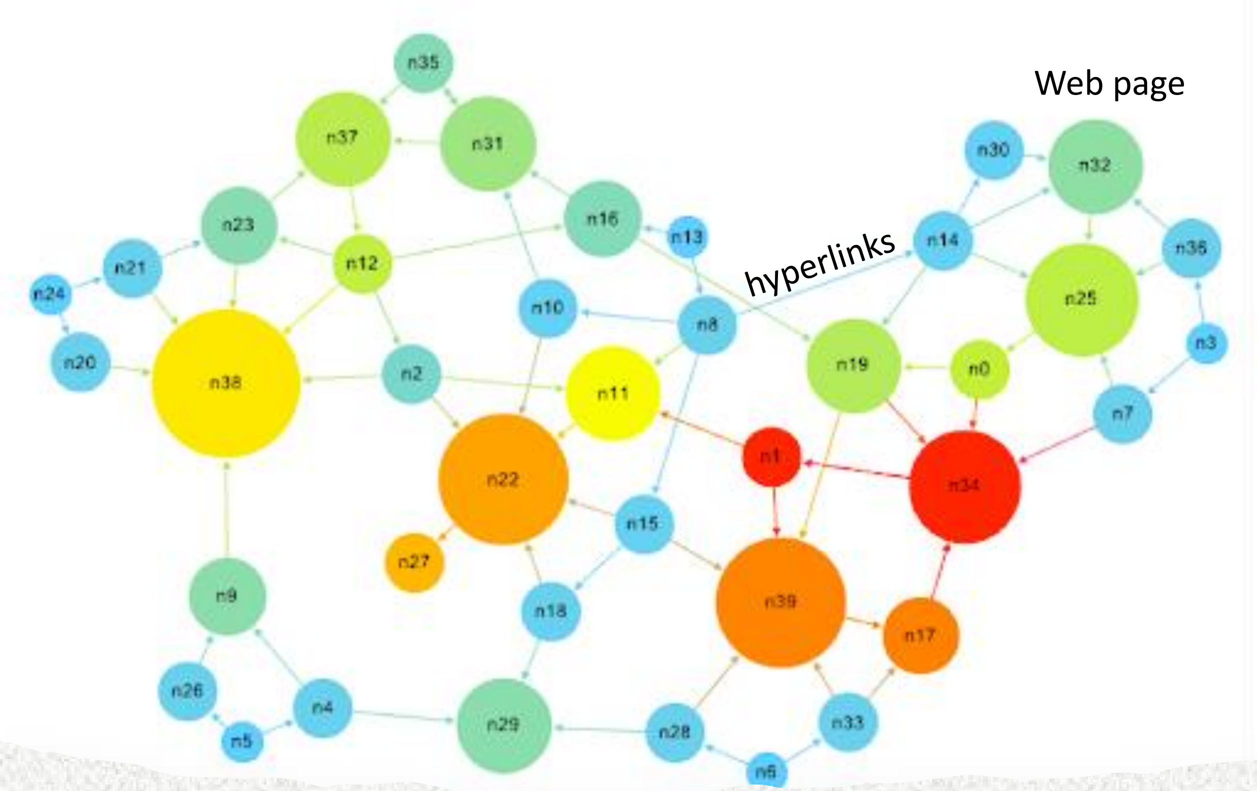
Network Analysis: sociometric factors

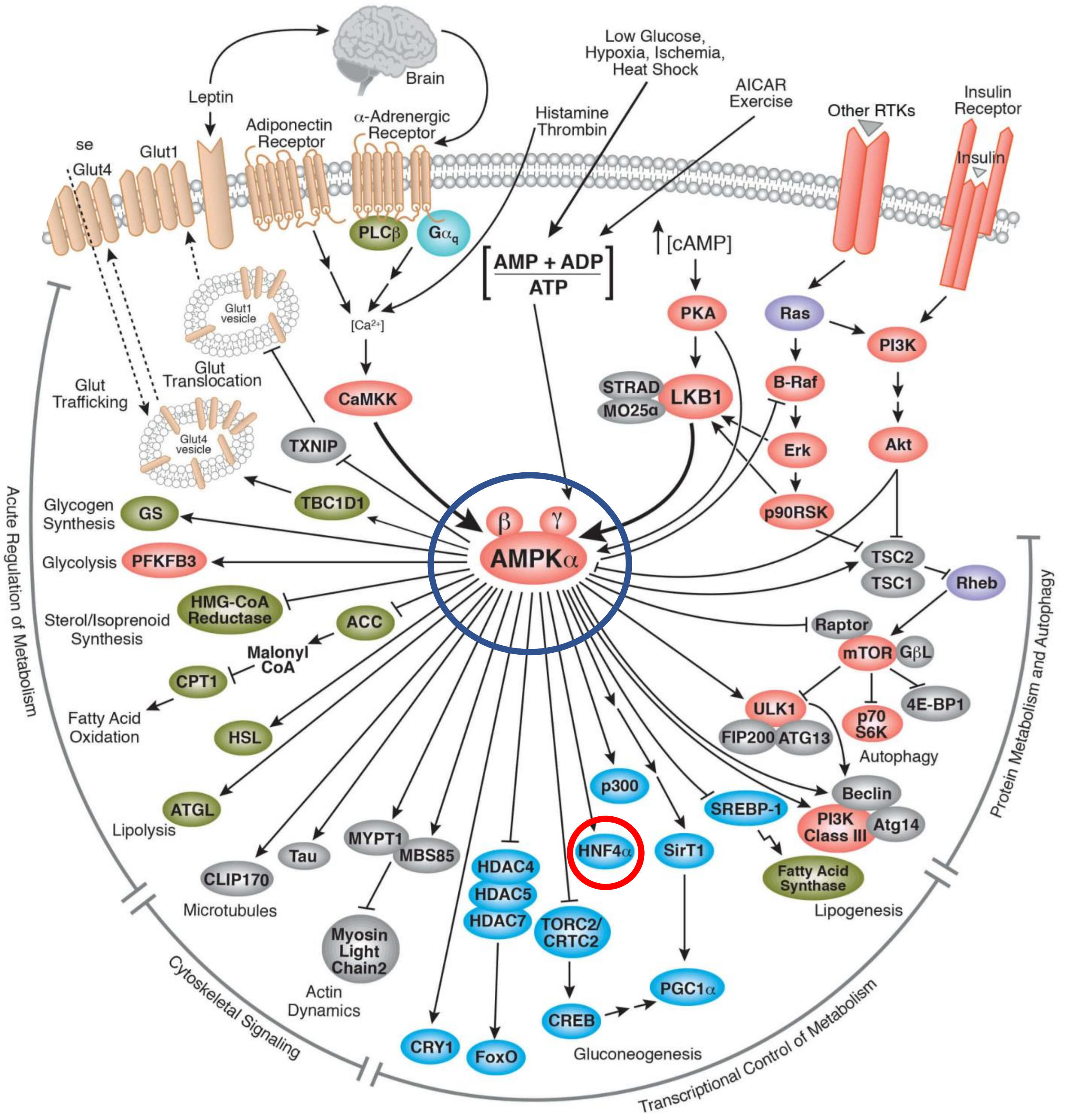
- **Creation of directed and undirected networks**
- **Sociometric factors**
- **Random networks and their properties**
