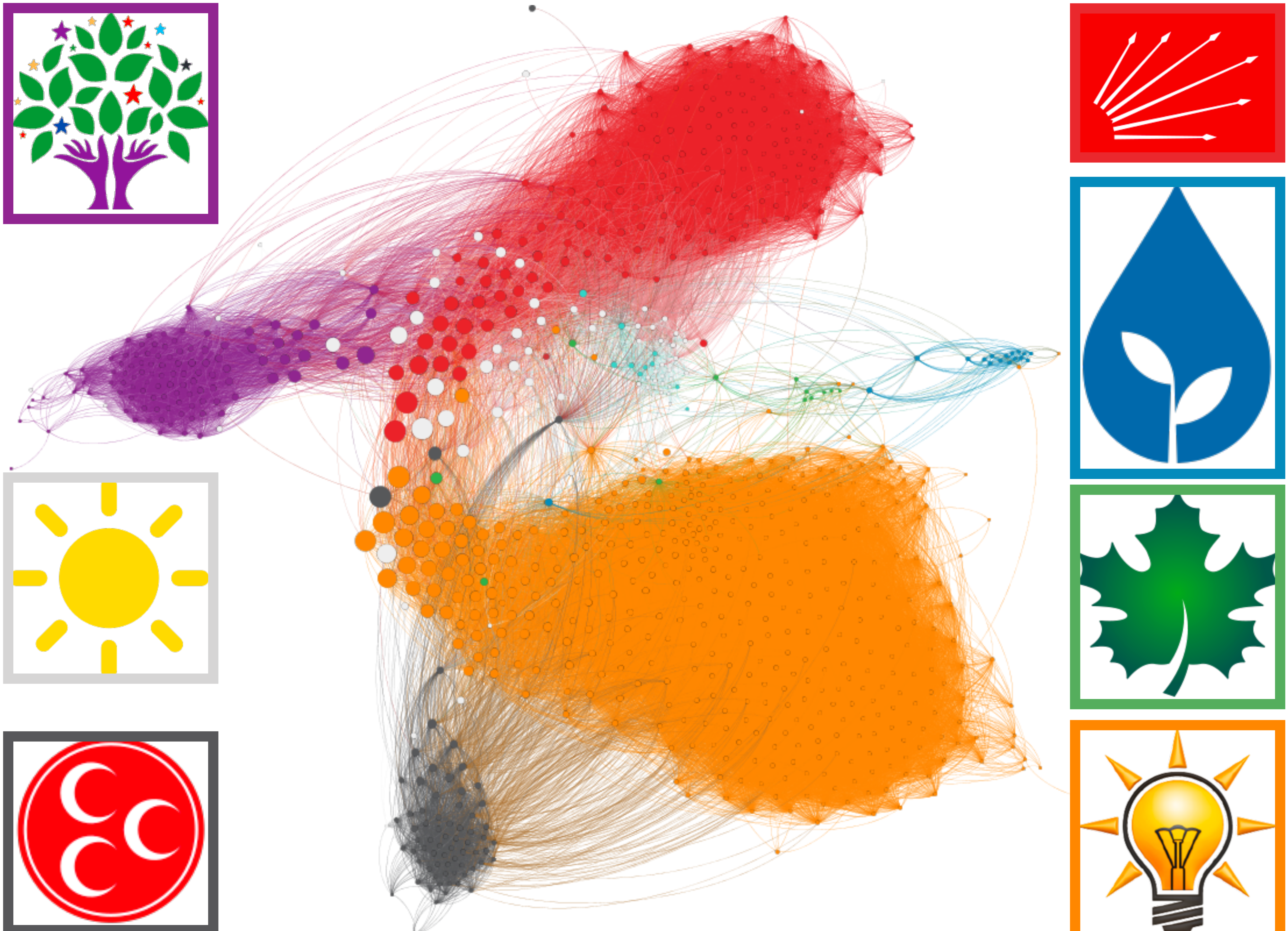
Ahn, Yong-Yeol, et al. "Flavor network and the principles of food pairing." Scientific reports 1 (2011): 196.



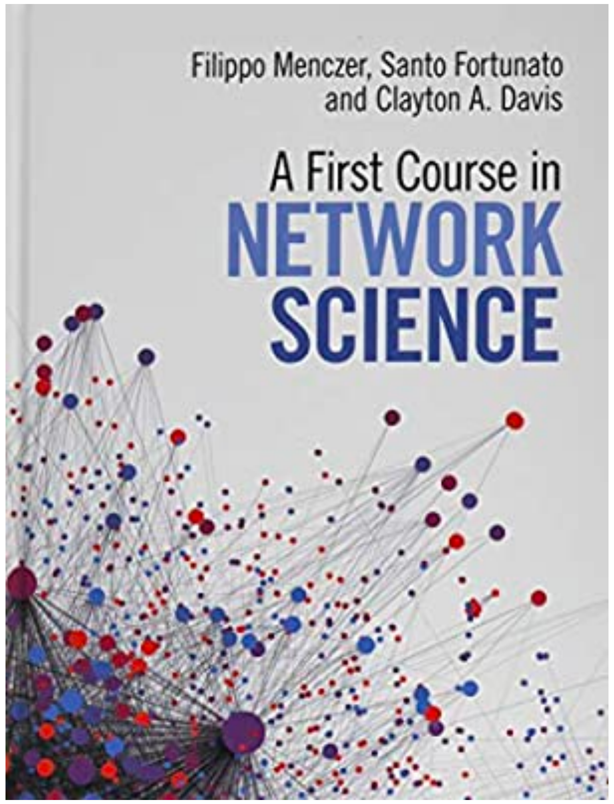
Goh, Kwang-Il, et al. "The human disease network." Proceedings of the National Academy of Sciences 104.21 (2007): 8685-8690.



Hagmann, Patric, et al. "Mapping the structural core of human cerebral cortex." PLoS Biol 6.7 (2008): e159.

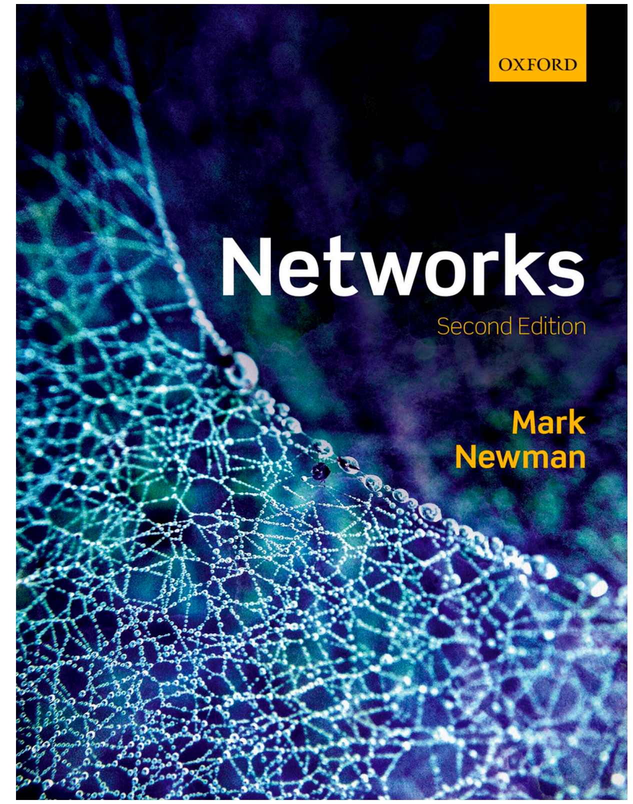


Recommended books on network science

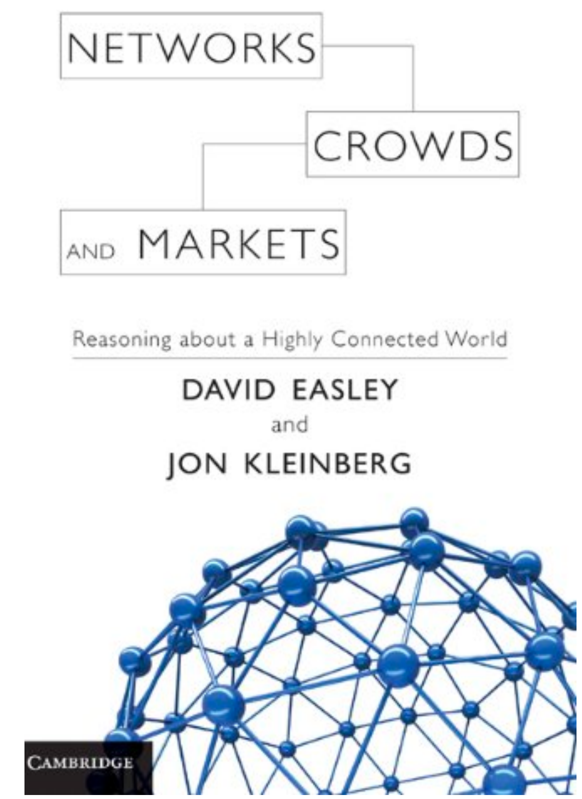


Menczer, Filippo, Santo Fortunato, and Clayton A. Davis.
A First Course in Network Science.
Cambridge University Press, 2020.

Barabasi, Albert-Laszlo.
Network science.
Cambridge university press, 2016.



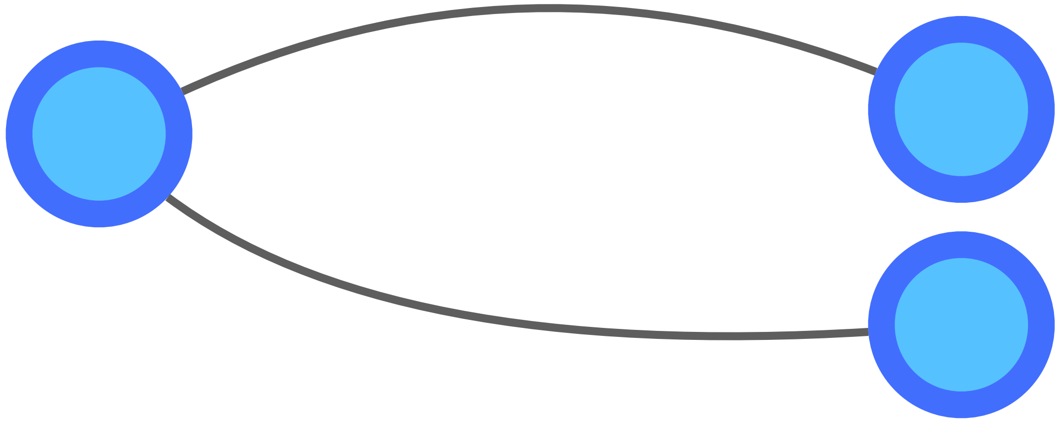
Newman, Mark.
Networks.
Oxford university press, 2018.



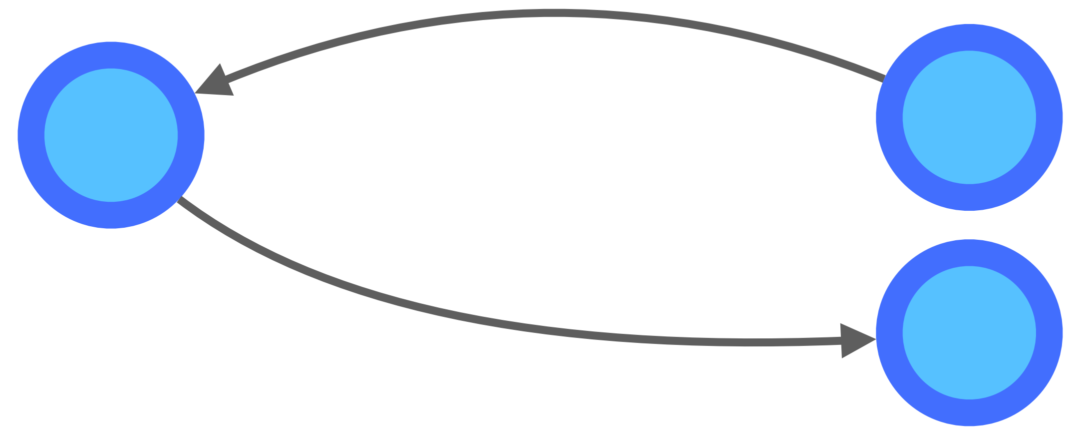
Easley, David, and Jon Kleinberg
Networks, crowds, and markets,
2012

Network types

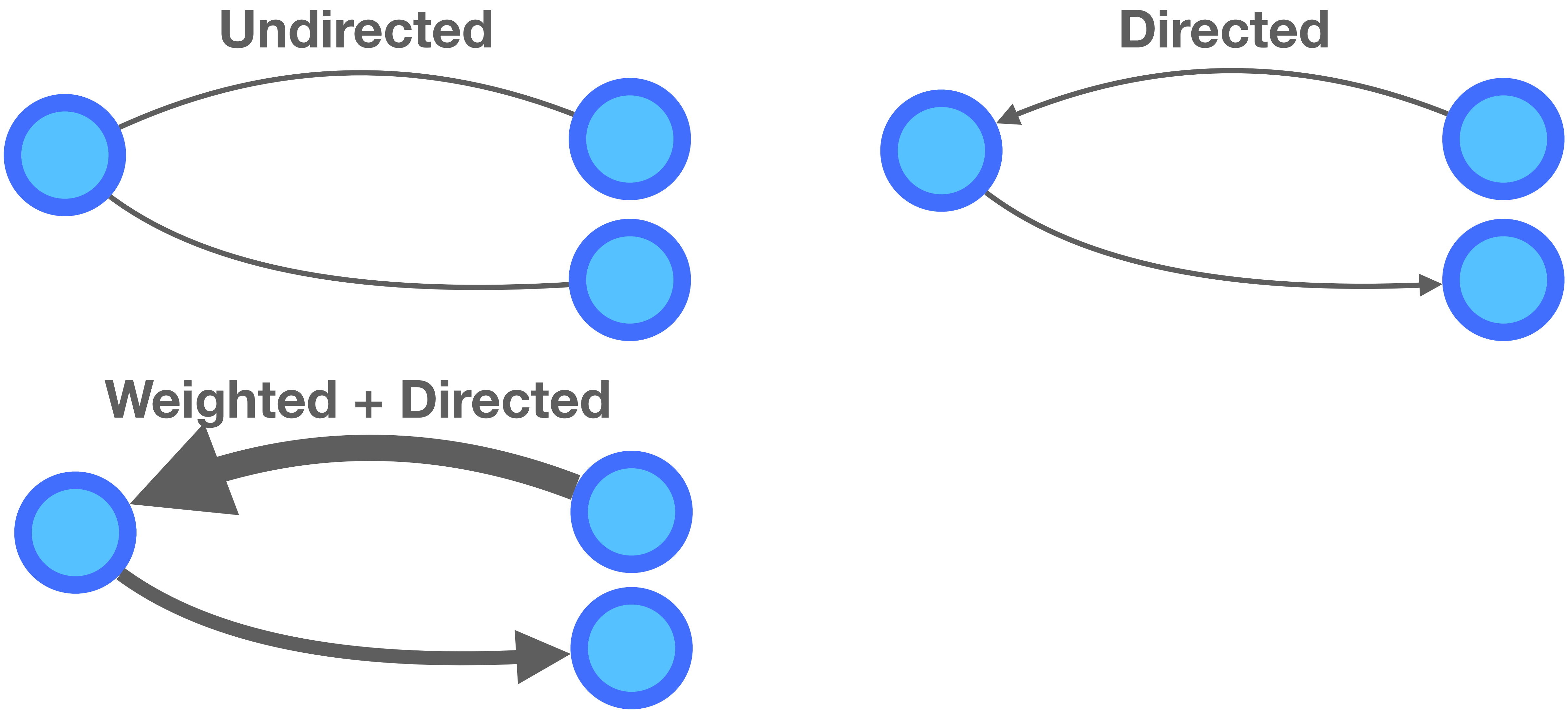
Undirected



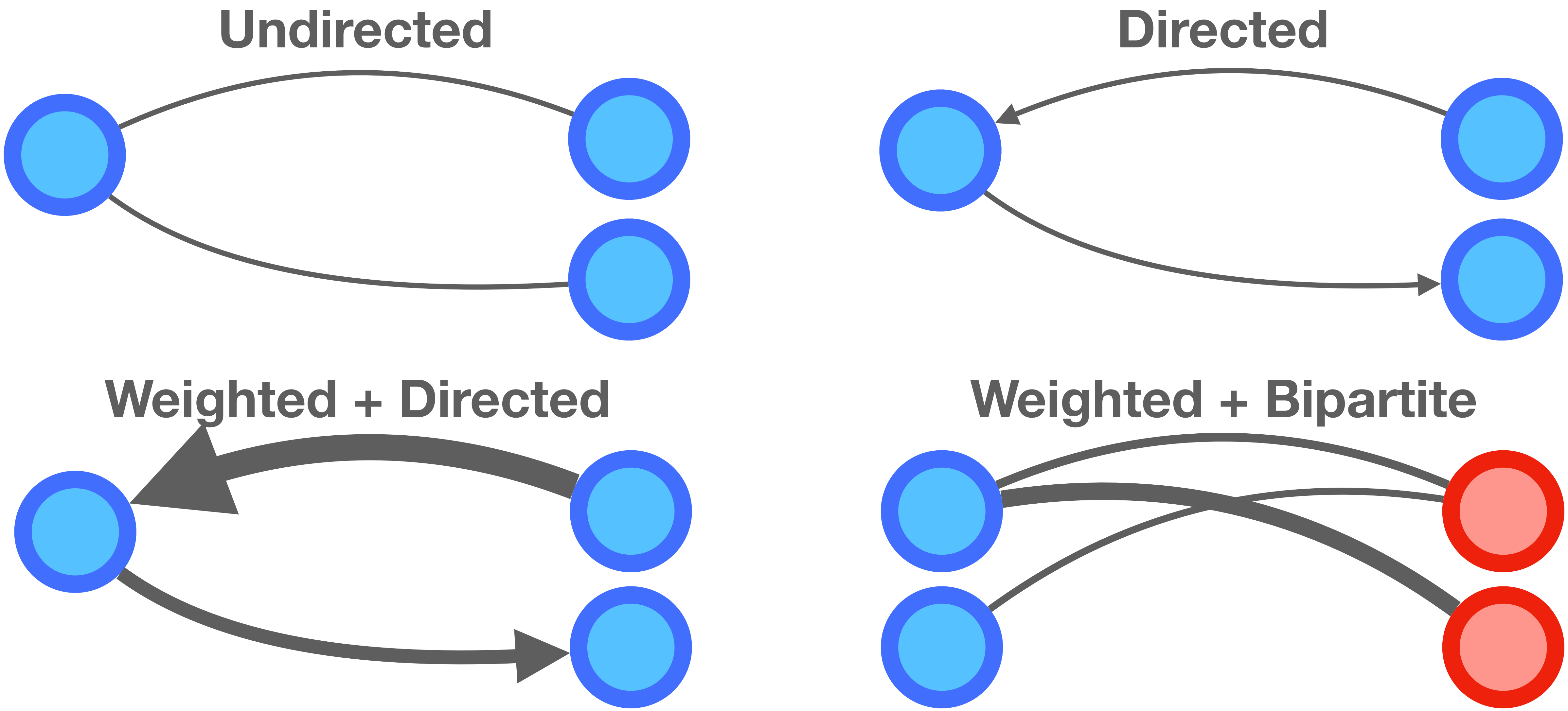
Directed



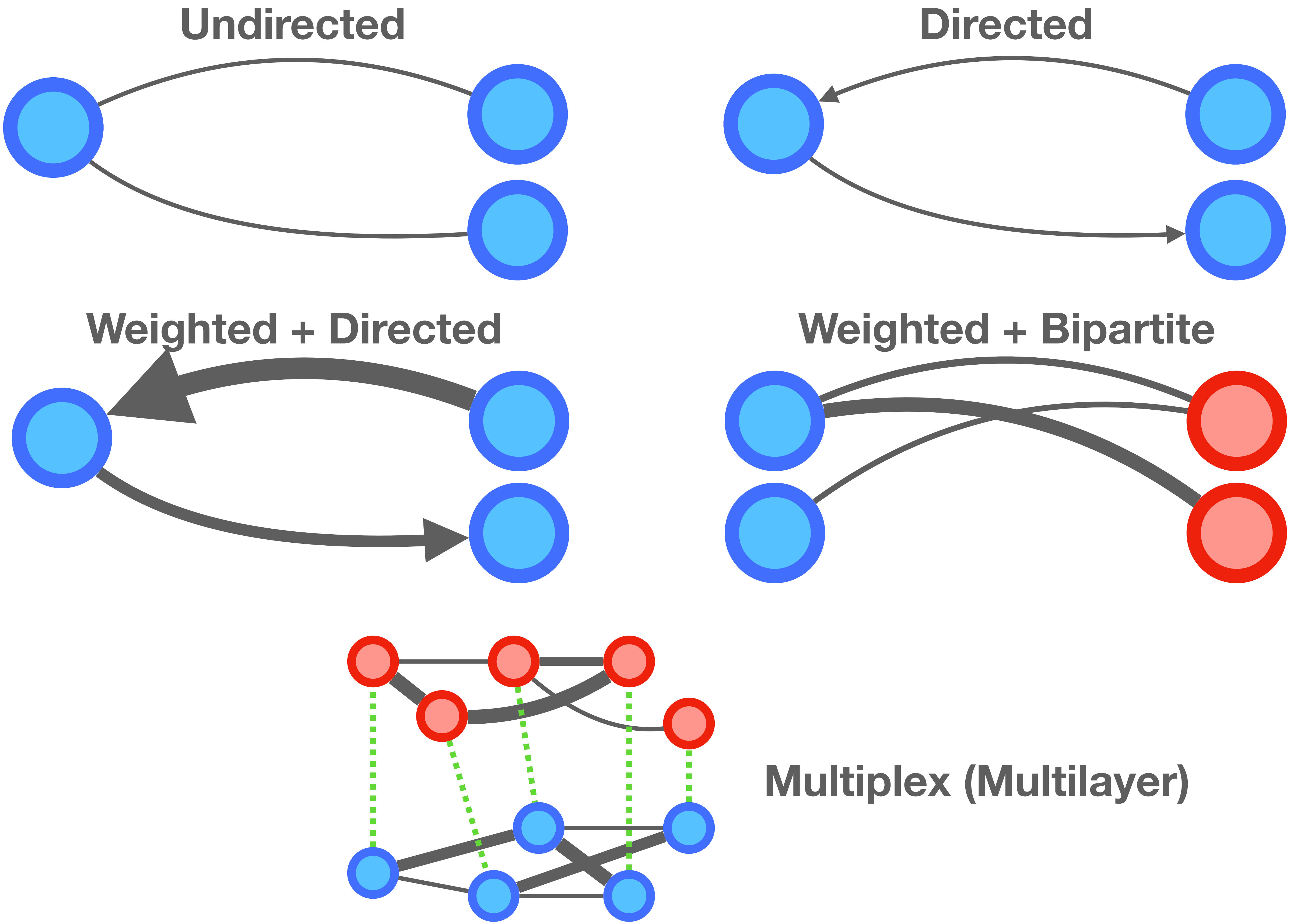
Network types



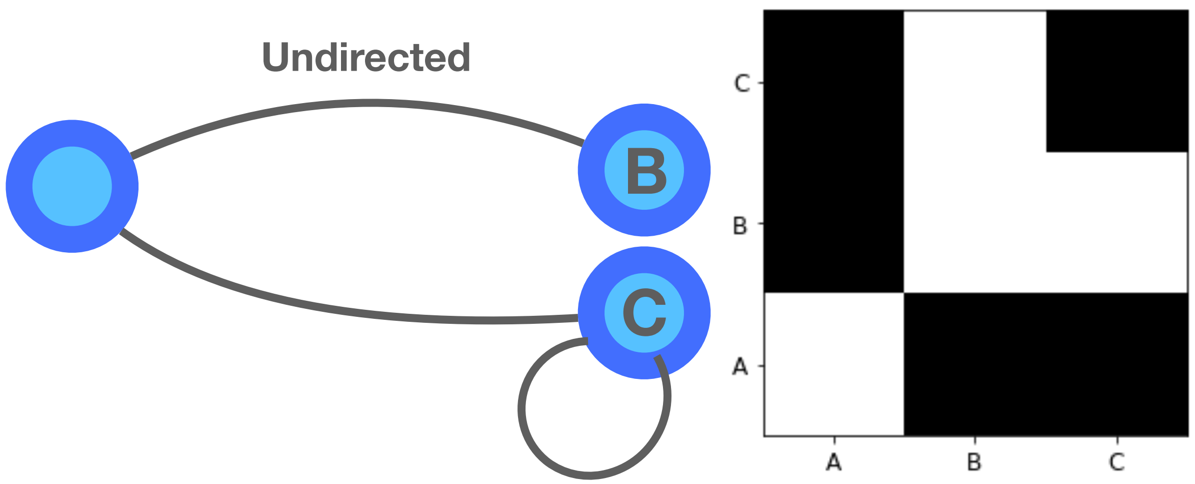
Network types

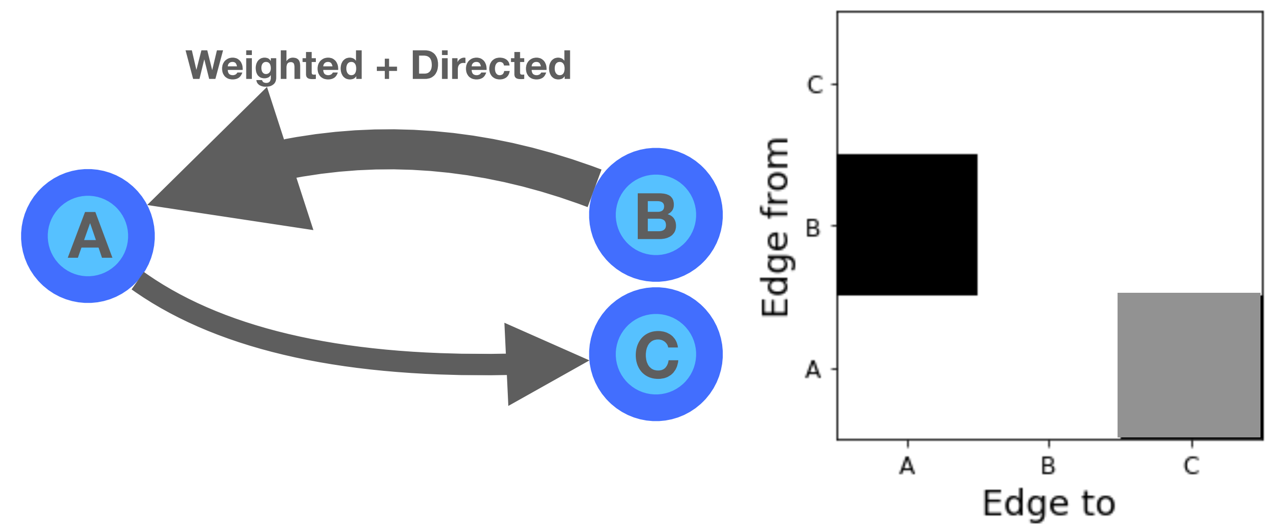


Network types

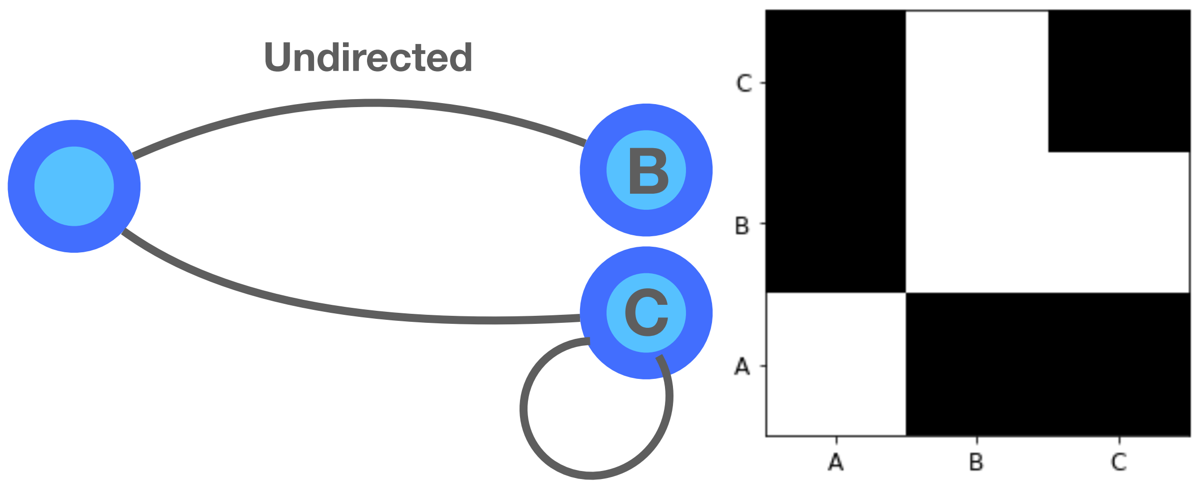


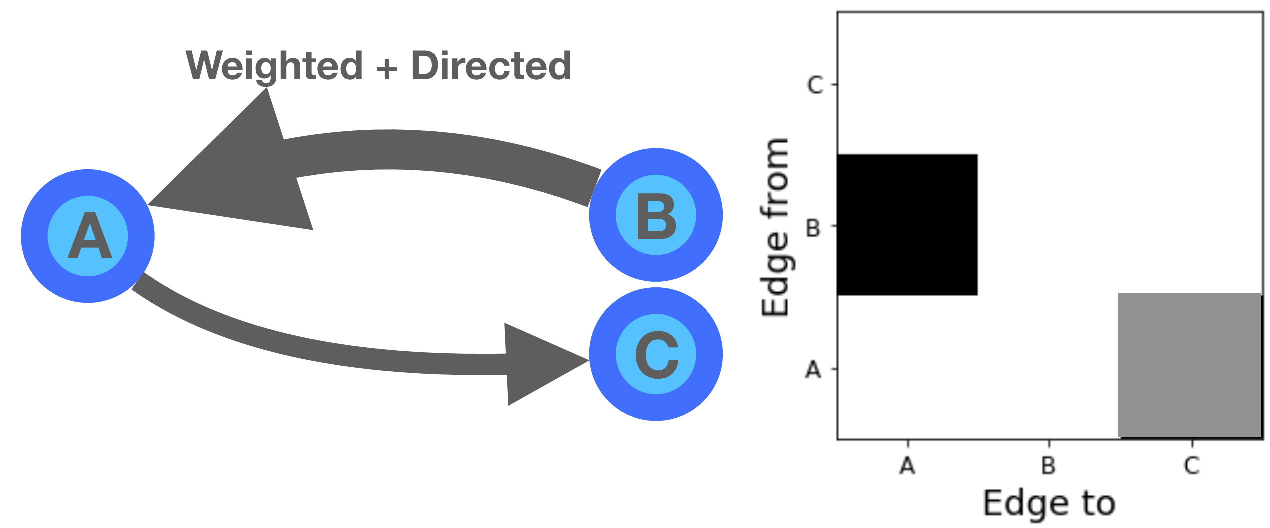
Representations for network data

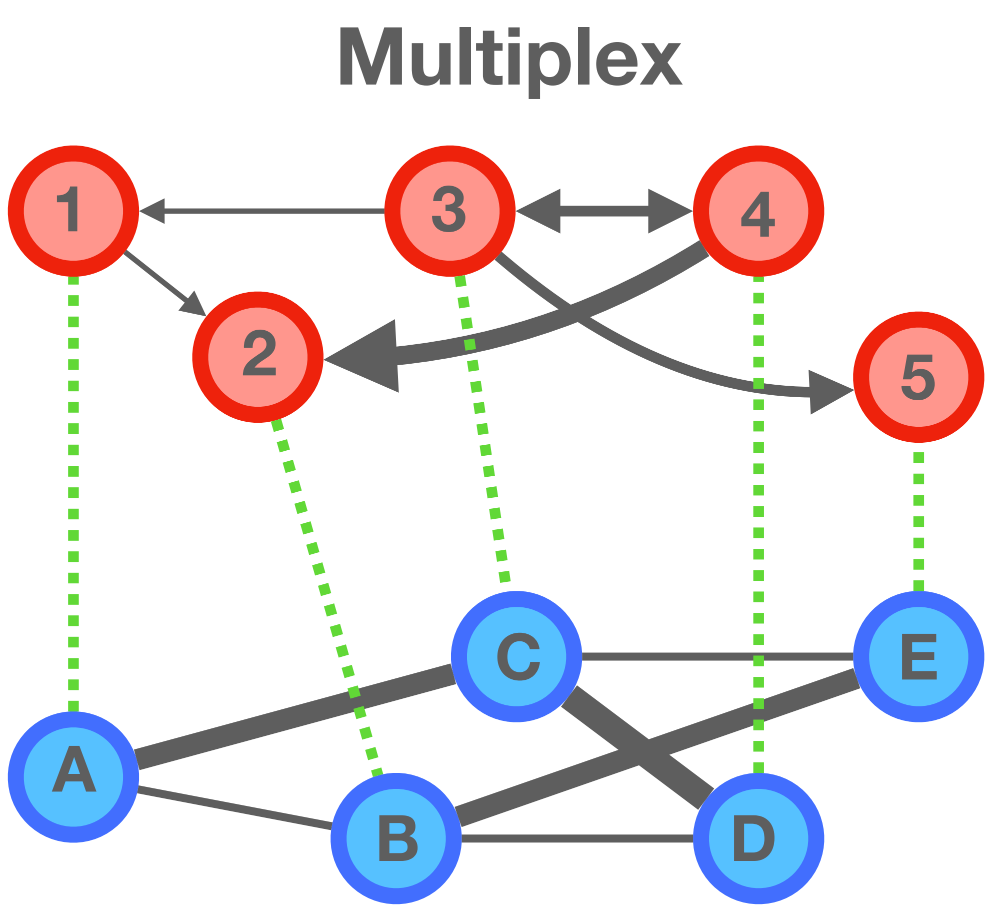




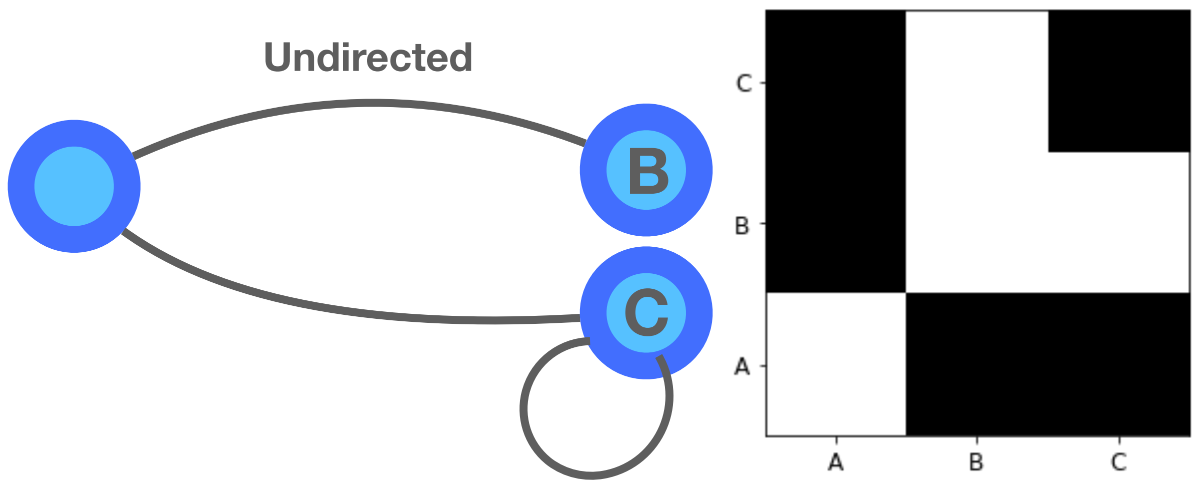
Representations for network data

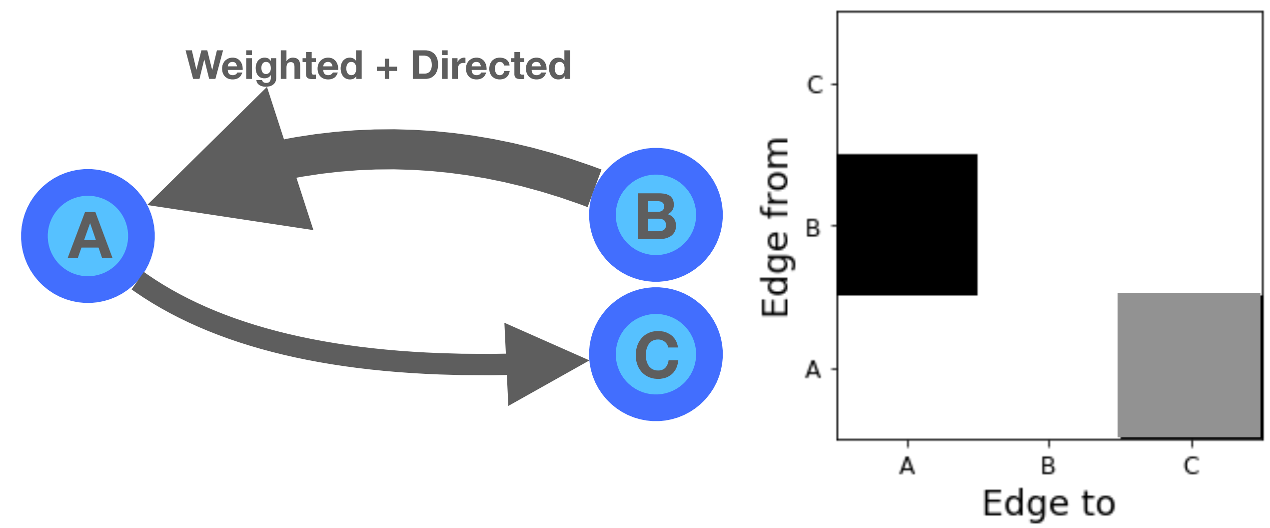


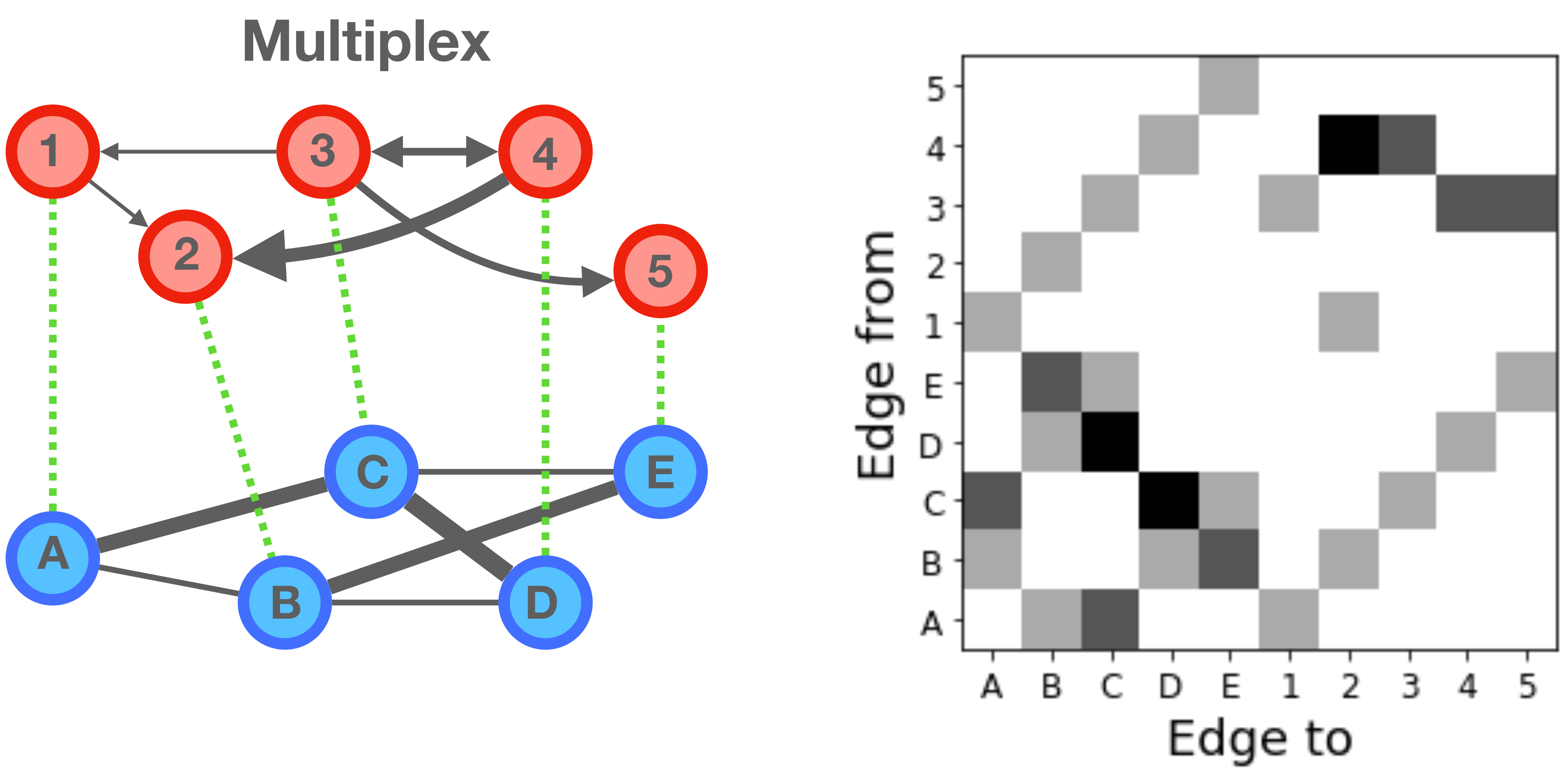




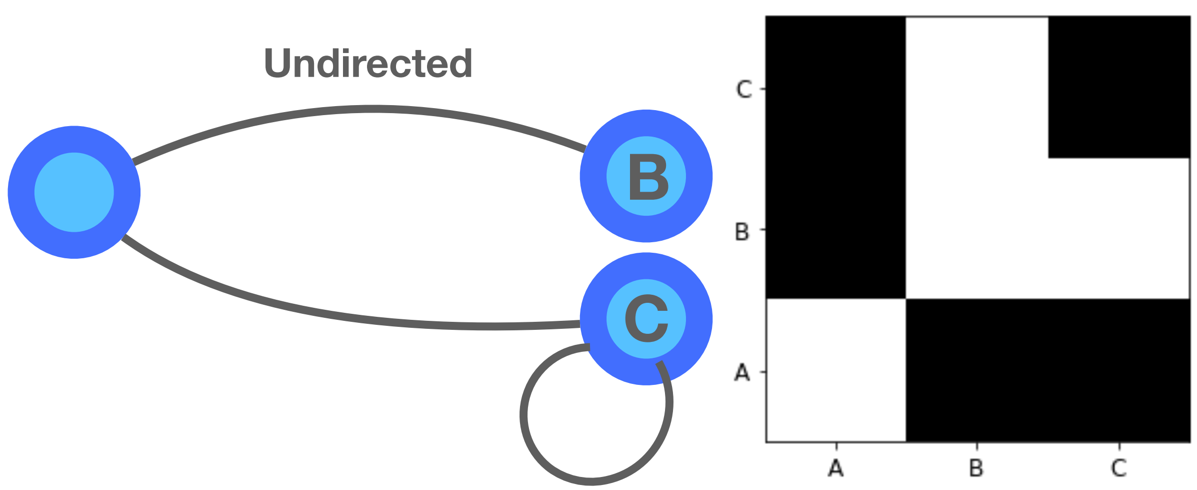
Representations for network data

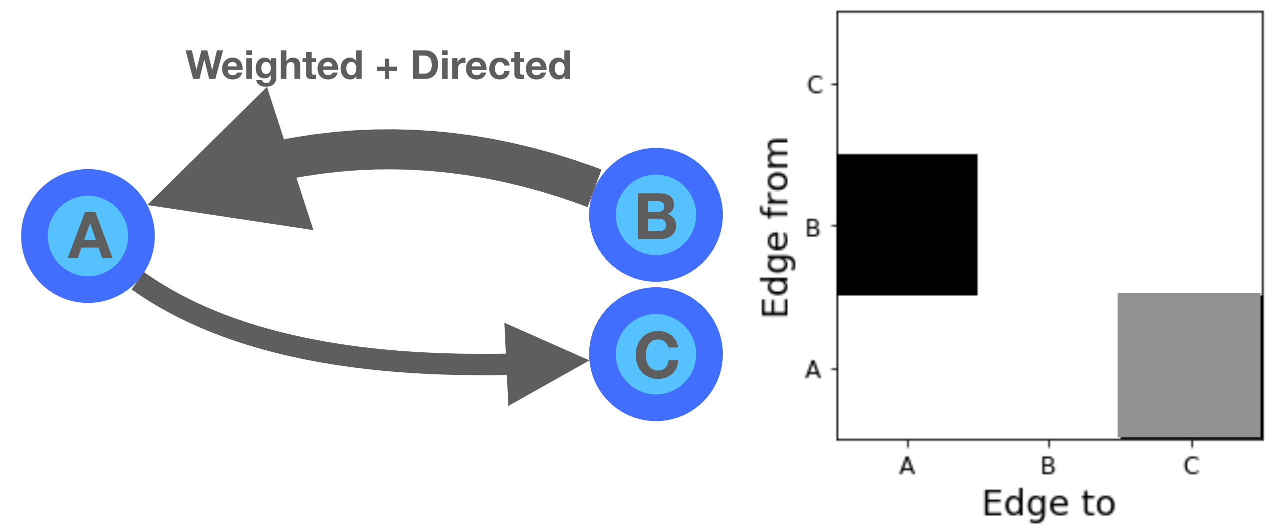


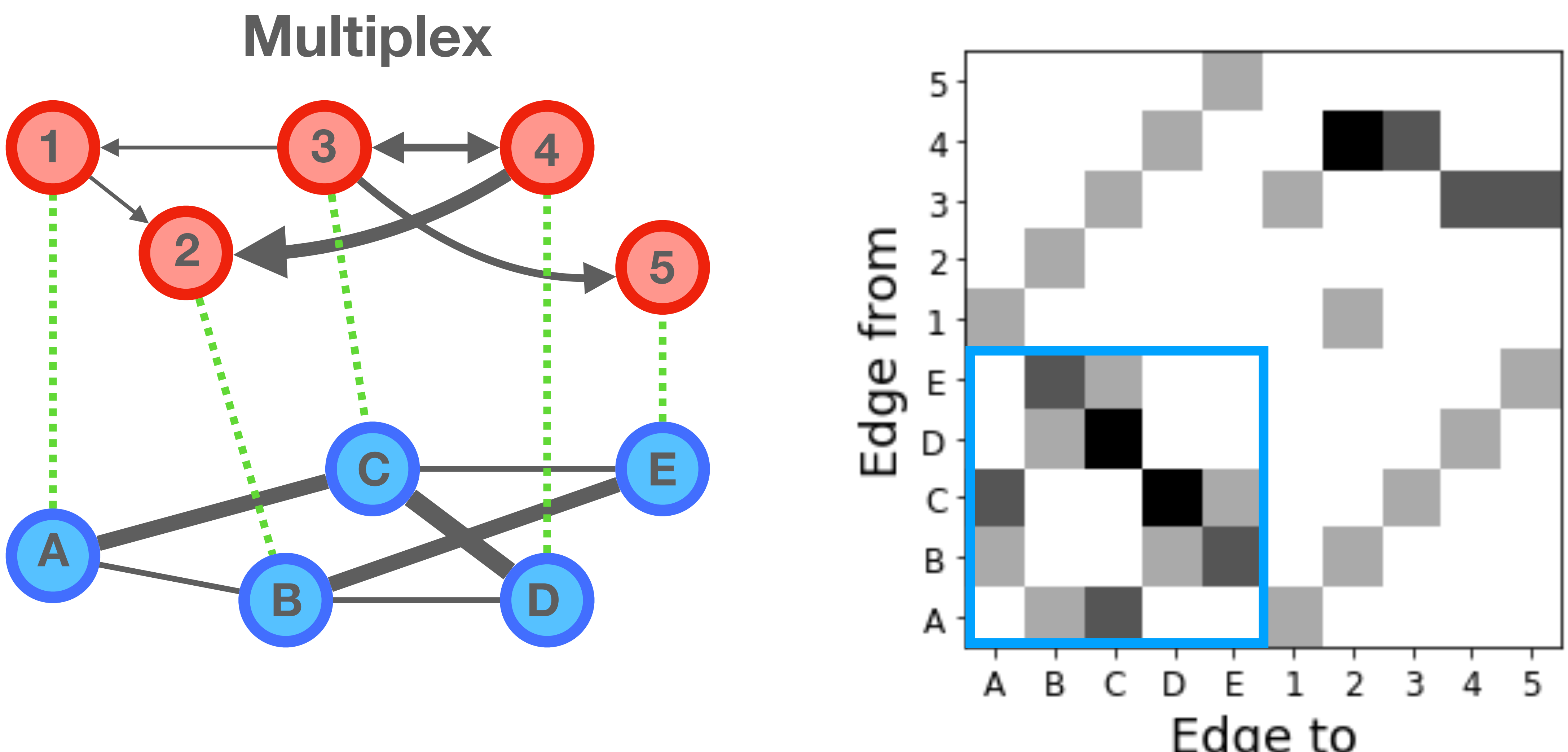




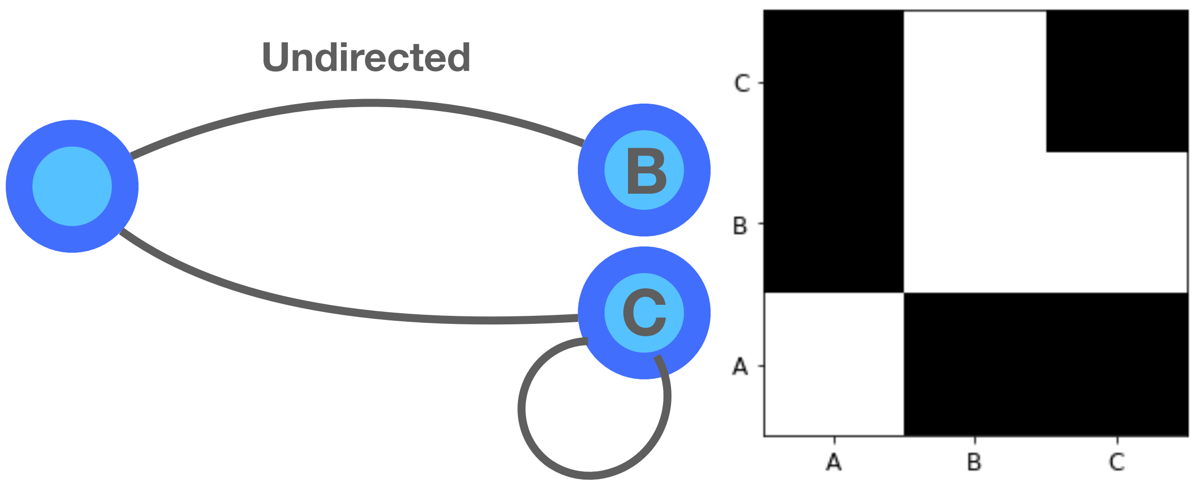
Representations for network data

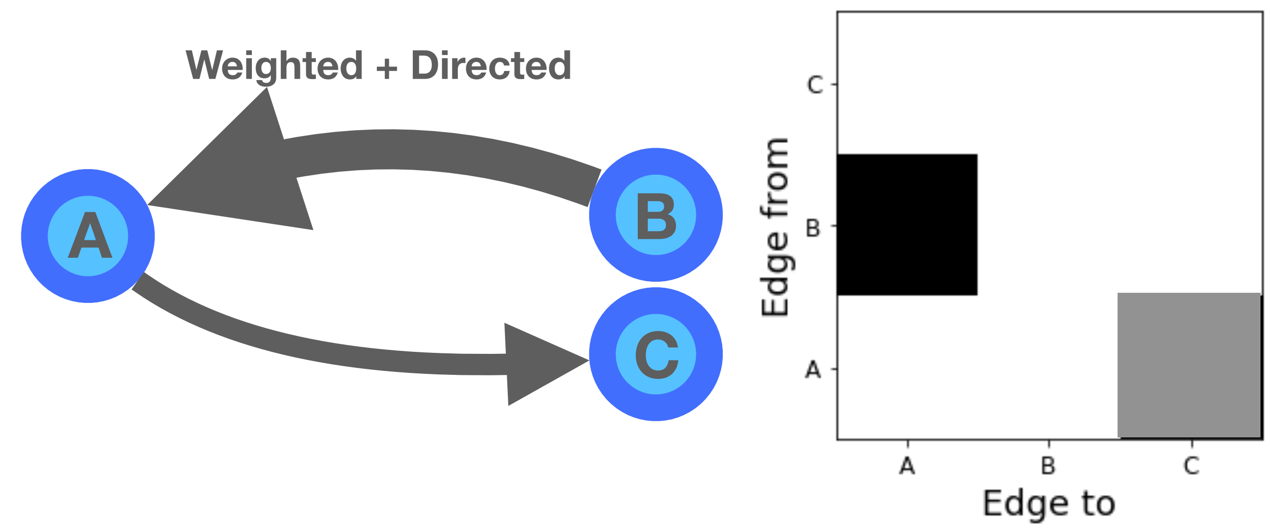


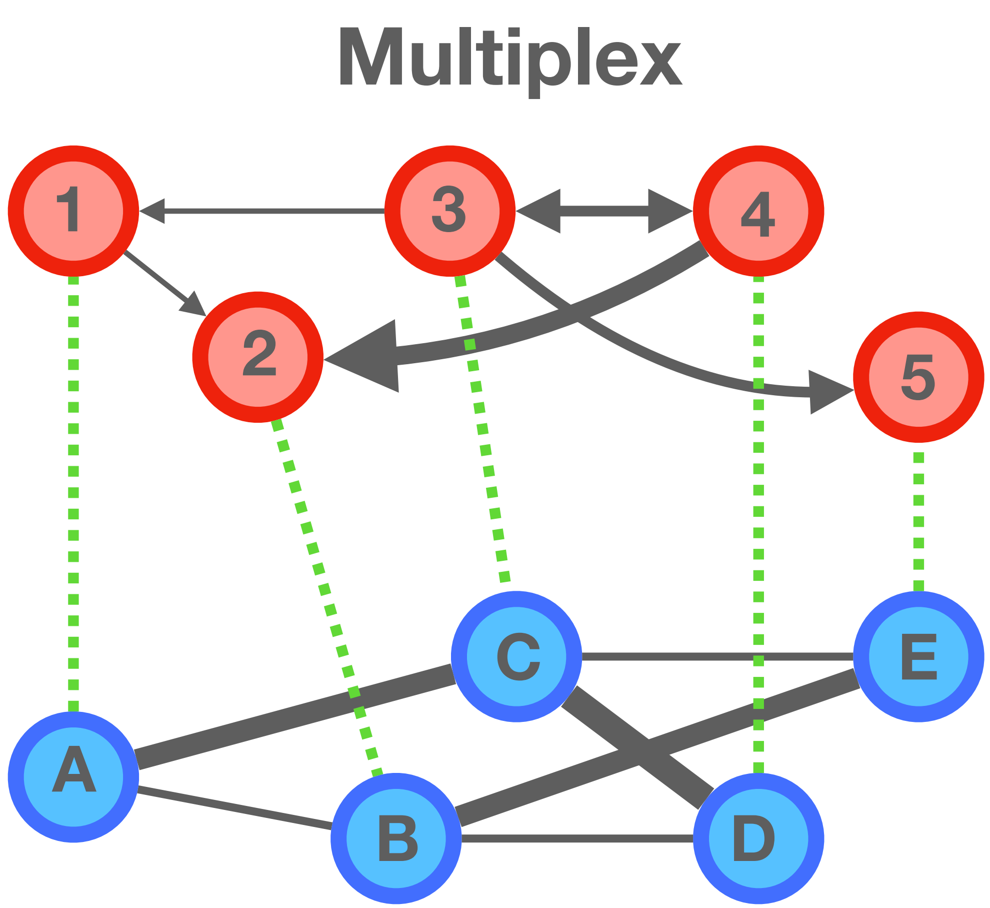


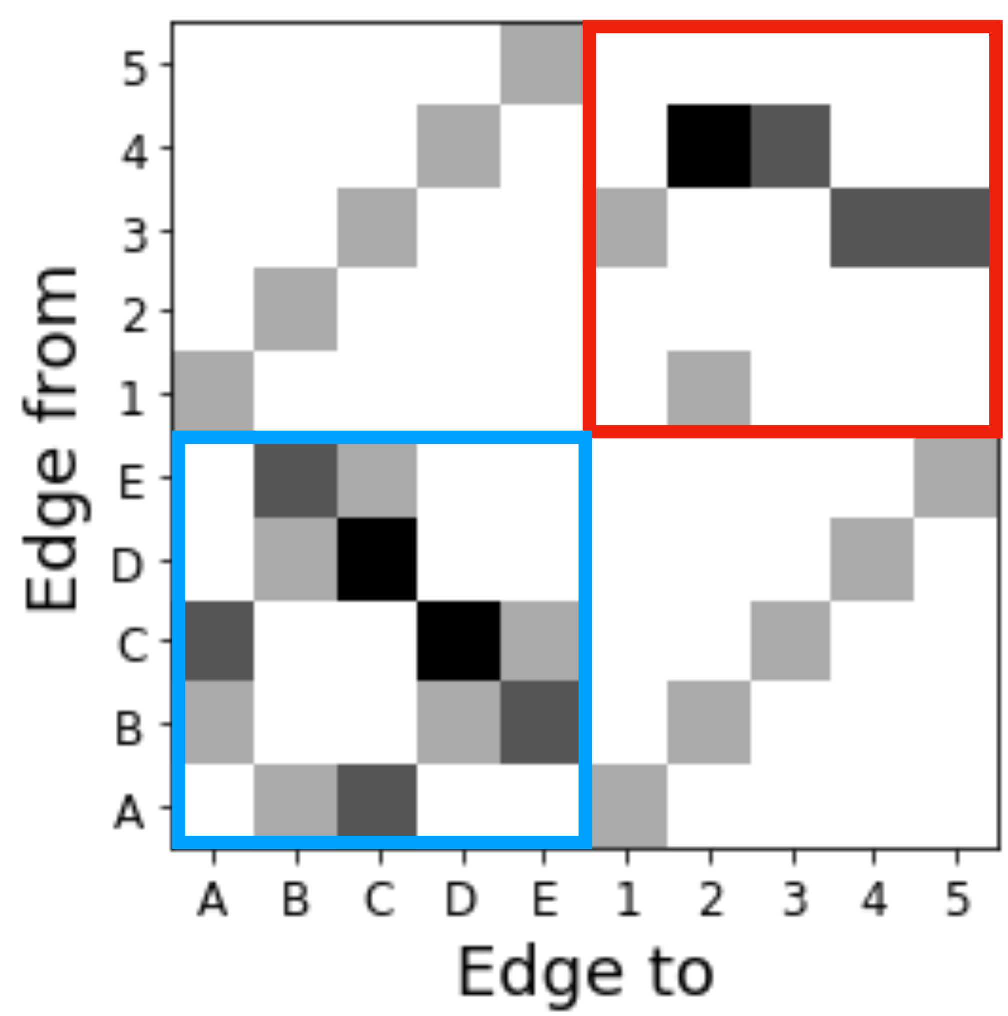


Representations for network data

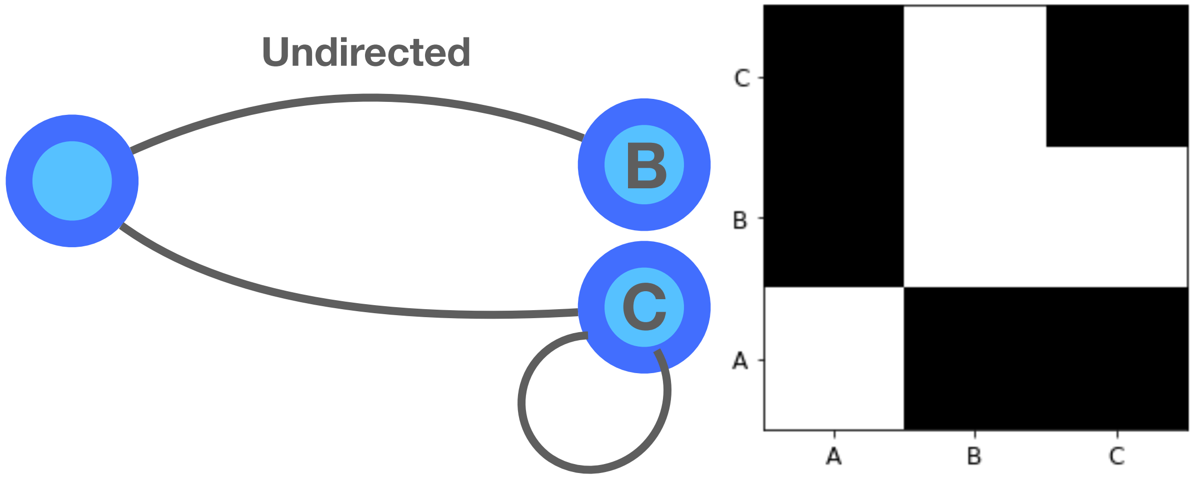


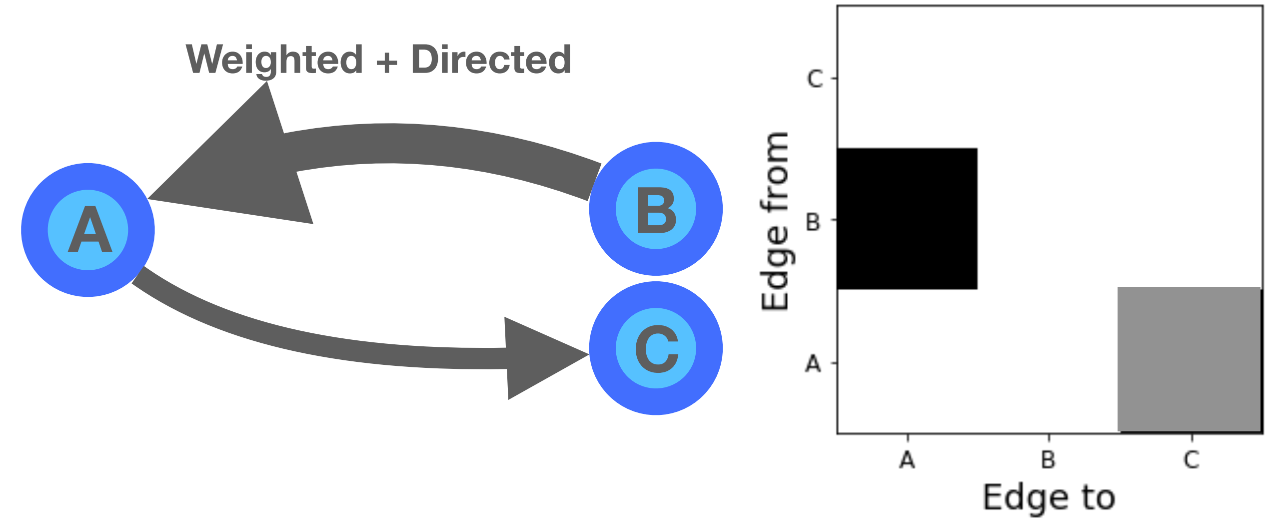


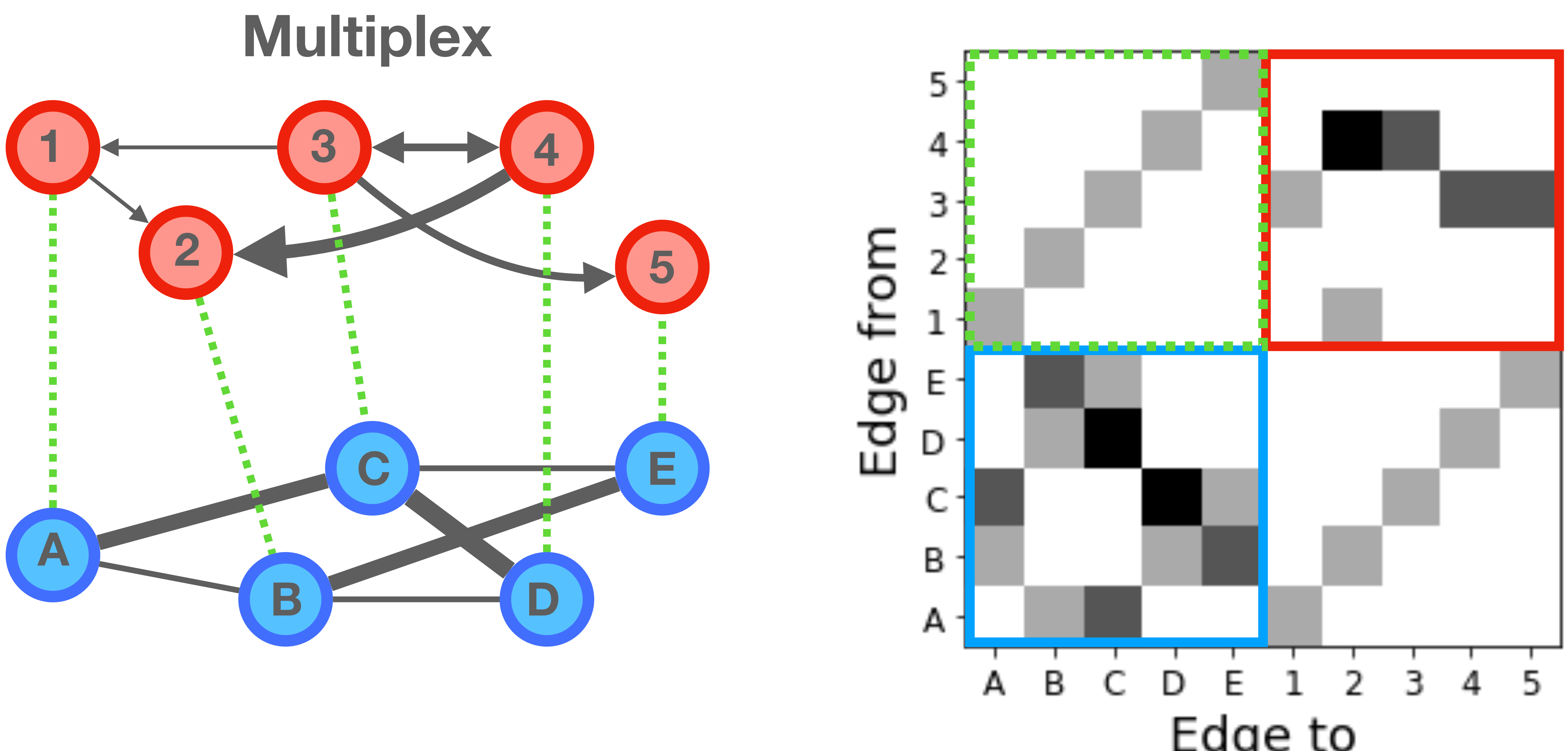




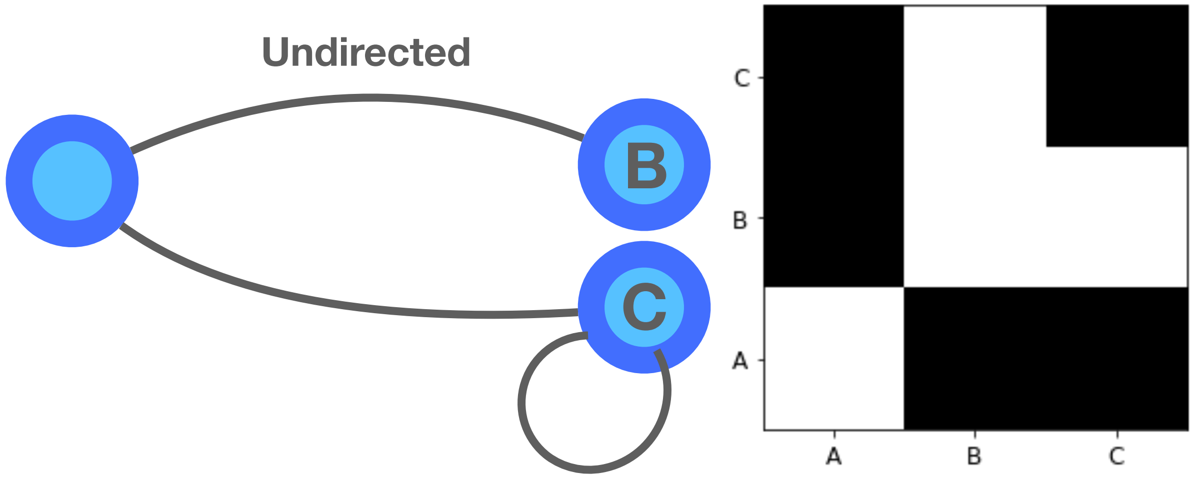