- **Small world phenomenon**



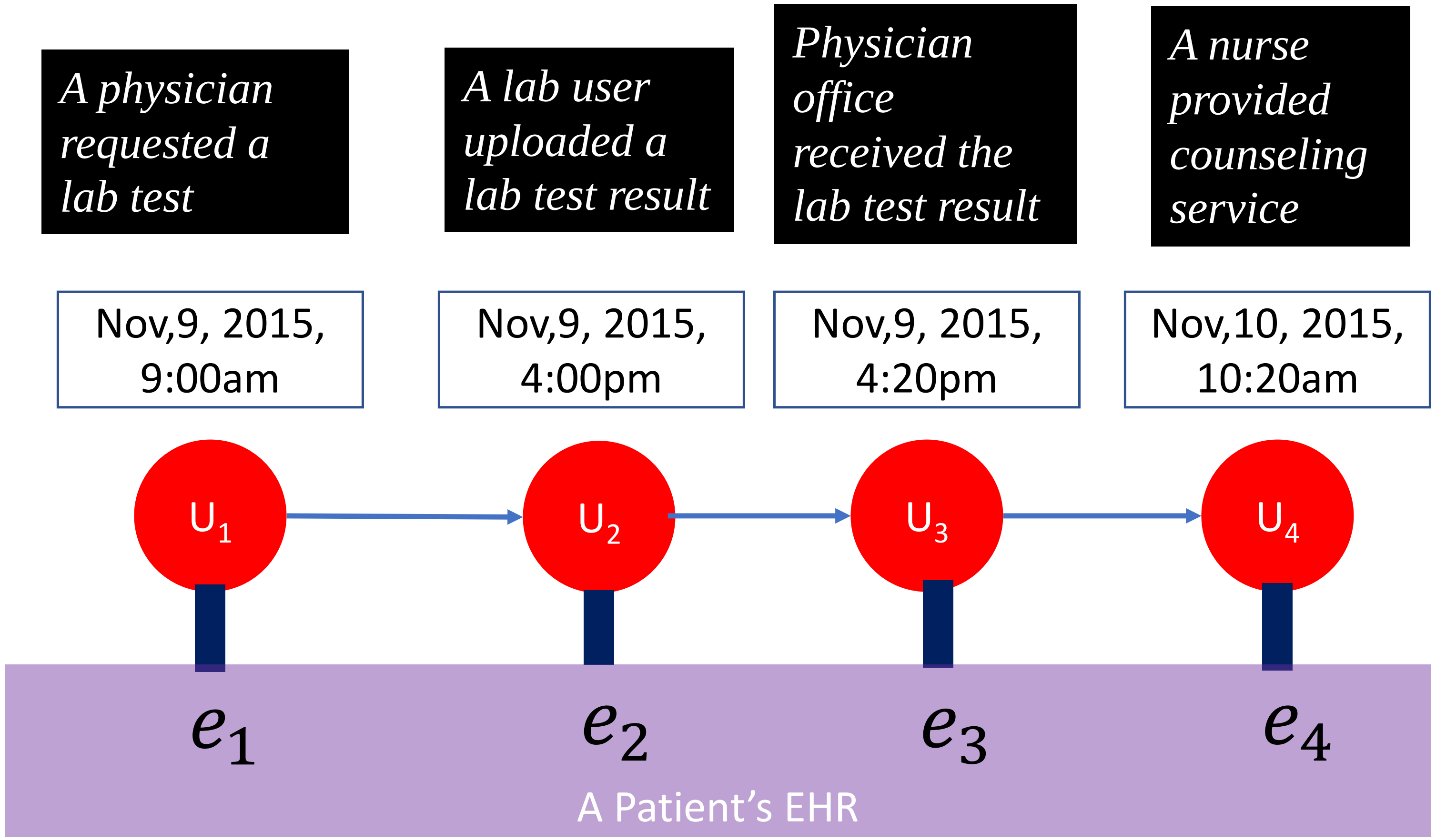
Directed networks – world wide web (hyperlinks between edges)

Metabolic networks (biochemical reactions between molecules)

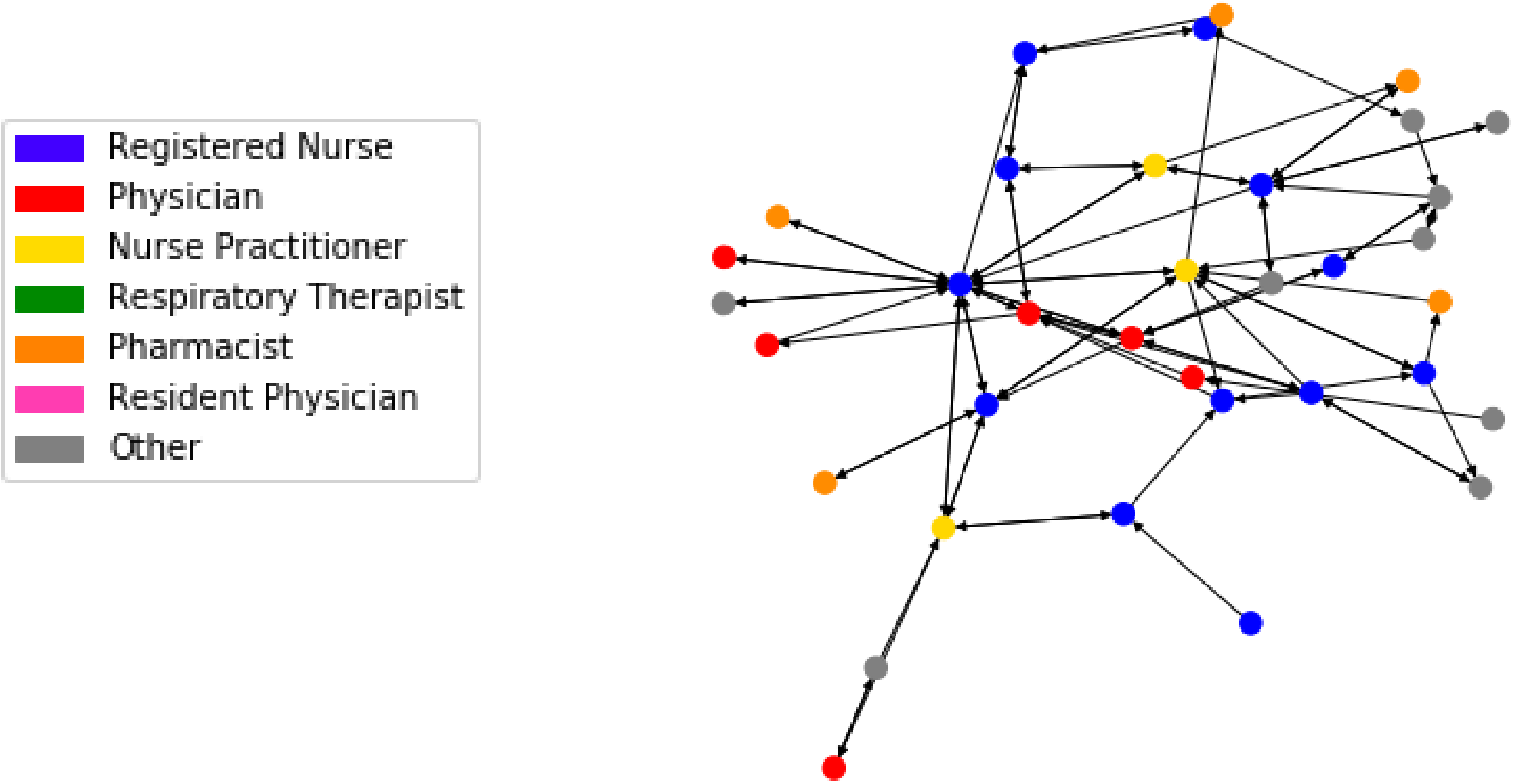
- The links are the biochemical reactions that take place between these molecules
- AMP-activated protein kinase (AMPK)
  - It is an enzyme that plays a role in cellular energy homeostasis, largely to activate glucose and fatty acid uptake and oxidation when cellular energy is low
  - AMPK directly impacts HNF4alpha and represses its transcriptional activity



EHR audit log data: User-EHR interactions
User: healthcare worker
EHR: electronic health record



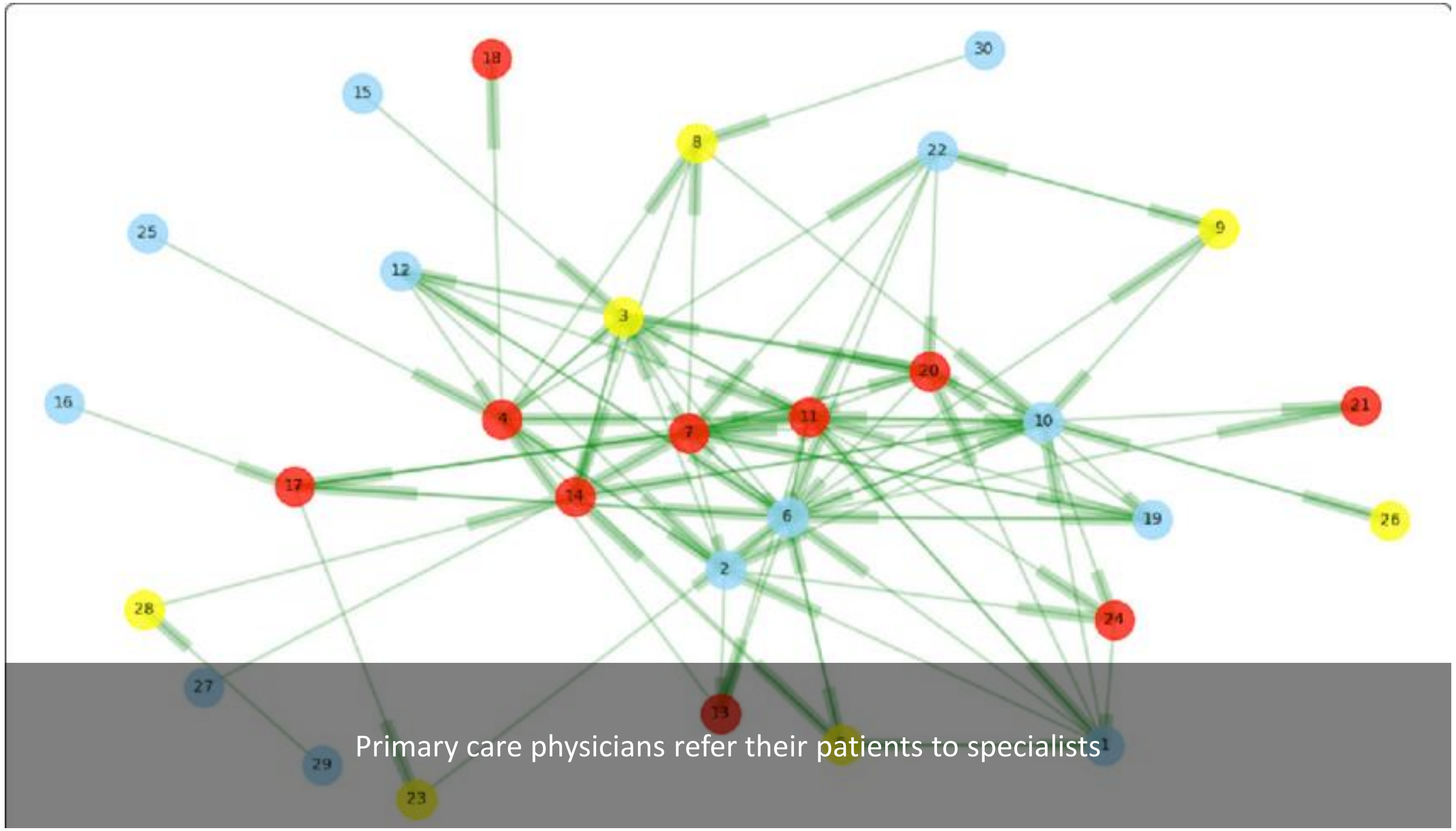
Networks of healthcare workers