Representations for network data

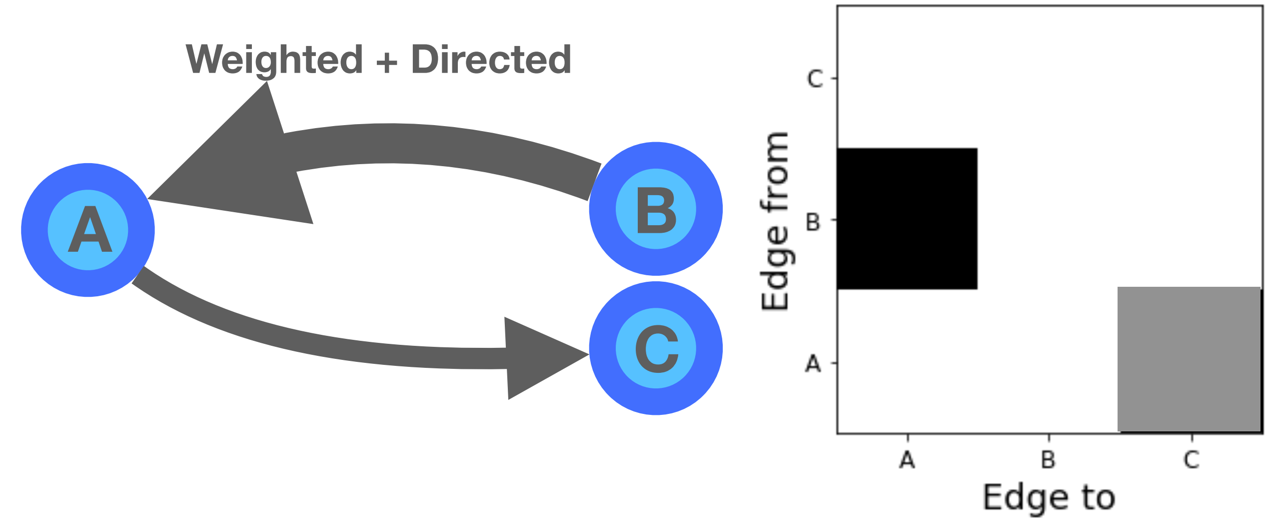


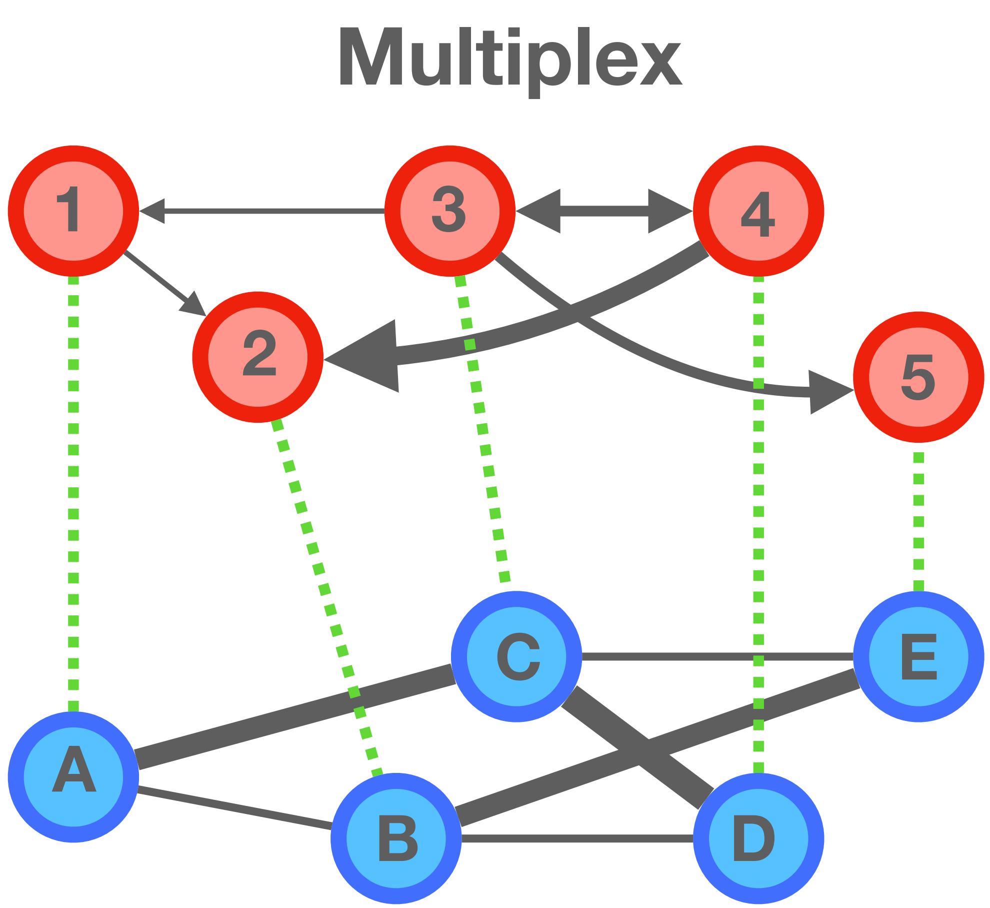


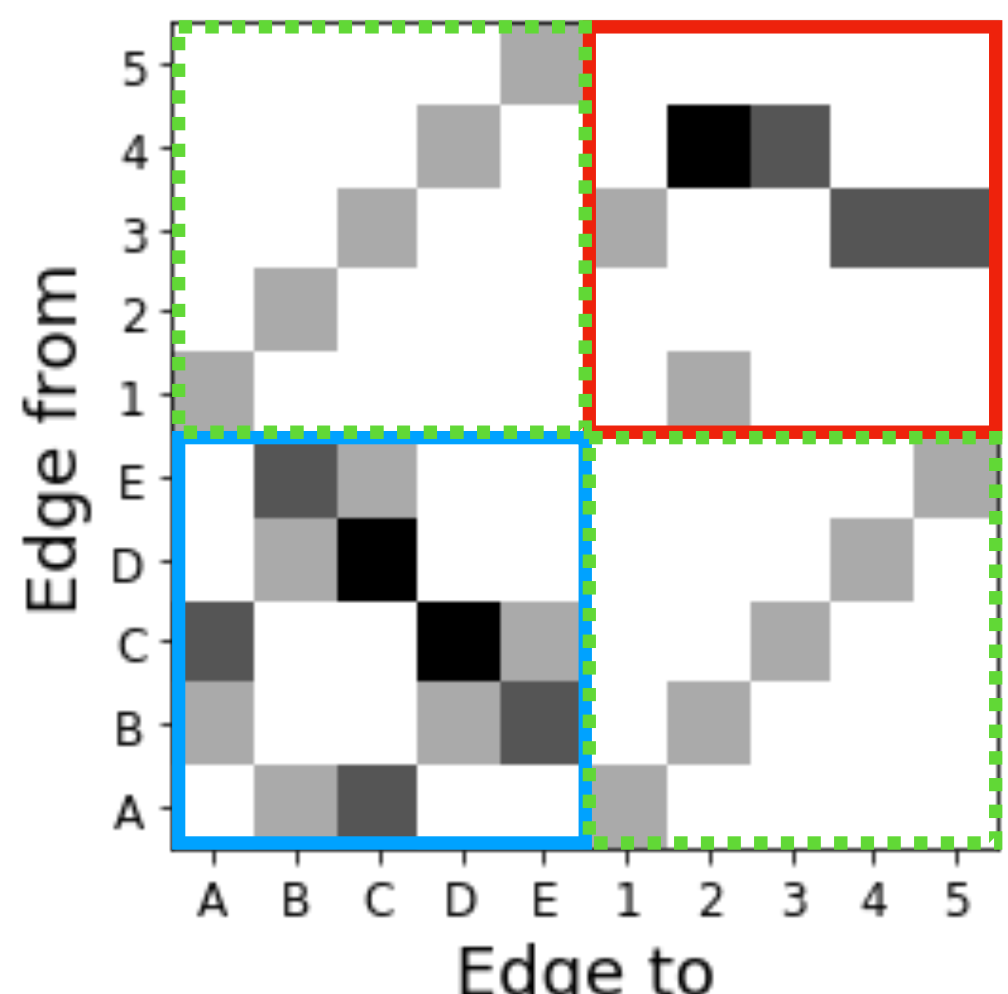


Representations for network data





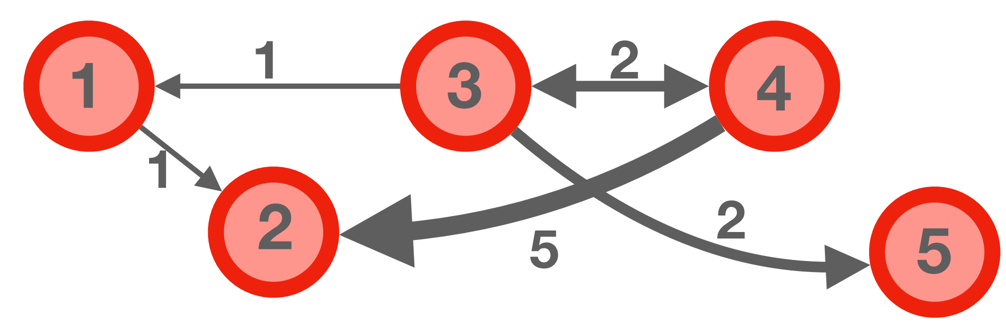


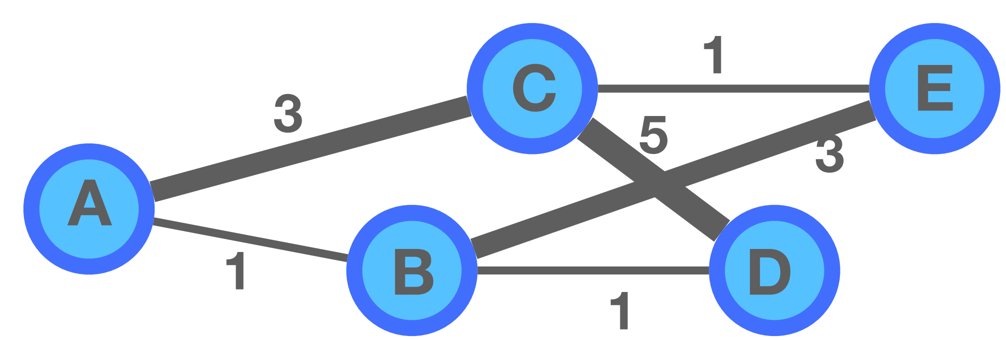


Network properties

- **Degree**: # of connections with other nodes
- **Strength**: sum of weights attached to a node
- **Density**: fraction between # of existing ties and # of all possible edges

$$L_{max} = \binom{N}{2} = N(N-1)/2$$





Network properties

- **Degree**: # of connections with other nodes

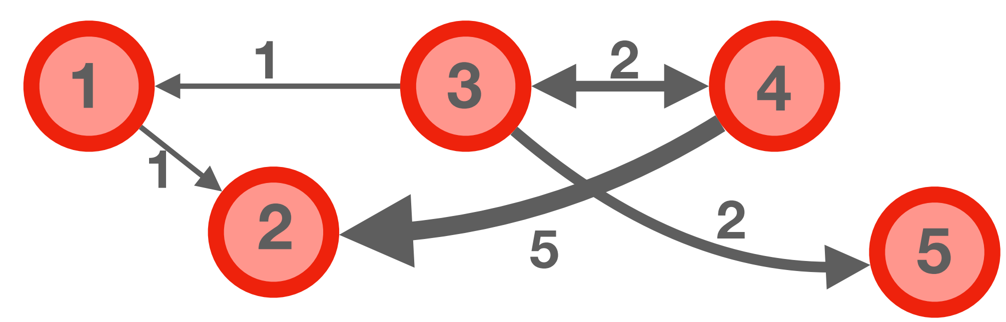
$d_A=2,\ d_C=3$

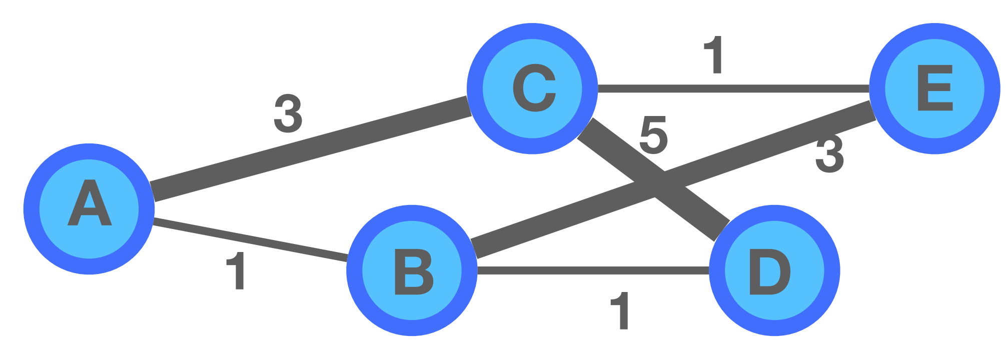
$d_{1,in}=1,\ d_{1,out}=1,\ d_{4,in}=1,\ d_{4,out}=2$

- **Strength**: sum of weights attached to a node

- **Density**: fraction between # of existing ties and # of all possible edges

$$L_{max} = \binom{N}{2} = N(N-1)/2$$





Network properties

- **Degree**: # of connections with other nodes

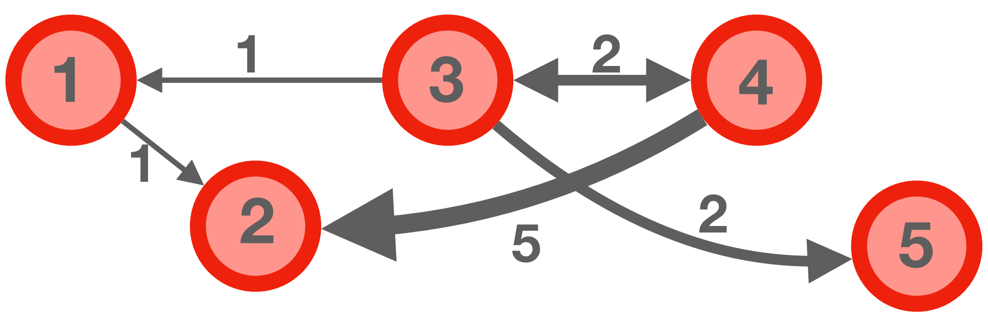
$d_A=2$, $d_C=3$
$d_{1,in}=1$, $d_{1,out}=1$, $d_{4,in}=1$, $d_{4,out}=2$

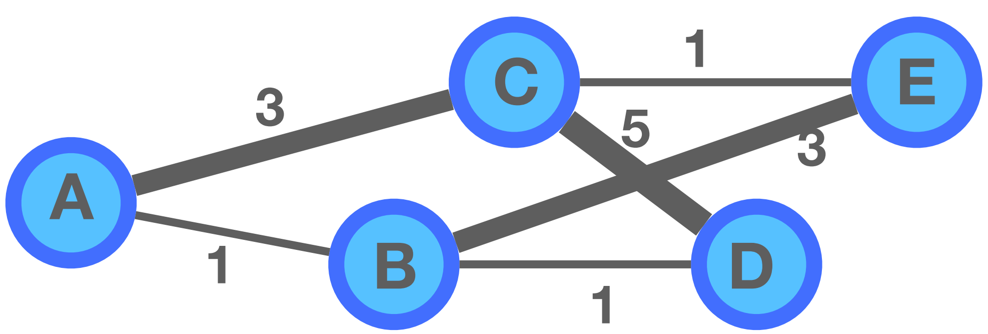
- **Strength**: sum of weights attached to a node

$s_A=4$, $s_C=9$
$s_{1,in}=1$, $s_{1,out}=1$, $s_{4,in}=2$, $s_{4,out}=7$

- **Density**: fraction between # of existing ties and # of all possible edges

$$L_{max} = \binom{N}{2} = N(N-1)/2$$





Network properties

- **Degree**: # of connections with other nodes

$d_A=2$, $d_C=3$
$d_{1,in}=1$, $d_{1,out}=1$, $d_{4,in}=1$, $d_{4,out}=2$

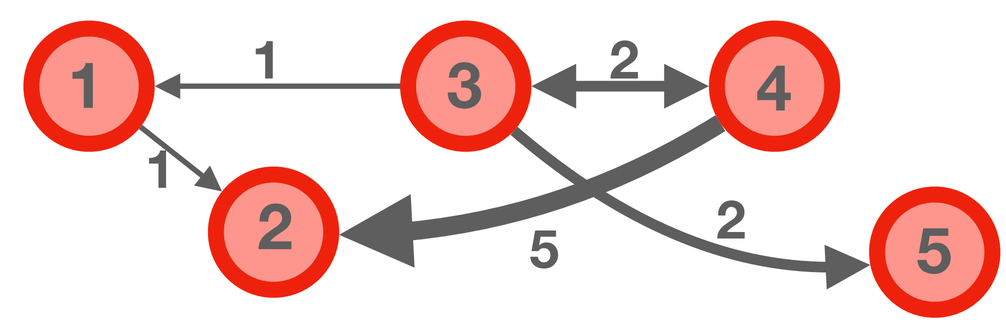
- **Strength**: sum of weights attached to a node

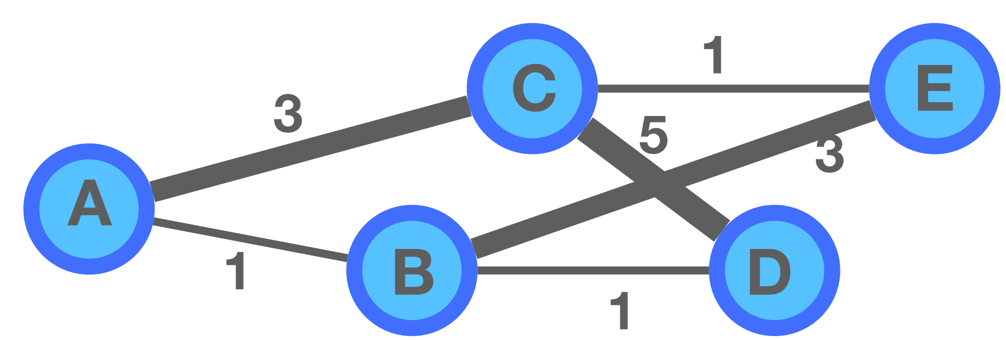
$s_A=4$, $s_C=9$
$s_{1,in}=1$, $s_{1,out}=1$, $s_{4,in}=2$, $s_{4,out}=7$

- **Density**: fraction between # of existing ties and # of all possible edges

$$L_{max} = \binom{N}{2} = N(N-1)/2$$

D = 6 * 2 / (5*(5-1)) = 0.6





Network properties

- **Degree**: # of connections with other nodes

$d_A=2$, $d_C=3$

$d_{1,in}=1$, $d_{1,out}=1$, $d_{4,in}=1$, $d_{4,out}=2$

- **Strength**: sum of weights attached to a node

$s_A=4$, $s_C=9$

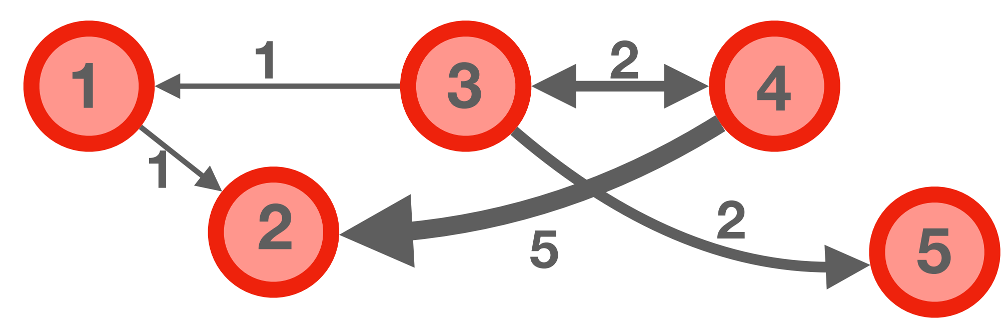
$s_{1,in}=1$, $s_{1,out}=1$, $s_{4,in}=2$, $s_{4,out}=7$

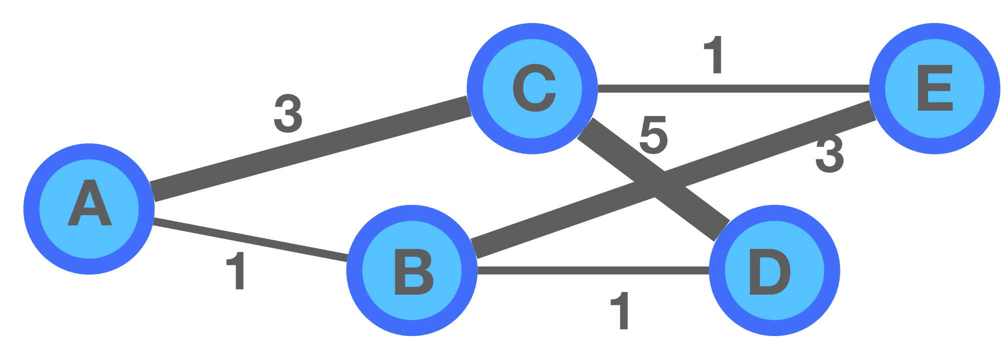
- **Density**: fraction between # of existing ties and # of all possible edges

$$L_{max} = \binom{N}{2} = N(N-1)/2$$

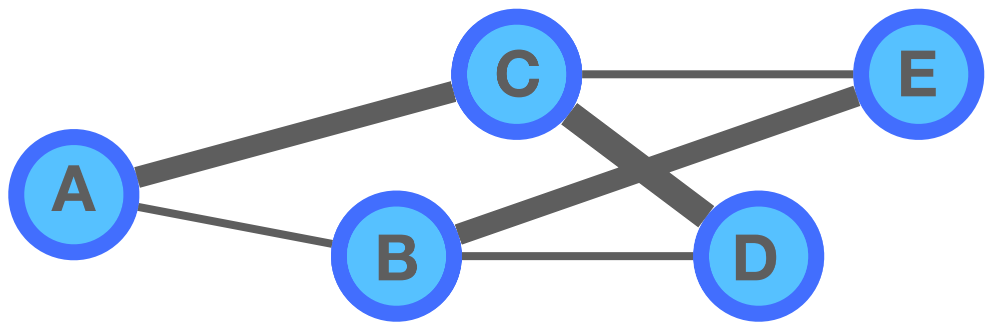
D = 6 * 2 / (5*(5-1)) = 0.6

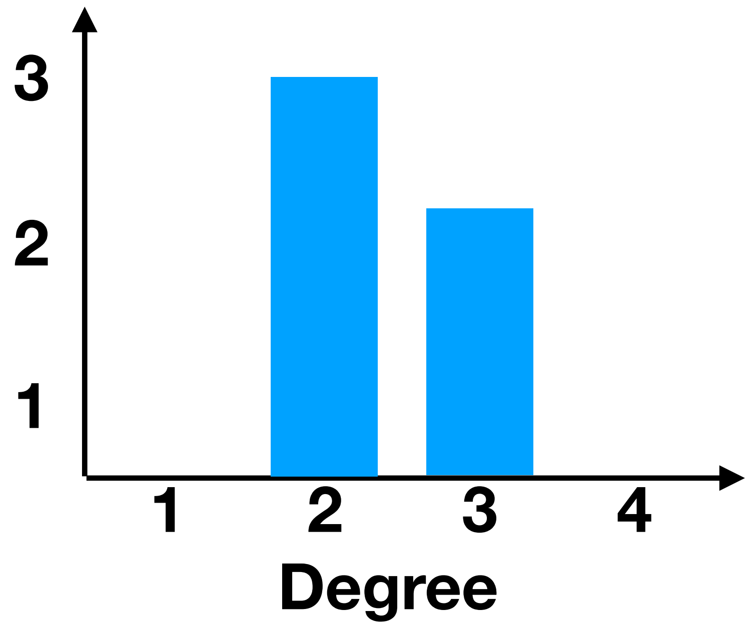
D = 6 / (5*(5-1)) = 0.3



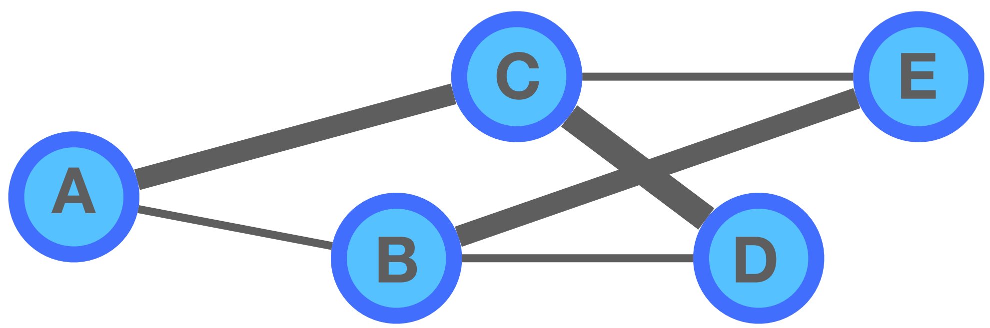


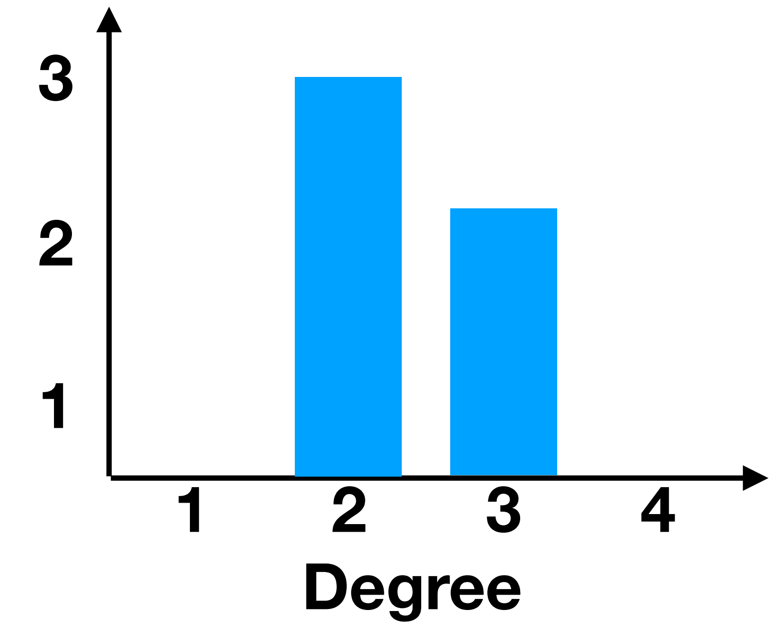
Degree distribution

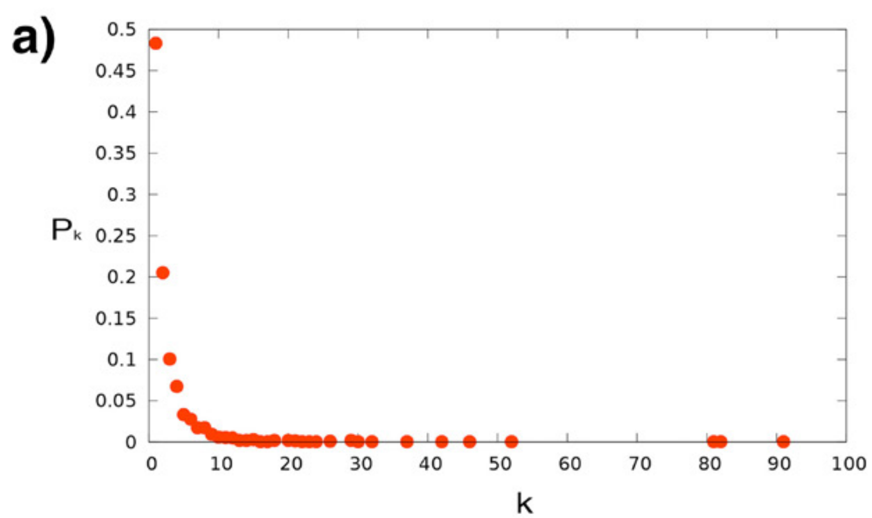


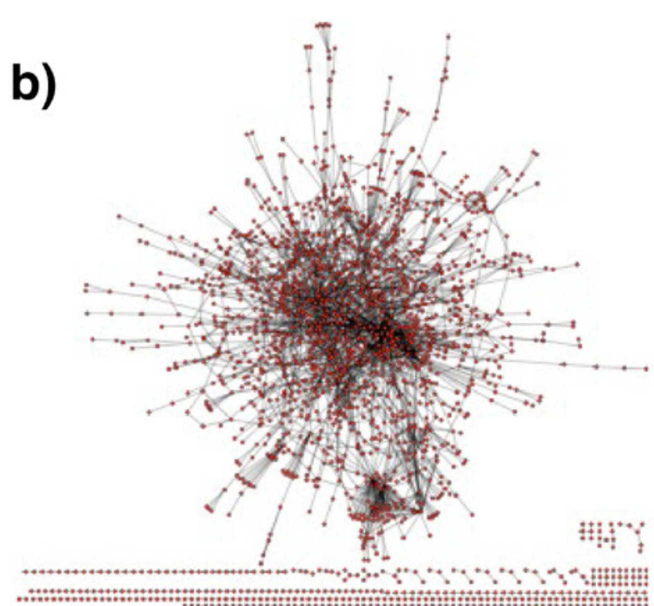


Degree distribution









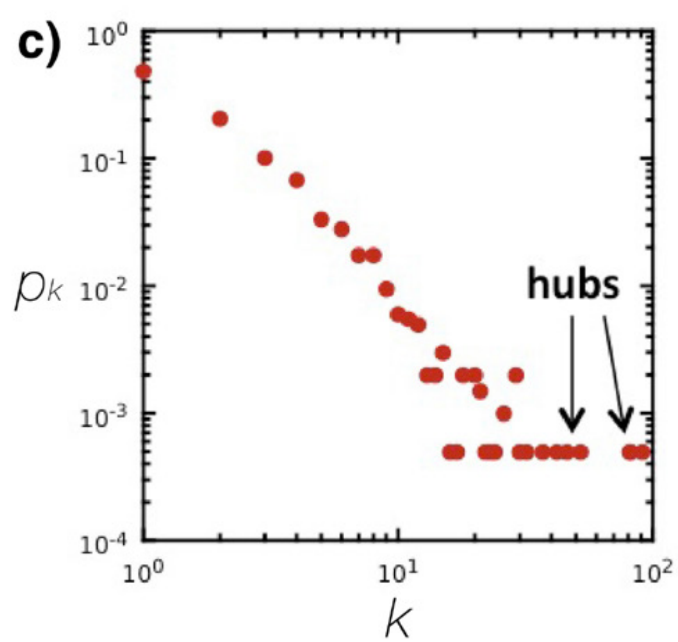
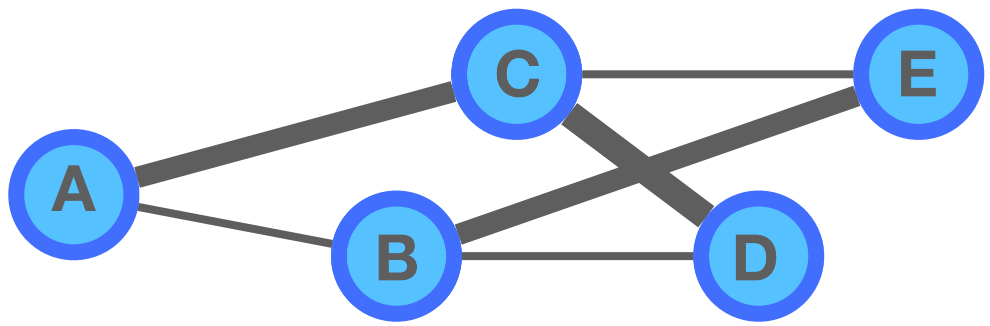
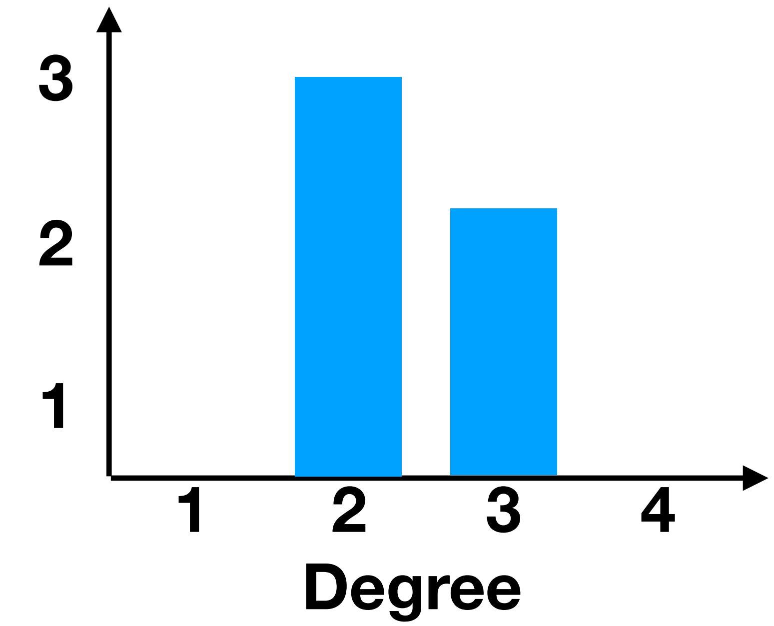
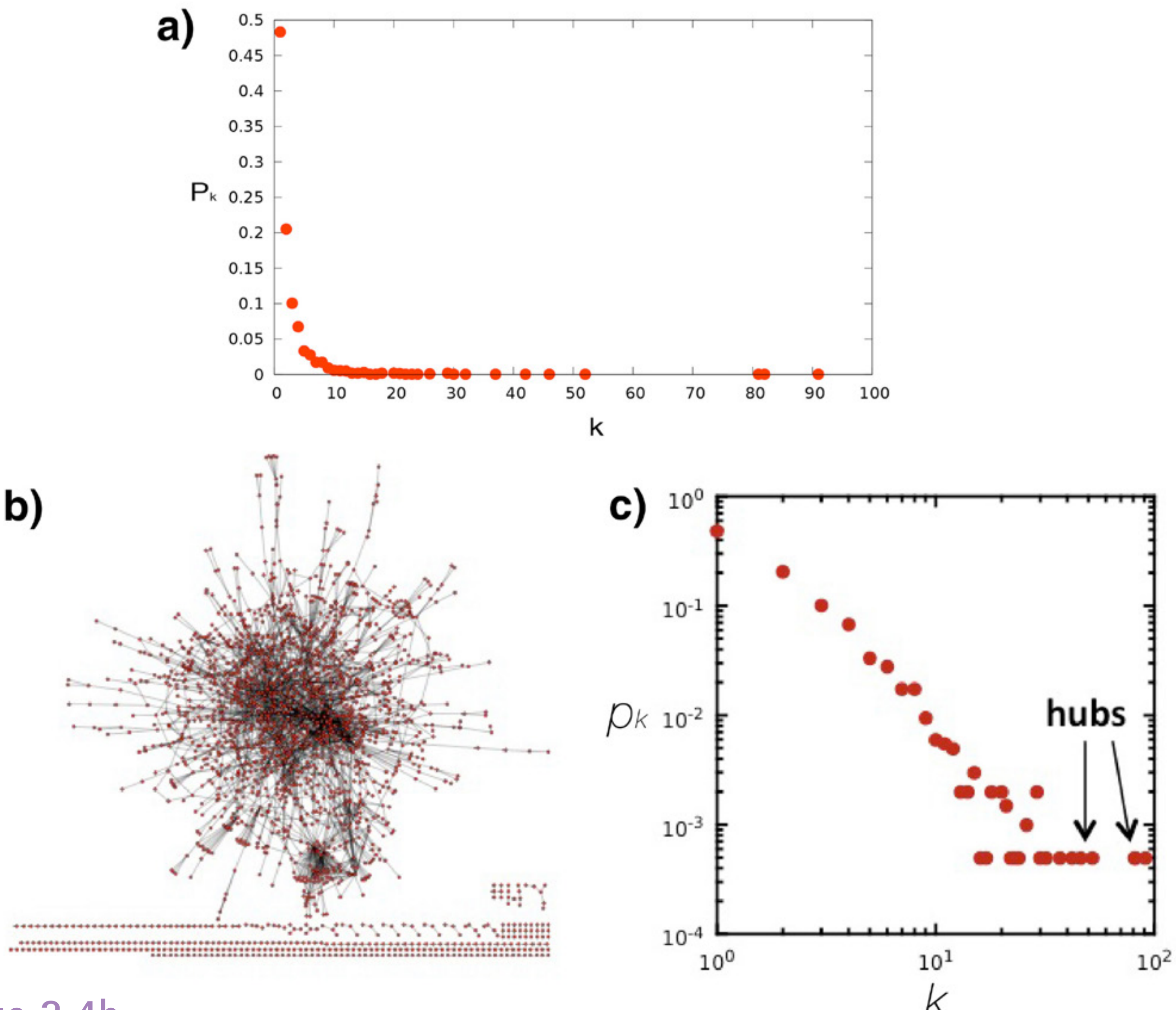


Image 2.4b

Degree distribution







Real networks are sparse!

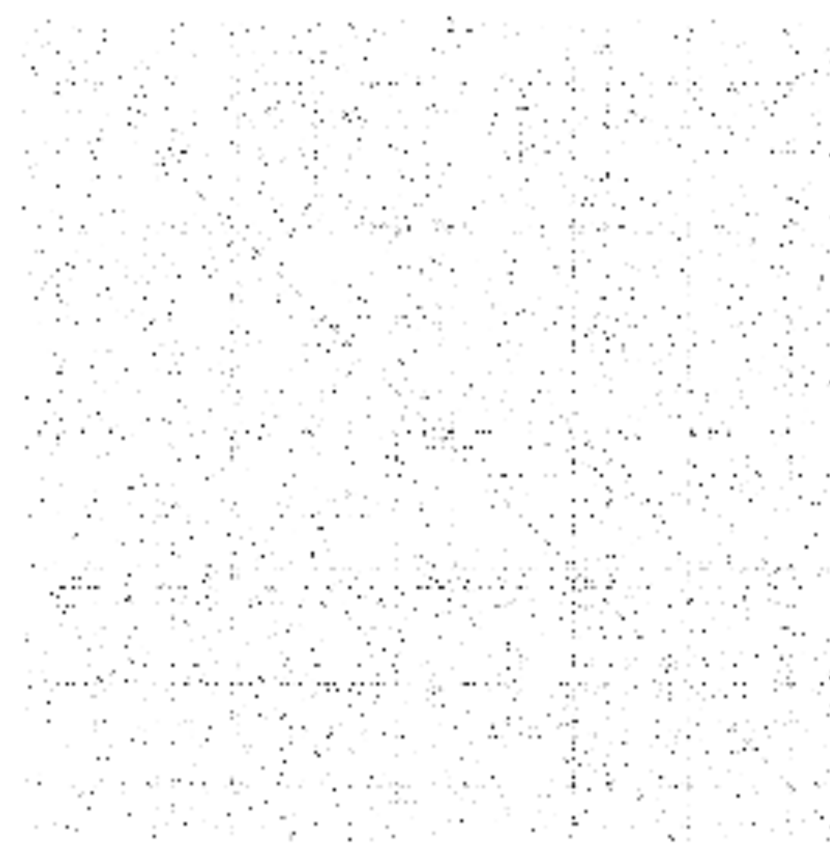


Image 2.4b

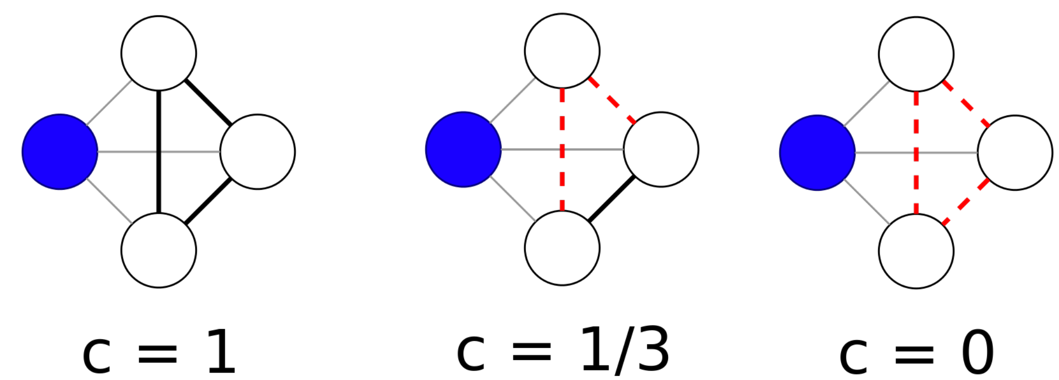
Network properties

Clustering coefficient: A measure of the degree to which nodes in a graph tend to cluster together. The global version was designed to give an overall indication of the clustering in the network, whereas the local clustering coef. gives an indication of the embeddedness of single nodes.

Network properties

Clustering coefficient: A measure of the degree to which nodes in a graph tend to cluster together. The global version was designed to give an overall indication of the clustering in the network, whereas the local clustering coef. gives an indication of the embeddedness of single nodes.

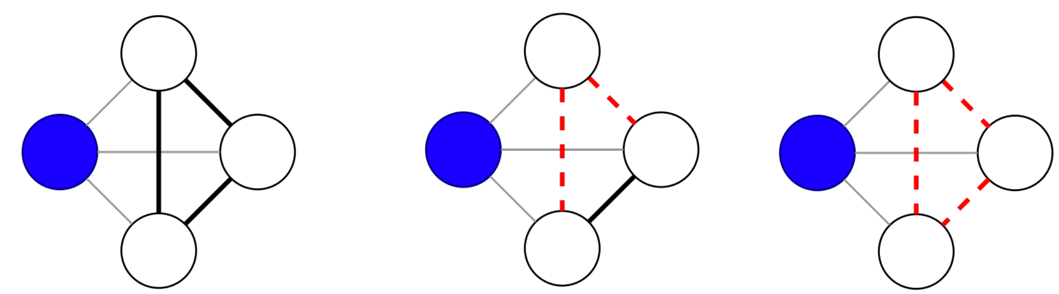
The **local clustering coefficient** C_i for a vertex v_i is given by the proportion of links between the vertices within its neighbourhood divided by the number of links that could possibly exist between them.



Network properties

Clustering coefficient: A measure of the degree to which nodes in a graph tend to cluster together. The global version was designed to give an overall indication of the clustering in the network, whereas the local clustering coef. gives an indication of the embeddedness of single nodes.

The **local clustering coefficient** C_i for a vertex v_i is given by the proportion of links between the vertices within its neighbourhood divided by the number of links that could possibly exist between them.



c = 1   c = 1/3   c = 0

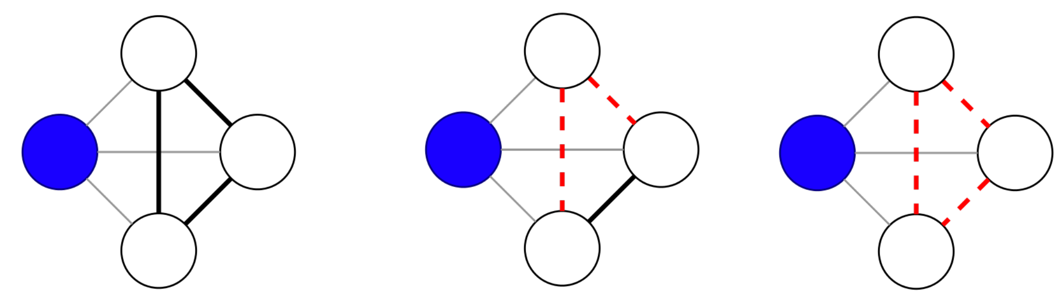
Global clustering coefficient

$$C = \frac{3 \text{ x number of triangles}}{\text{number of all triplets}}$$

Network properties

Clustering coefficient: A measure of the degree to which nodes in a graph tend to cluster together. The global version was designed to give an overall indication of the clustering in the network, whereas the local clustering coef. gives an indication of the embeddedness of single nodes.

The **local clustering coefficient** C_i for a vertex v_i is given by the proportion of links between the vertices within its neighbourhood divided by the number of links that could possibly exist between them.



c = 1   c = 1/3   c = 0

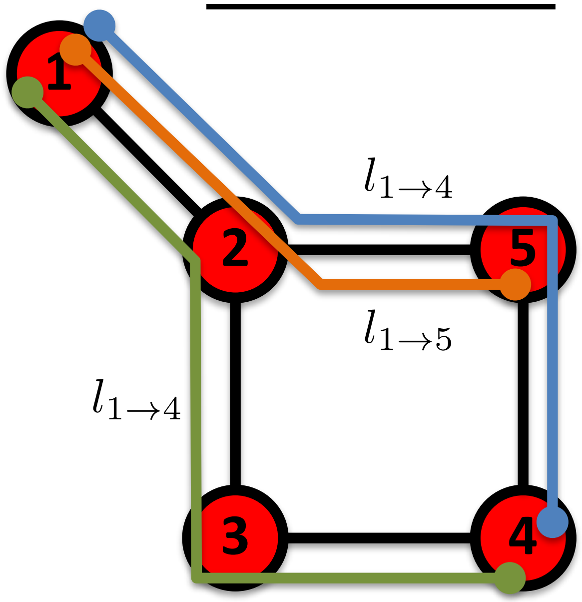
Global clustering coefficient

$$C = \frac{\texttt{3 x number of triangles}}{\texttt{number of all triplets}}$$

Avg. clustering coefficient

$$\bar{C} = \frac{1}{n}\sum_{i=1}^{n} C_i$$

Shortest Path



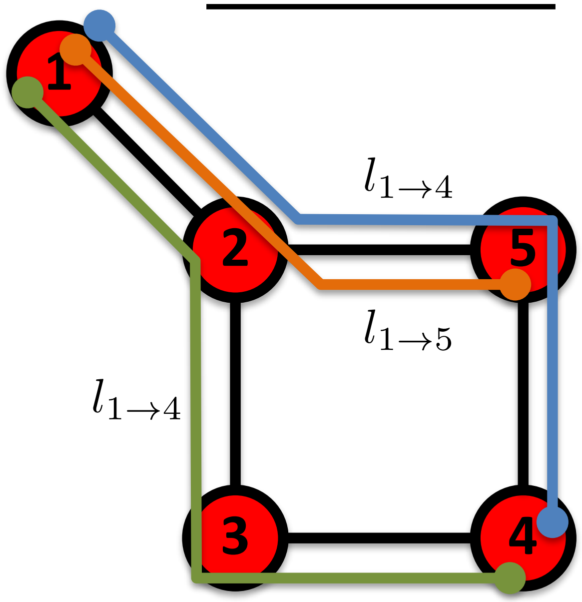
$$l_{1\to 4} = 3$$

$$l_{1\to 5} = 2$$

The path with the shortest length between two nodes (distance).

Slide from the Network Science book

Shortest Path

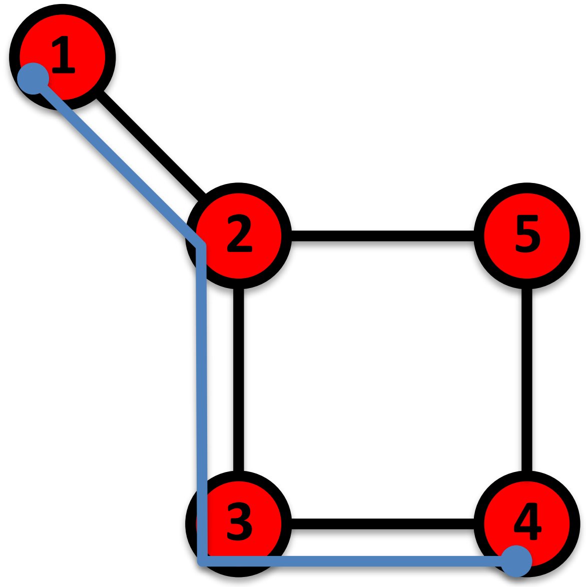


$l_{1\to4} = 3$

$l_{1\to5} = 2$

The path with the shortest length between two nodes (distance).

Diameter

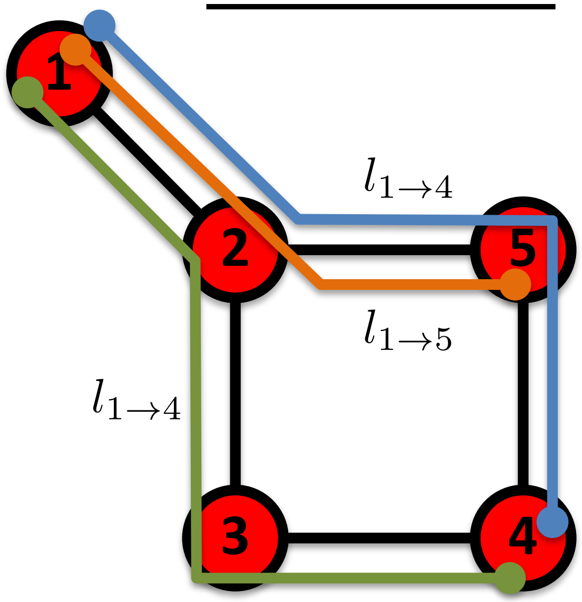


$l_{1\to4} = 3$

The longest shortest path in a graph

Slide from the Network Science book

Shortest Path

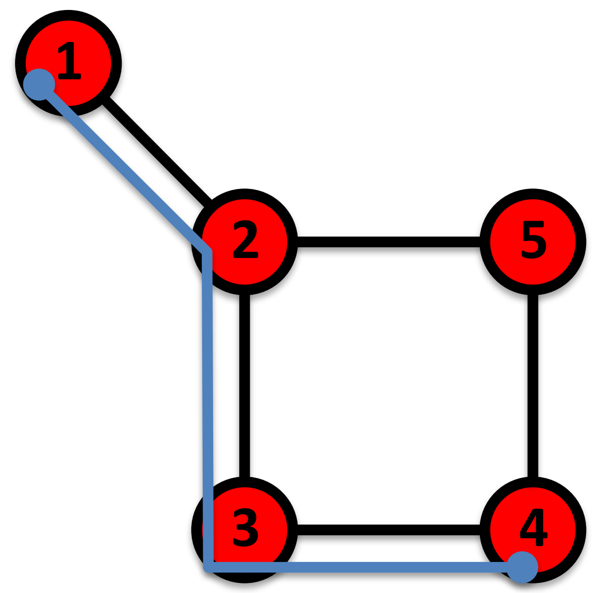


$l_{1\to4} = 3$

$l_{1\to5} = 2$

The path with the shortest length between two nodes (distance).

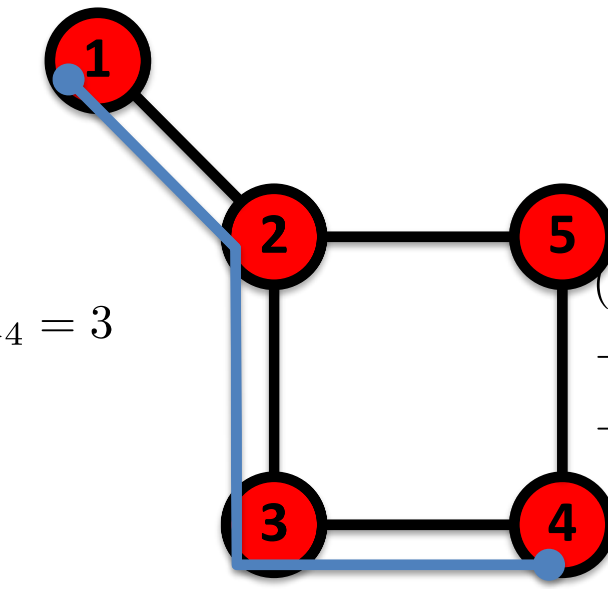
Diameter



$l_{1\to4} = 3$

The longest shortest path in a graph

Average Path Length

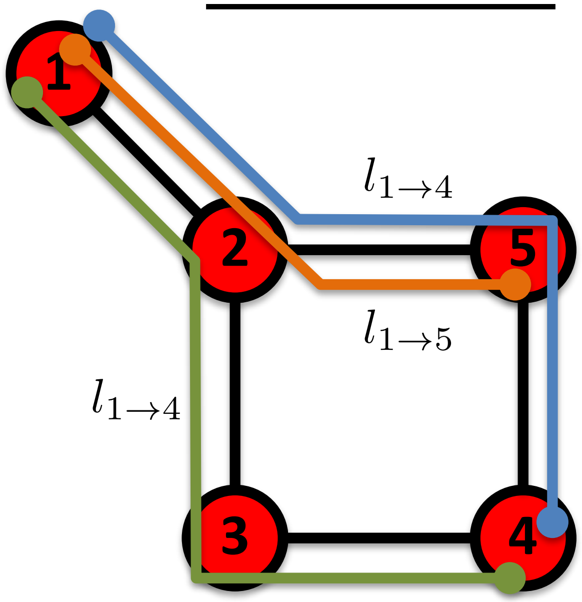


$$\begin{gathered}(l_{1\to2} + l_{1\to3} + l_{1\to4} + \\ + l_{1\to5} + l_{2\to3} + l_{2\to4} + \\ + l_{2\to5} + l_{3\to4} + l_{3\to5} + \\ + l_{4\to5})/10 = 1.6\end{gathered}$$

The average of the shortest paths for all pairs of nodes.

Slide from the Network Science book

Shortest Path

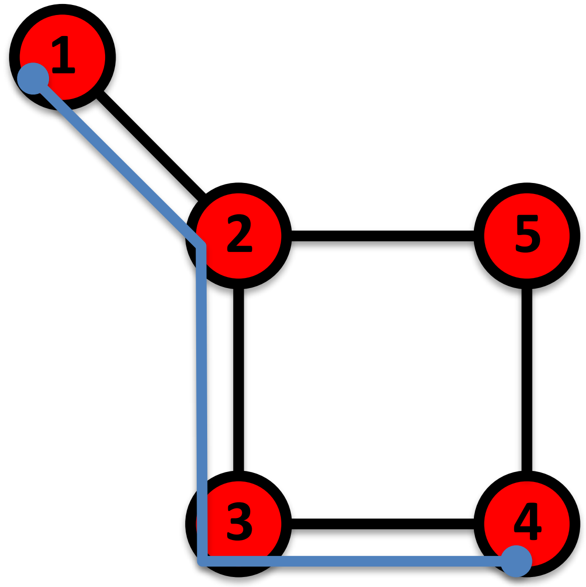


$l_{1\to4} = 3$

$l_{1\to5} = 2$

The path with the shortest length between two nodes (distance).

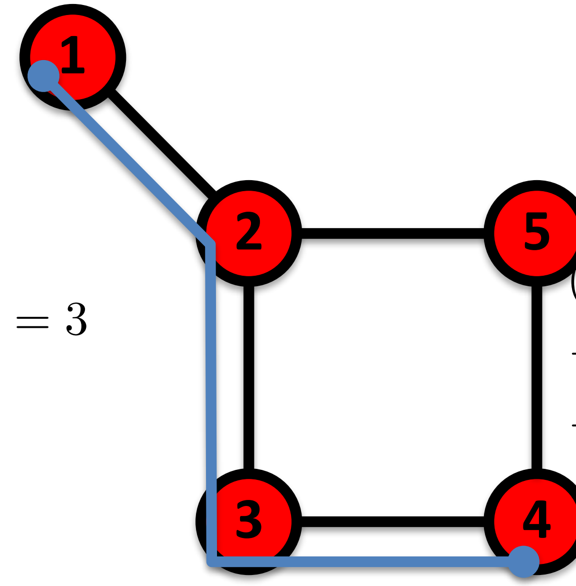
Diameter



$l_{1\to4} = 3$

The longest shortest path in a graph

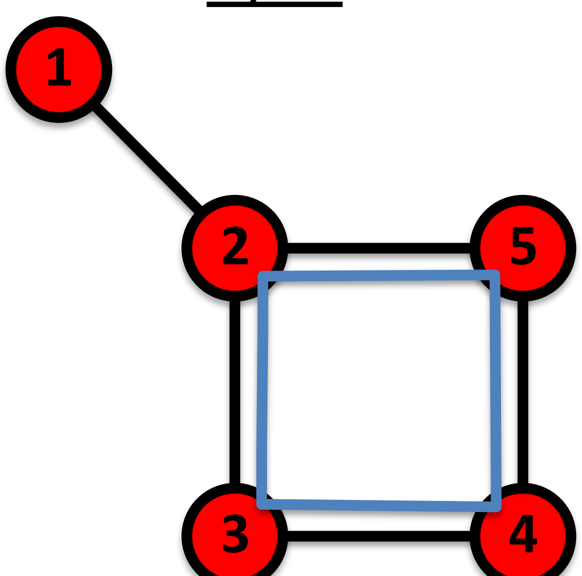
Average Path Length



$$\begin{aligned}&(l_{1\to2} + l_{1\to3} + l_{1\to4} +\\ &+ l_{1\to5} + l_{2\to3} + l_{2\to4} +\\ &+ l_{2\to5} + l_{3\to4} + l_{3\to5} +\\ &+ l_{4\to5})/10 = 1.6\end{aligned}$$

The average of the shortest paths for all pairs of nodes.

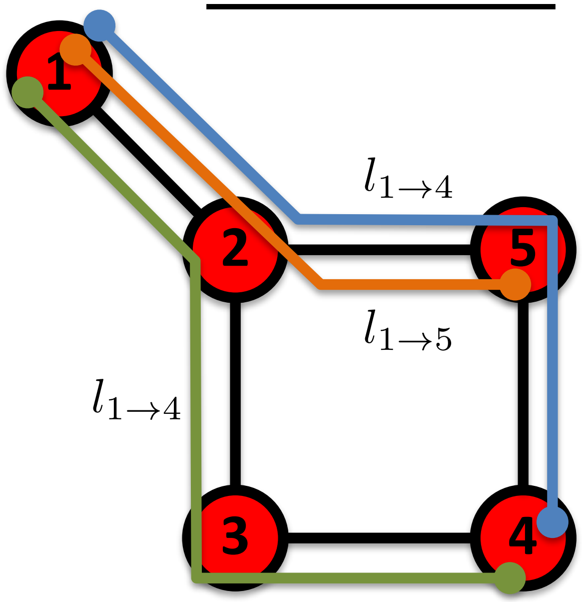
Cycle



A path with the same start and end node.

Slide from the Network Science book

Shortest Path

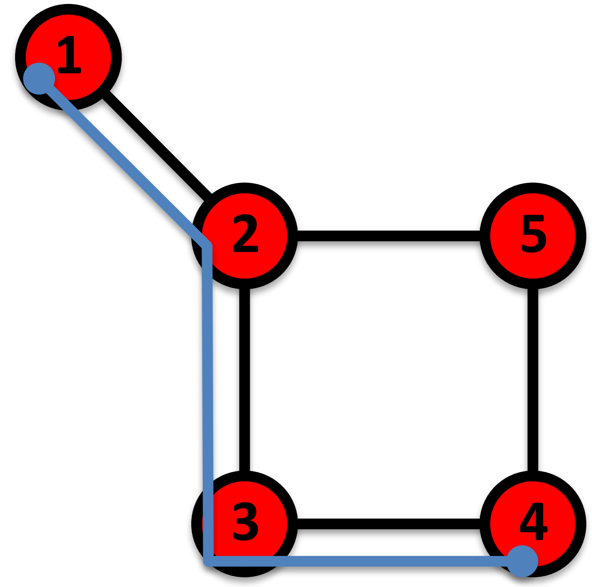


$l_{1\to 4} = 3$

$l_{1\to 5} = 2$

The path with the shortest length between two nodes (distance).

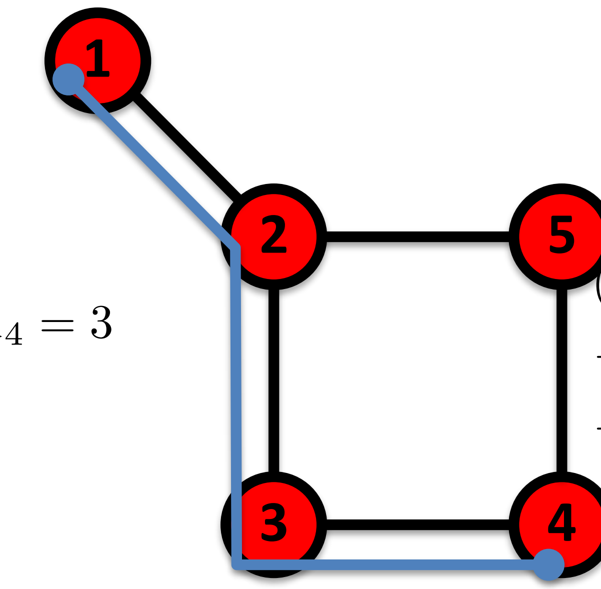
Diameter



$l_{1\to 4} = 3$

The longest shortest path in a graph

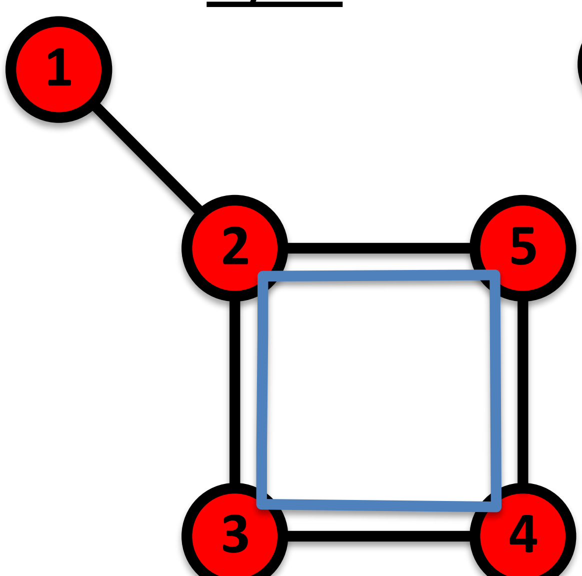
Average Path Length



$$(l_{1\to 2} + l_{1\to 3} + l_{1\to 4} + l_{1\to 5} + l_{2\to 3} + l_{2\to 4} + l_{2\to 5} + l_{3\to 4} + l_{3\to 5} + l_{4\to 5})/10 = 1.6$$

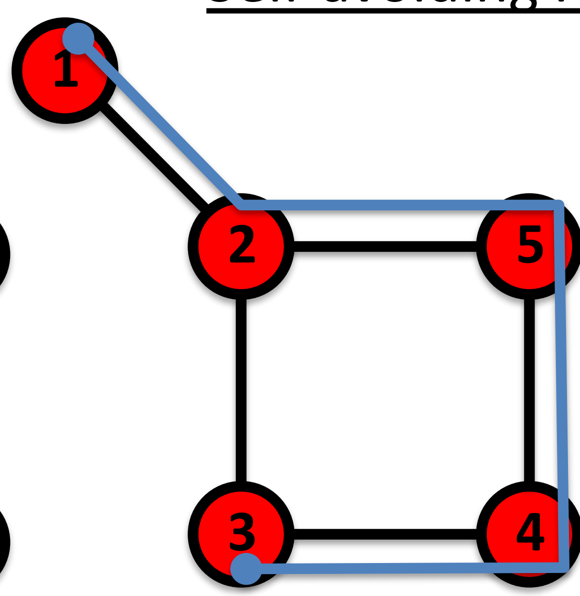
The average of the shortest paths for all pairs of nodes.

Cycle



A path with the same start and end node.

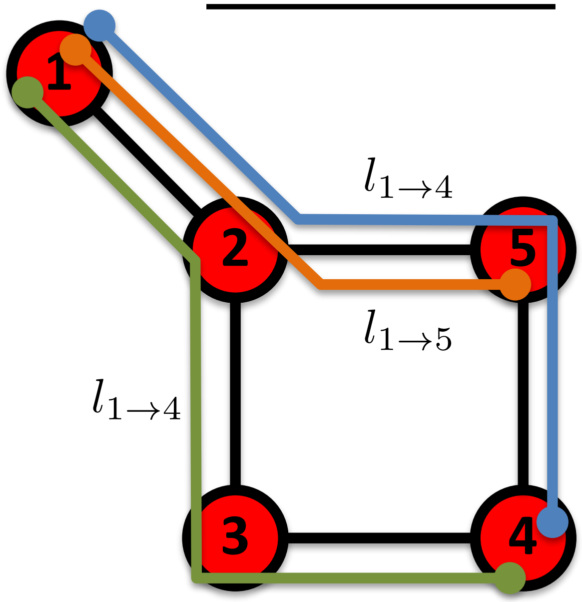
Self-avoiding Path



A path that does not intersect itself.

Slide from the Network Science book

Shortest Path

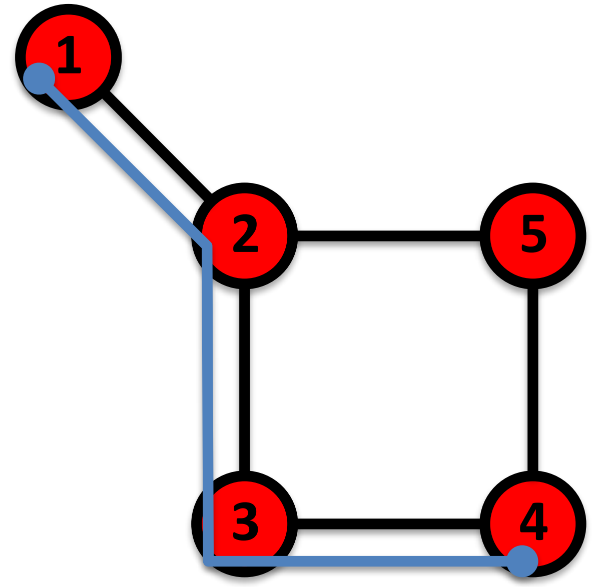


$l_{1\to 4} = 3$

$l_{1\to 5} = 2$

The path with the shortest length between two nodes (distance).

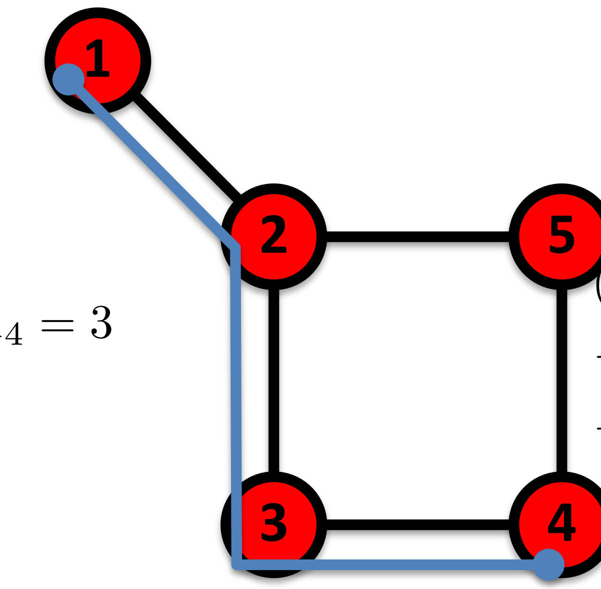
Diameter



$l_{1\to 4} = 3$

The longest shortest path in a graph

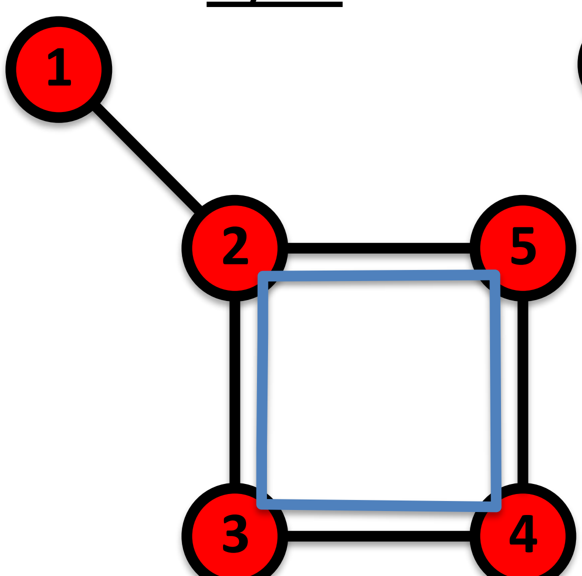
Average Path Length



$$(l_{1\to 2} + l_{1\to 3} + l_{1\to 4} + \\ + l_{1\to 5} + l_{2\to 3} + l_{2\to 4} + \\ + l_{2\to 5} + l_{3\to 4} + l_{3\to 5} + \\ + l_{4\to 5})/10 = 1.6$$

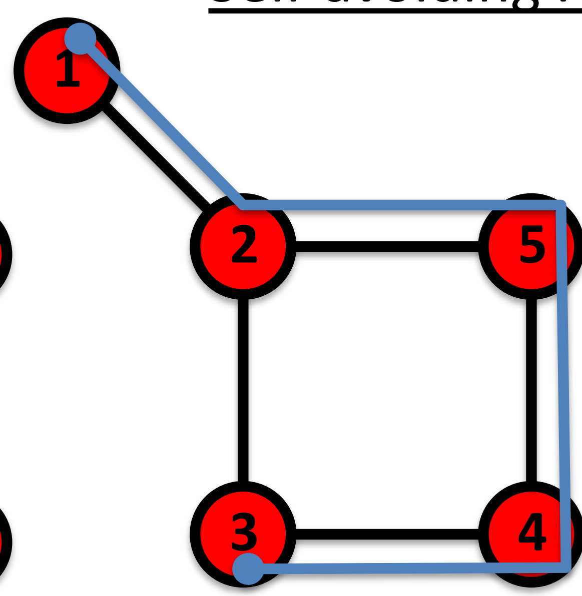
The average of the shortest paths for all pairs of nodes.

Cycle



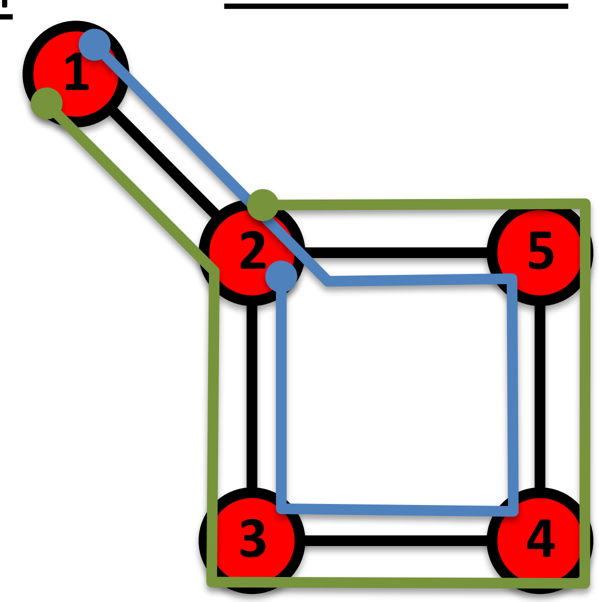
A path with the same start and end node.

Self-avoiding Path



A path that does not intersect itself.

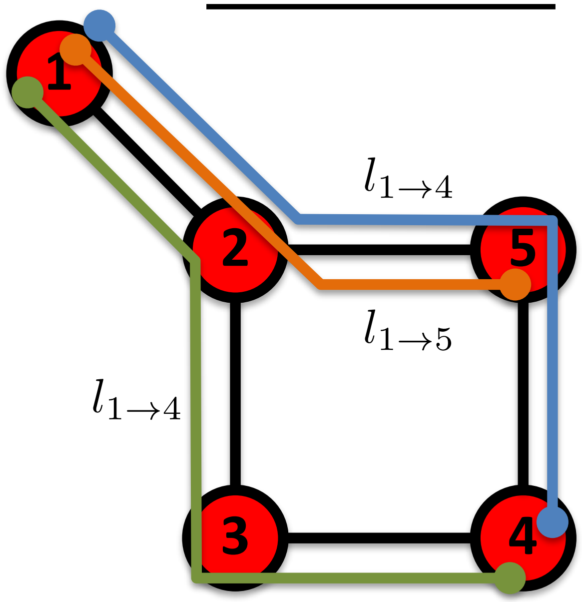
Eulerian Path



A path that traverses each link exactly once.

Slide from the Network Science book

Shortest Path

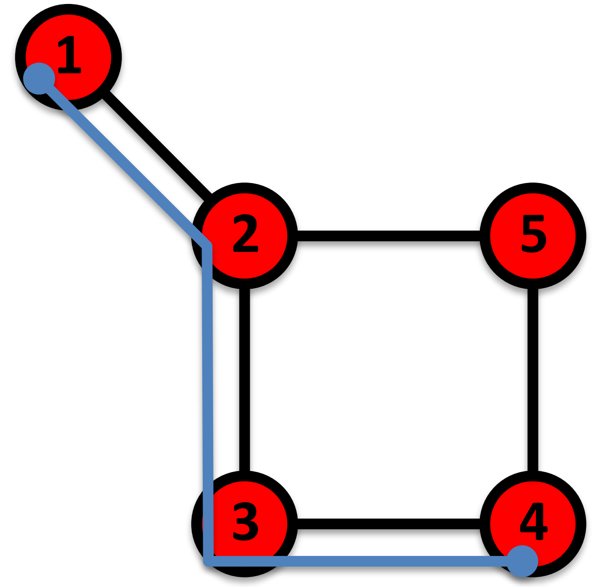


$l_{1\to4} = 3$

$l_{1\to5} = 2$

The path with the shortest length between two nodes (distance).

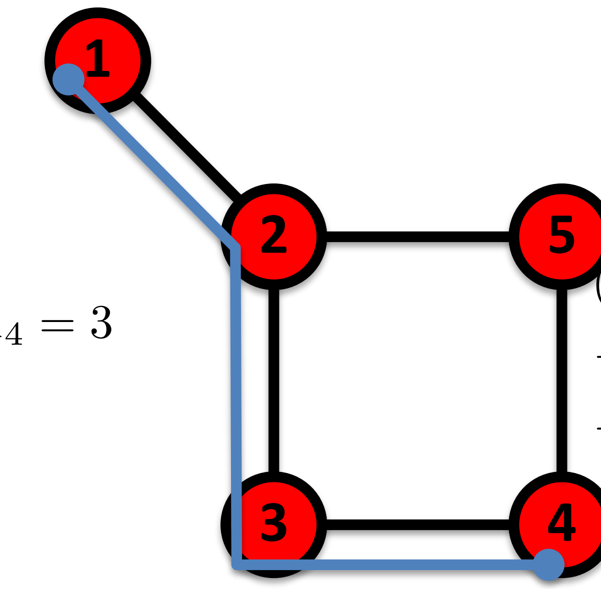
Diameter



$l_{1\to4} = 3$

The longest shortest path in a graph

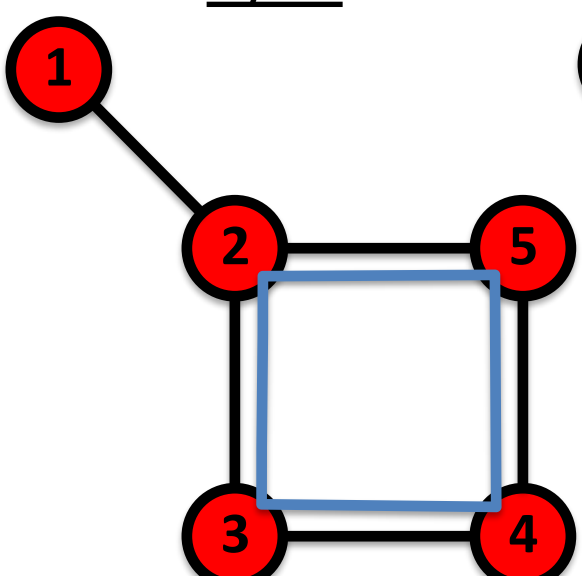
Average Path Length



$$(l_{1\to2} + l_{1\to3} + l_{1\to4} + l_{1\to5} + l_{2\to3} + l_{2\to4} + l_{2\to5} + l_{3\to4} + l_{3\to5} + l_{4\to5})/10 = 1.6$$

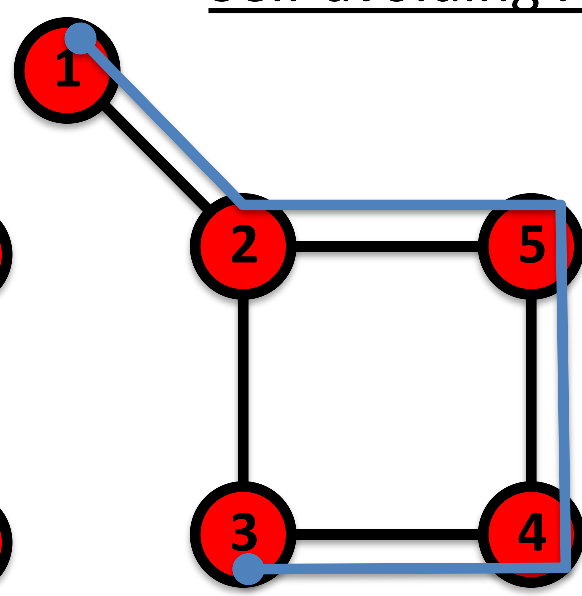
The average of the shortest paths for all pairs of nodes.

Cycle



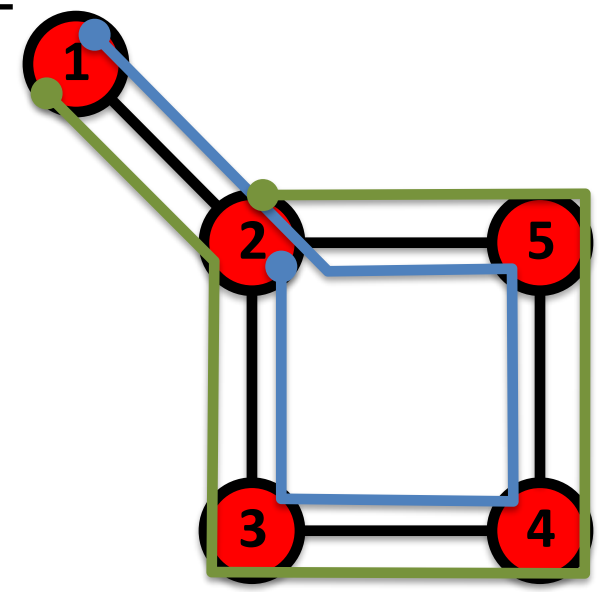
A path with the same start and end node.

Self-avoiding Path



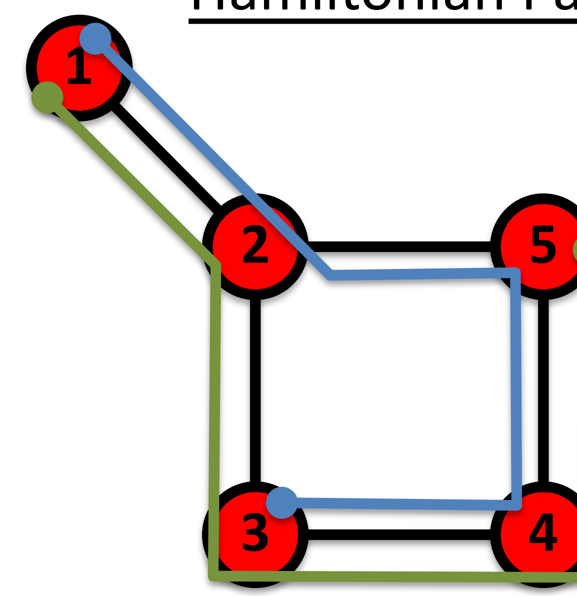
A path that does not intersect itself.

Eulerian Path



A path that traverses each link exactly once.

Hamiltonian Path

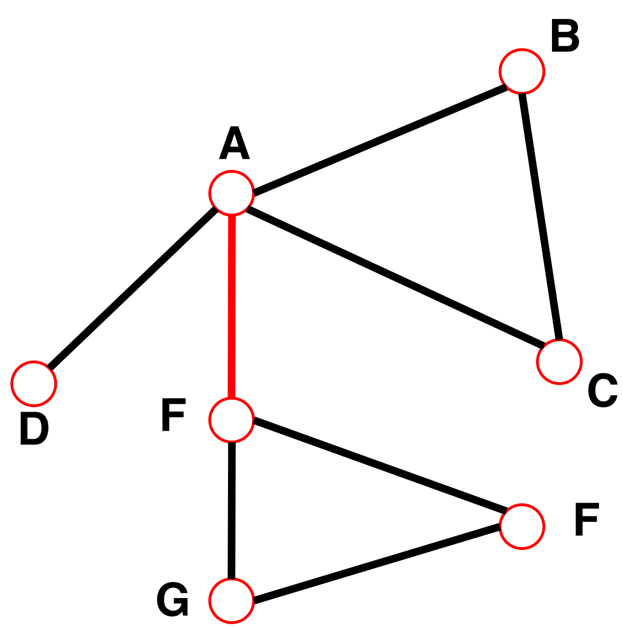


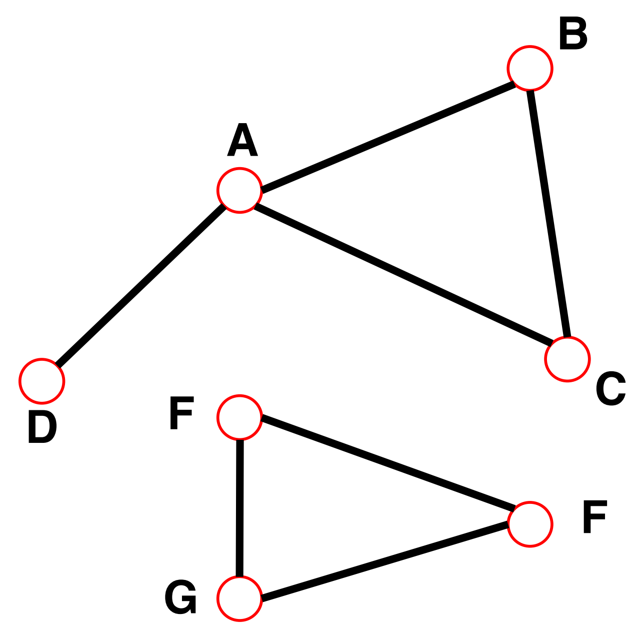
A path that visits each node exactly once.

Slide from the Network Science book

CONNECTIVITY OF UNDIRECTED GRAPHS

Connected (undirected) graph: any two vertices can be joined by a path.
A disconnected graph is made up by two or more connected components.





Largest Component:
Giant Component

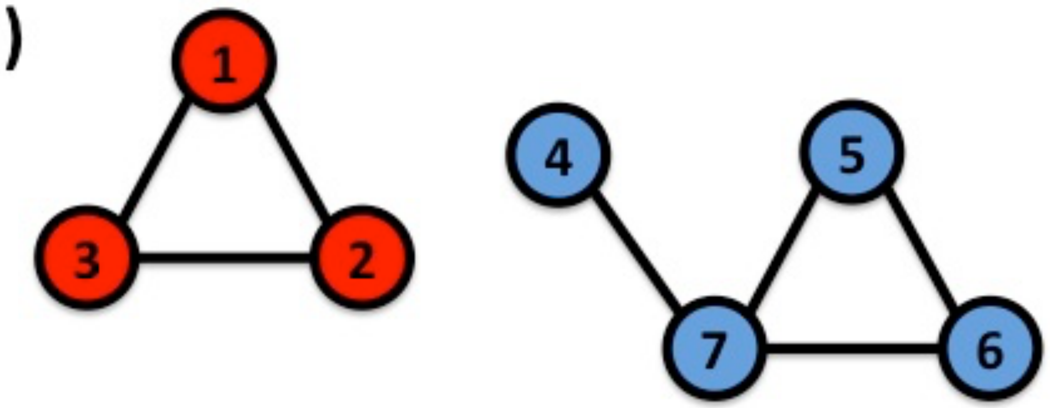
The rest: **Isolates**

Bridge: if we erase it, the graph becomes disconnected.

Slide from the Network Science book

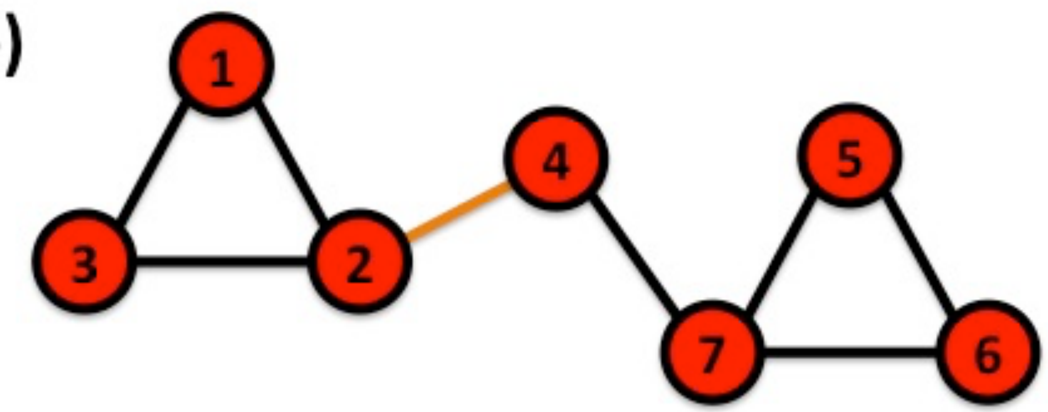
The adjacency matrix of a network with several components can be written in a block-diagonal form, so that nonzero elements are confined to squares, with all other elements being zero:

(a)



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

(b)



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

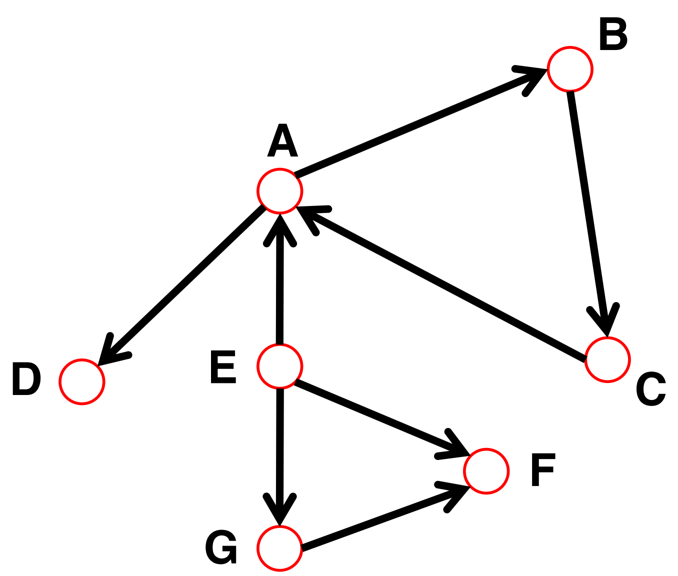
Network Science: Graph Theory

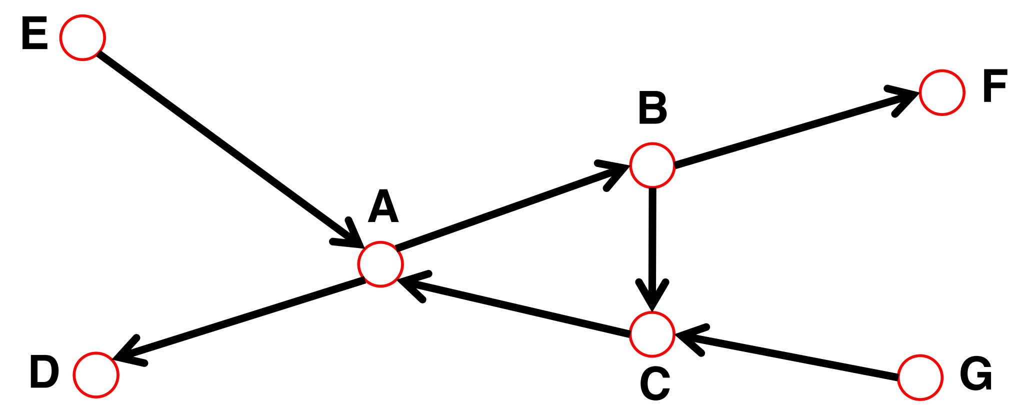
CONNECTIVITY OF DIRECTED GRAPHS

Strongly connected directed graph: has a path from each node to every other node and vice versa (e.g. AB path and BA path).

Weakly connected directed graph: it is connected if we disregard the edge directions.

Strongly connected components can be identified, but not every node is part of a nontrivial strongly connected component.





In-component: nodes that can reach the scc,

Out-component: nodes that can be reached from the scc.

Slide from the Network Science book

Which node in the network is the most “interesting”?

Which node in the network is the most “interesting”?

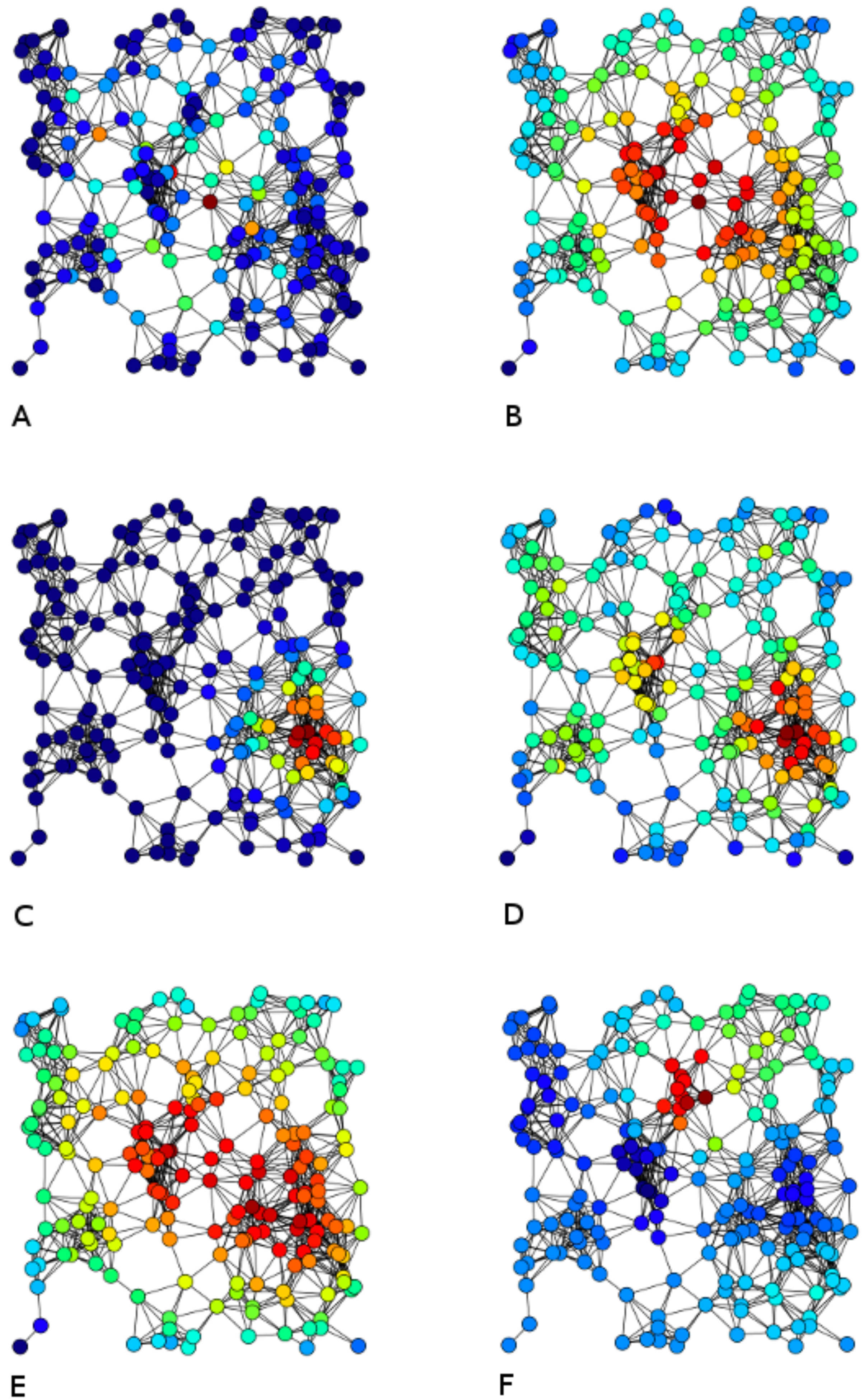
- What is the definition of “interesting”?

Which node in the network is the most “interesting”?

- What is the definition of “interesting”?
- How relationships are defined?

Which node in the network is the most “interesting”?

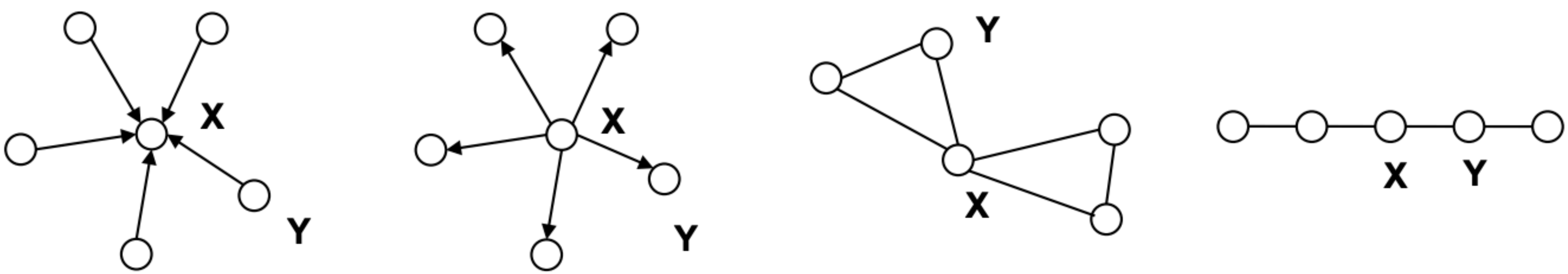
- What is the definition of “interesting”?
- How relationships are defined?
- In networks we define centrality measures to rank nodes according to some definitions for “interesting” or “central”
  - How influential a person is in a social network?
  - How critical a road in case of an emergency?
  - Which drugs to use for threading a disease?



Depending on the definition different nodes turn out to be more central.

A) Betweenness centrality
B) Closeness centrality
C) Eigenvector centrality
D) Degree centrality
E) Harmonic centrality
F) Katz centrality

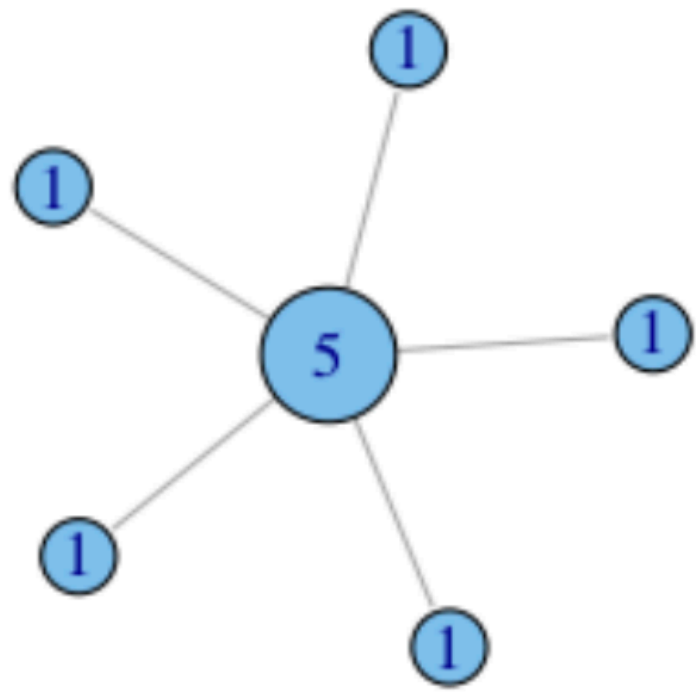
In each of the following networks, X has higher centrality than Y according to a particular measure



indegree outdegree betweenness closeness

Degree Centrality (Undirected)

He or she who has many friends is most important.



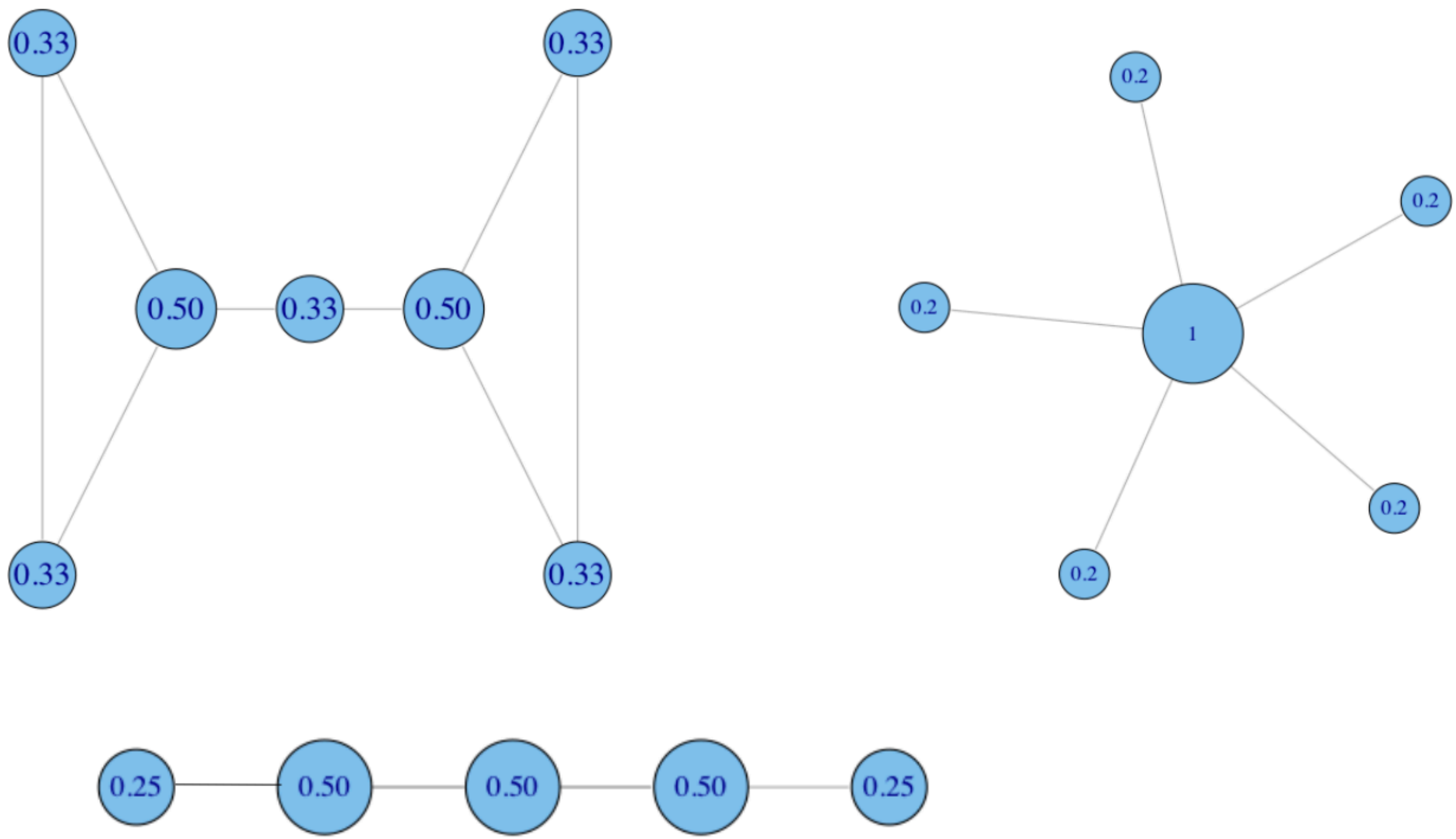
When is the number of connections the best centrality measure?

- people who will do favors for you
- people you can talk to / have coffee with

https://cs.brynmawr.edu/Courses/cs380/spring2013/section02/slides/05_Centrality.pdf

Normalization

divide degree by the max. possible, i.e. (N-1)



https://cs.brynmawr.edu/Courses/cs380/spring2013/section02/slides/05_Centrality.pdf

Graph centralization

- We can extend the definition of node centrality to whole graph.
- This quantifies the variation of centrality among nodes

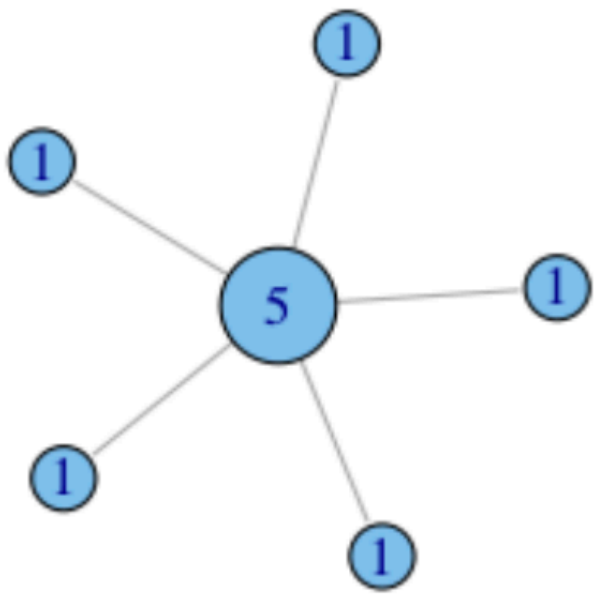
$$C_D(G) = \frac{\boxed{\sum_{i=1}^{|V|} [C_D(v*) - C_D(v_i)]}}{\boxed{|V|^2 - 3|V| + 2}}$$

For each node how far their centrality measure from the maximum value observed on the graph

theoretically largest such sum of differences in any network of the same size: (N-1)*(N-1-1)

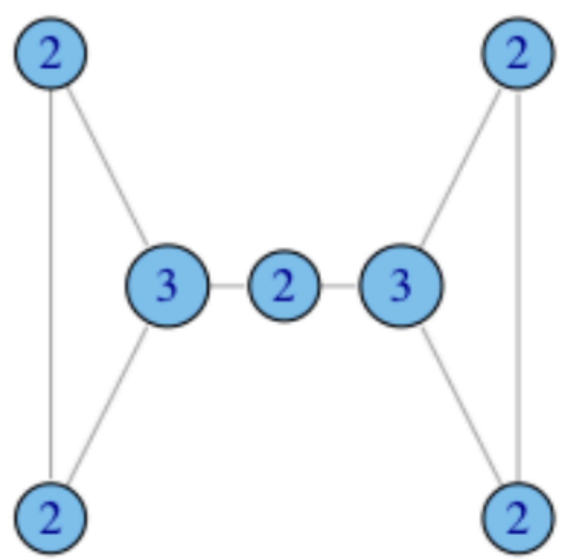
Remind that node has 0 difference with itself

Example degree centralization



$C_D = 1.0$

$C_D = 0.167$

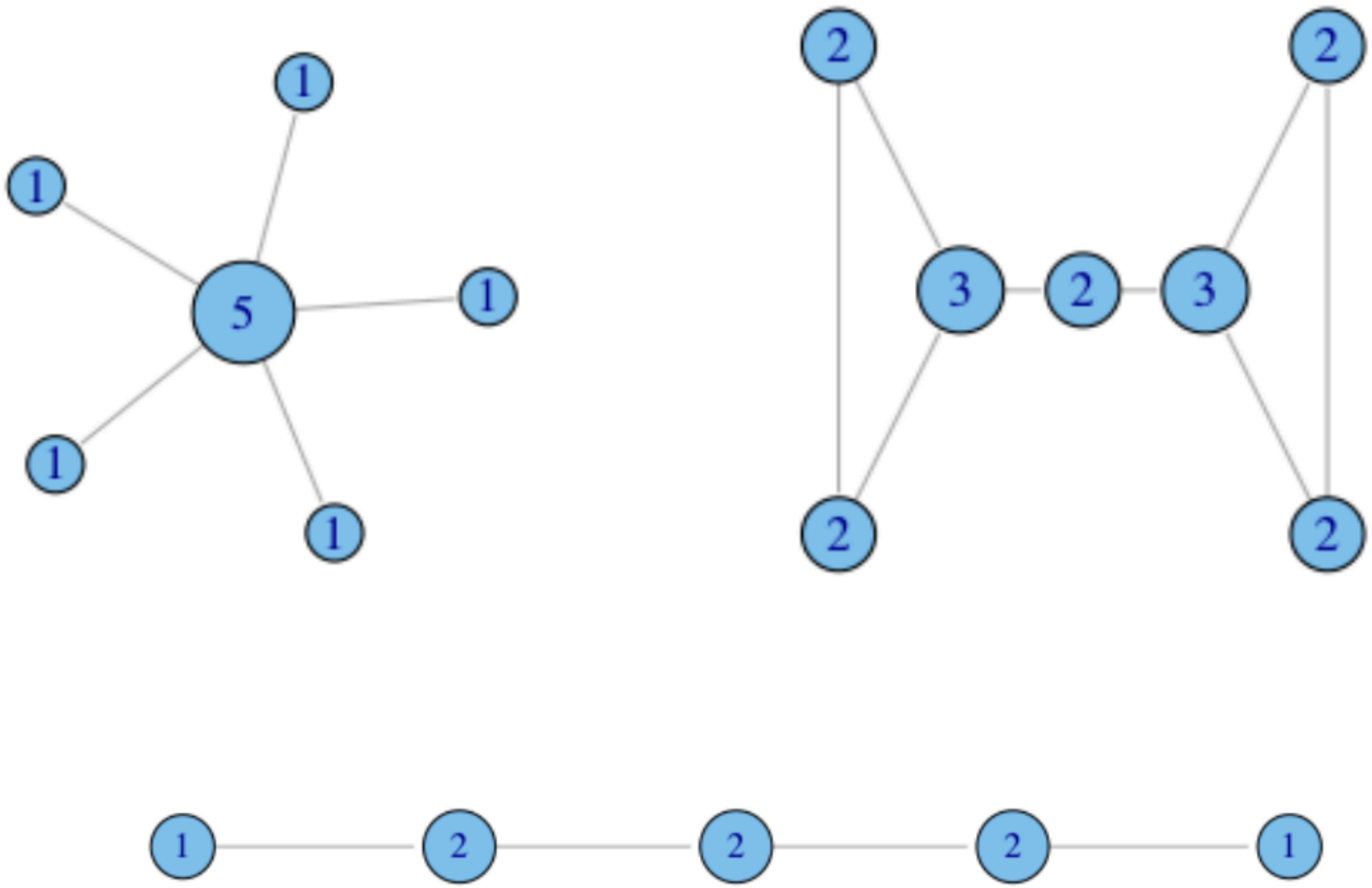


$C_D = 0.167$

https://cs.brynmawr.edu/Courses/cs380/spring2013/section02/slides/05_Centrality.pdf

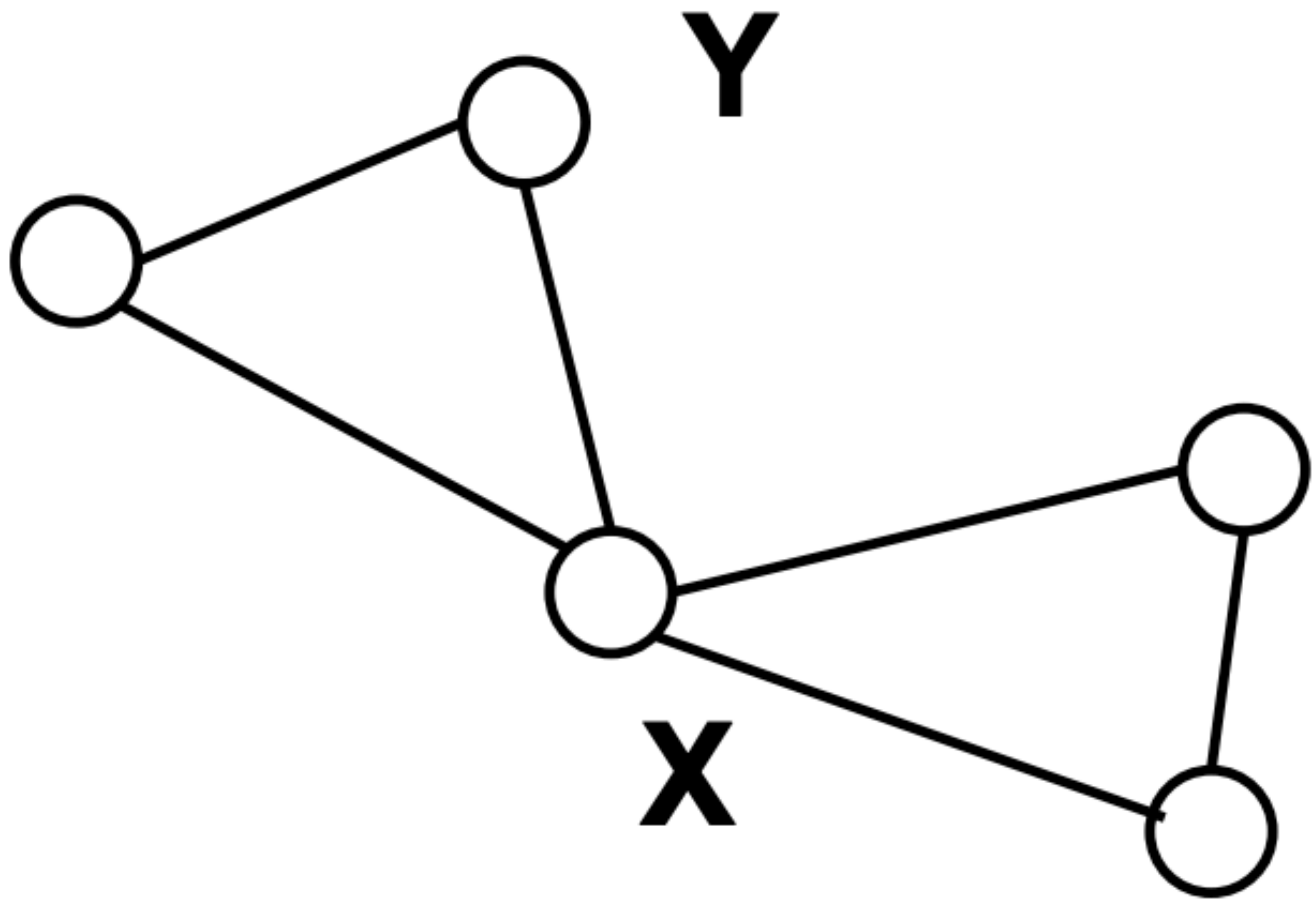
What are we missing by looking only degree?

In what ways does degree fail to capture centrality in the following graphs?



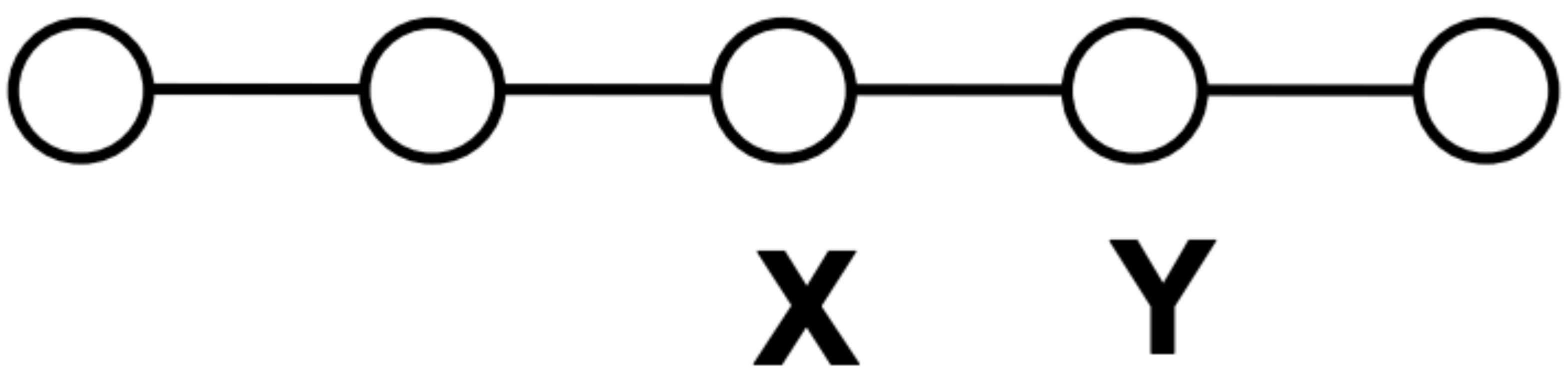
https://cs.brynmawr.edu/Courses/cs380/spring2013/section02/slides/05_Centrality.pdf

Brokerage not captured by degree



Betweenness centrality intuition

How many pairs of individuals would have to go through you in order to reach one another in the minimum number of hops?



Linton Freeman

Article Talk

From Wikipedia, the free encyclopedia

Linton Clarke Freeman (1927 – August 17, 2018) was an American structuralist sociologist known for his pioneering work in social networks. He was an emeritus professor of Sociology at the University of California, Irvine.[1] Freeman developed the first measure of betweenness centrality. He was the founding editor of the journal *Social Networks*[2] which began publishing in 1979.[3]

https://cs.brynmawr.edu/Courses/cs380/spring2013/section02/slides/05_Centrality.pdf

Betweenness centrality

$$C_B(v) = \sum_{s \neq v \neq t \in V} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

# of shortest path connecting s and t
that goes through v

# of shortest path connecting s and t

It's common to normalize betweenness centrality by dividing
# of pair of vertices excluding v: (N-1)(N-2)

Betweenness centrality

$$C_B(v) = \sum_{s \neq v \neq t \in V} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

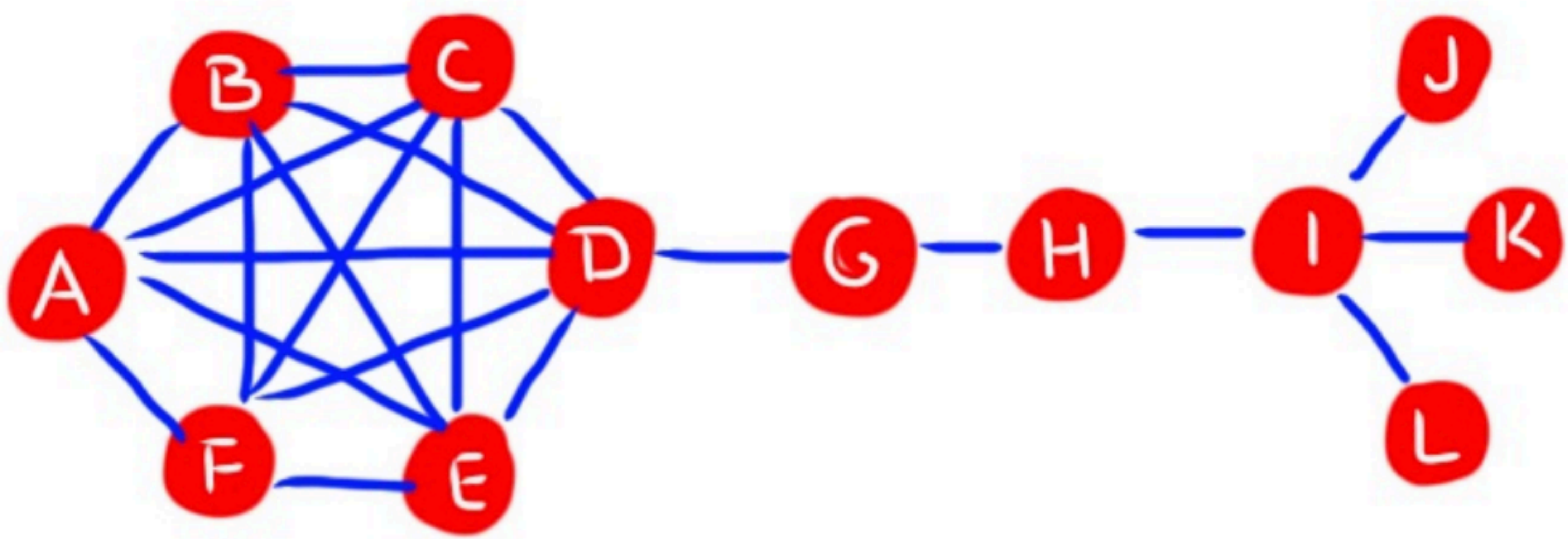
# of shortest path connecting s and t that goes through v

# of shortest path connecting s and t

It's common to normalize betweenness centrality by dividing # of pair of vertices excluding v: (N-1)(N-2)

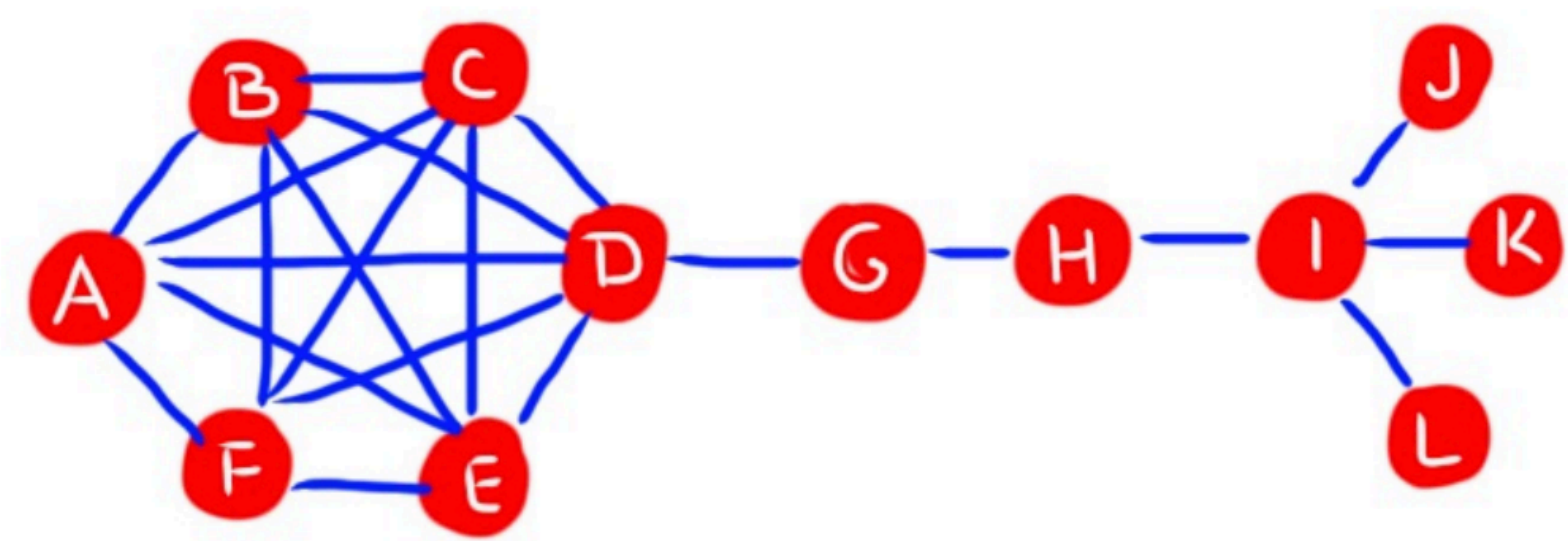
One caveat!

From a calculation aspect, both betweenness and closeness centralities of all vertices in a graph involve calculating the shortest paths between all pairs of vertices on a graph, which requires **O(V^3)** time with the Floyd–Warshall algorithm. However, on sparse graphs, Johnson's algorithm may be more efficient, taking **O(V^2 log(V)+VE)** time.



https://cs.brynmawr.edu/Courses/cs380/spring2013/section02/slides/05_Centrality.pdf

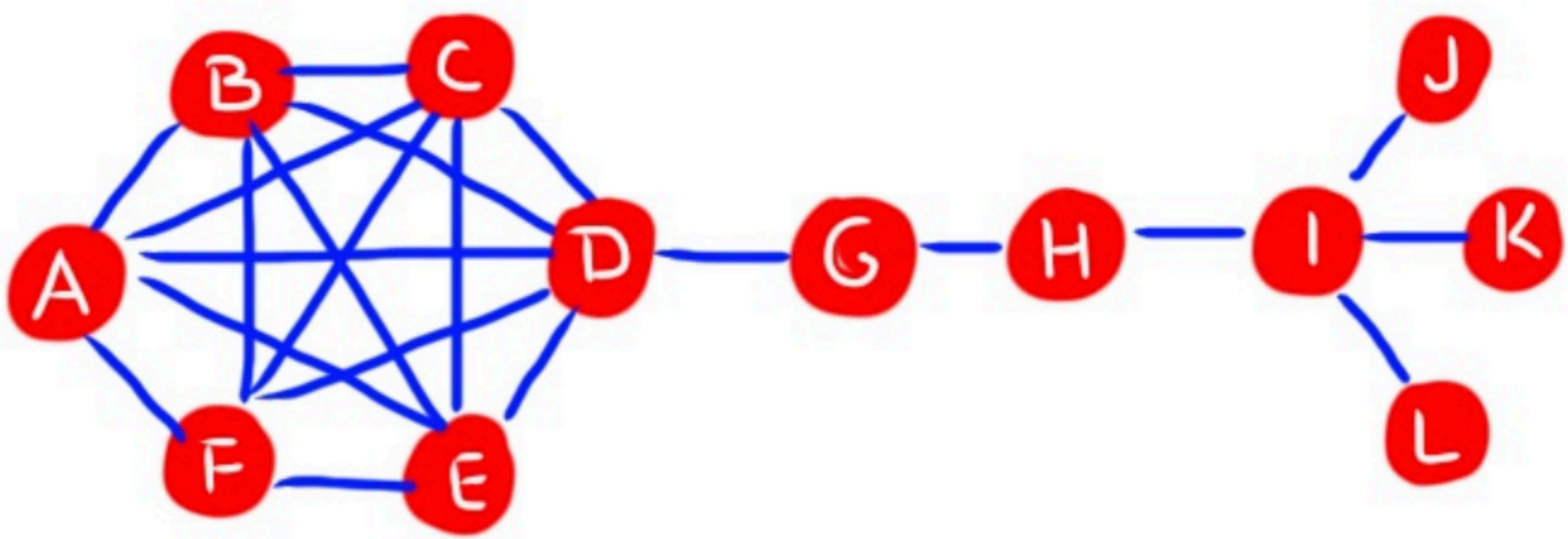
Find a node that has high betweenness but low degree



https://cs.brynmawr.edu/Courses/cs380/spring2013/section02/slides/05_Centrality.pdf

Find a node that has high betweenness but low degree

Find a node that has low betweenness but high degree



https://cs.brynmawr.edu/Courses/cs380/spring2013/section02/slides/05_Centrality.pdf

Closeness centrality

Closeness centrality (or closeness) of a node is the average length of the shortest path between the node and all other nodes in the graph. Thus the more central a node is, the closer it is to all other nodes.

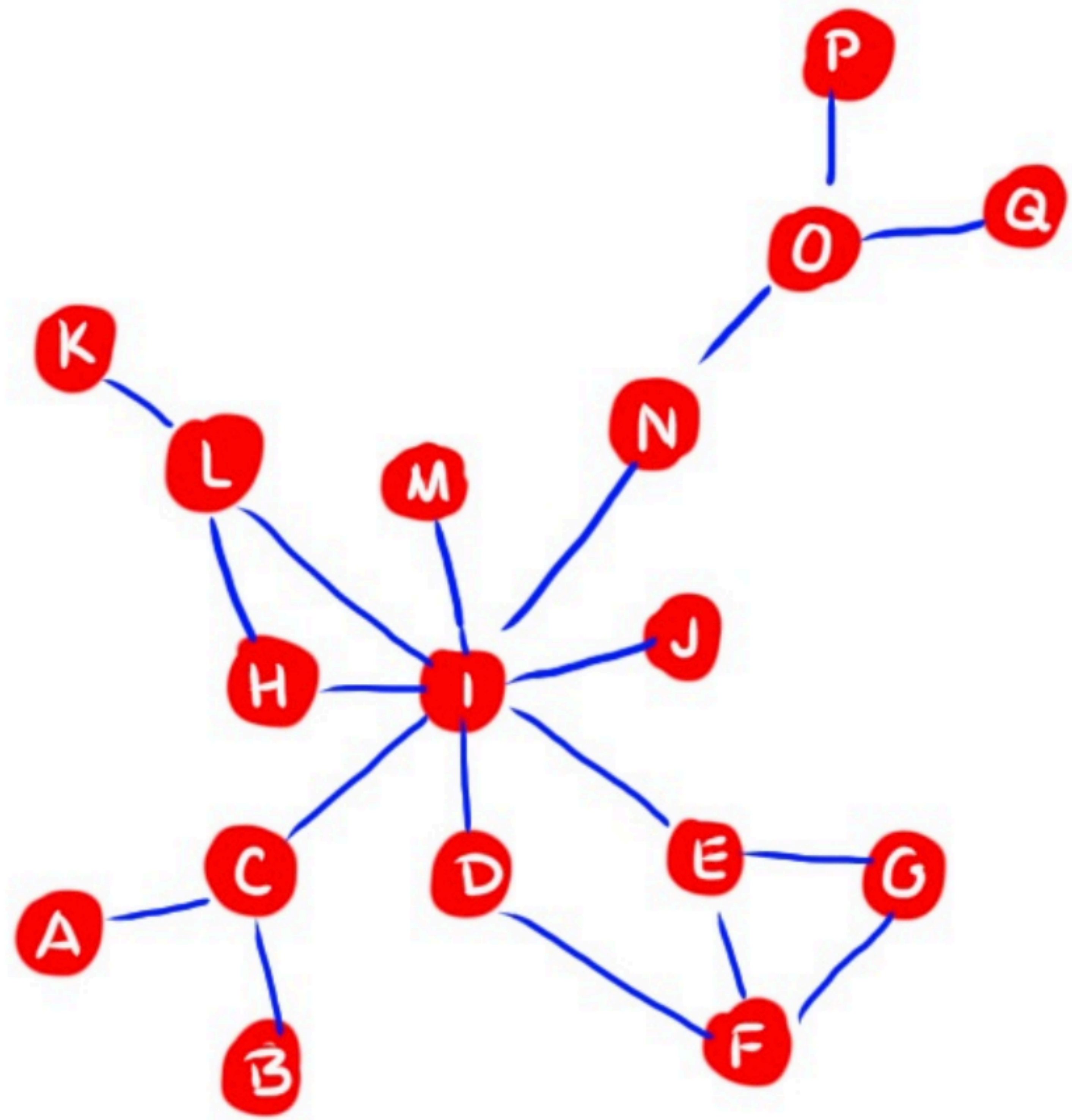
$$C(x) = \frac{1}{\sum_y d(y,x)}$$

Distance between nodes x and y

$$C'(x) = \frac{C(x)}{N-1}$$

Normalized closeness centrality

Which node has relatively high degree but low closeness?



https://cs.brynmawr.edu/Courses/cs380/spring2013/section02/slides/05_Centrality.pdf


Let's play a game