Health insurance claims

- A request for payment that you or your health care provider submits to your health insurer when you get items or services you think are covered
- A claim contain patient, physician, healthcare service, and referral information

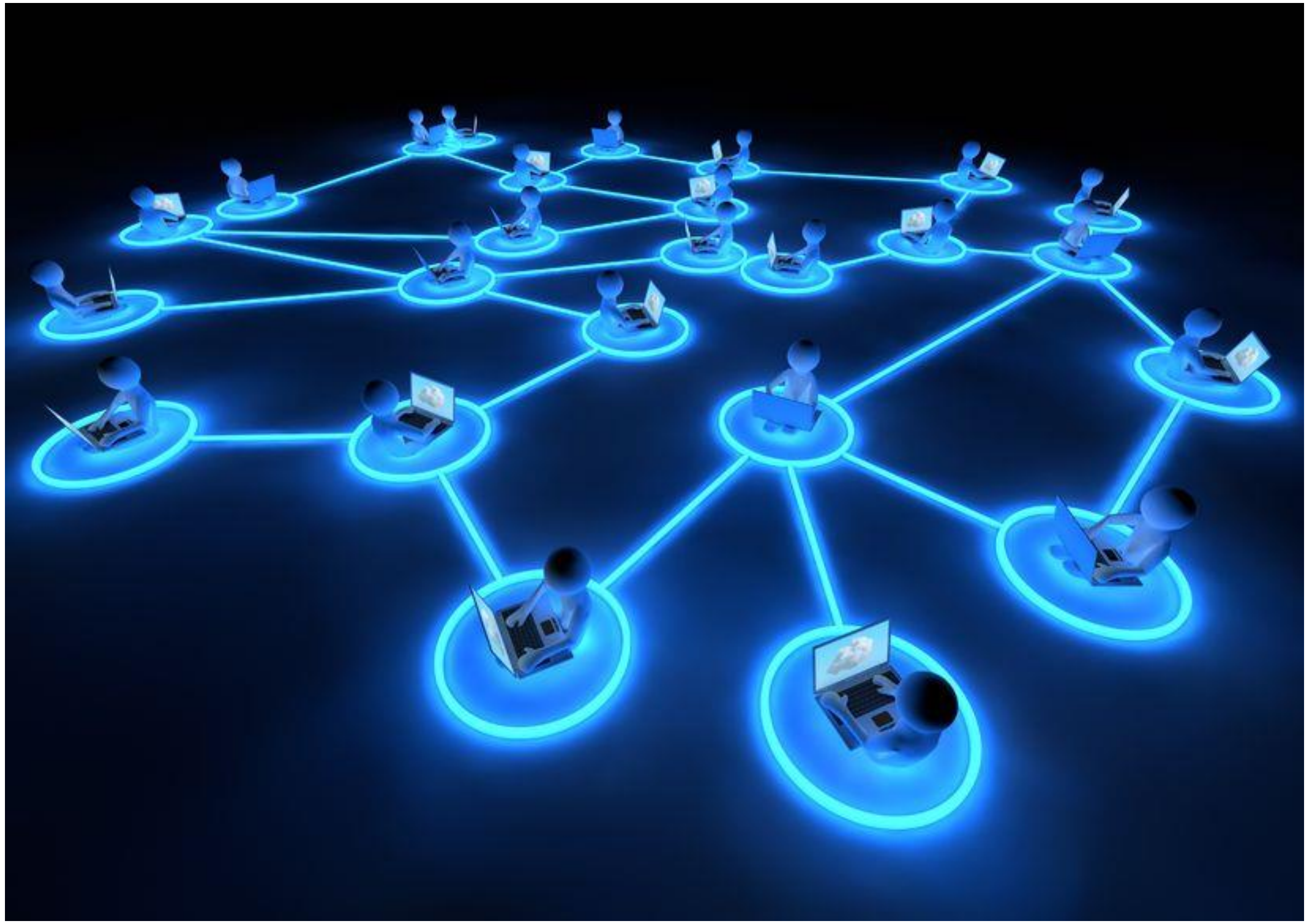
Patient referral network



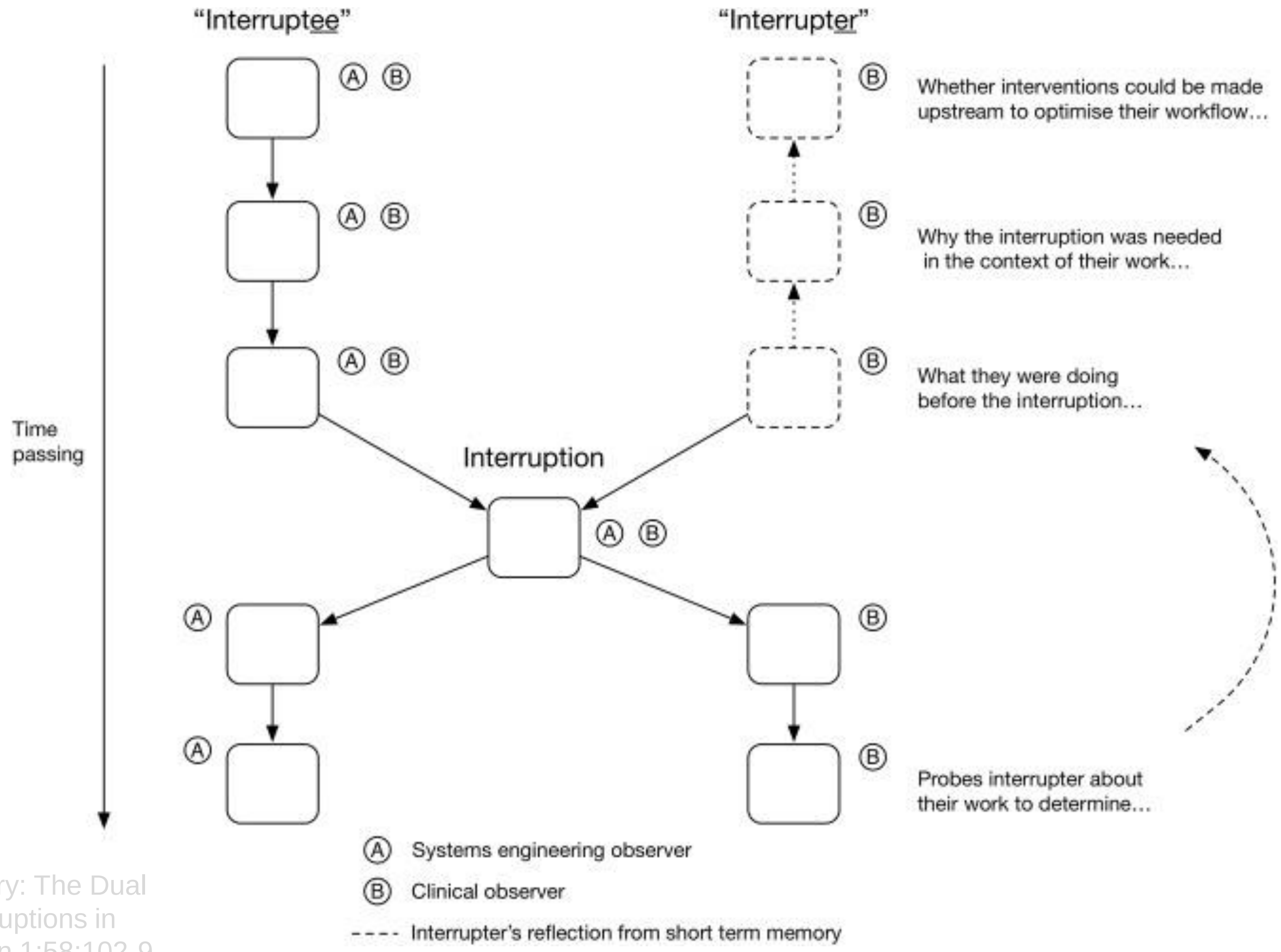
Primary care physicians refer their patients to specialists

Undirected networks

- The topology of connections in a computer network or a digital social network
- The graph is undirected because we can assume that if one device is connected to another, then the second one is also connected to the first
- The topology of digital social networks, where each friend of someone is that someone's friend



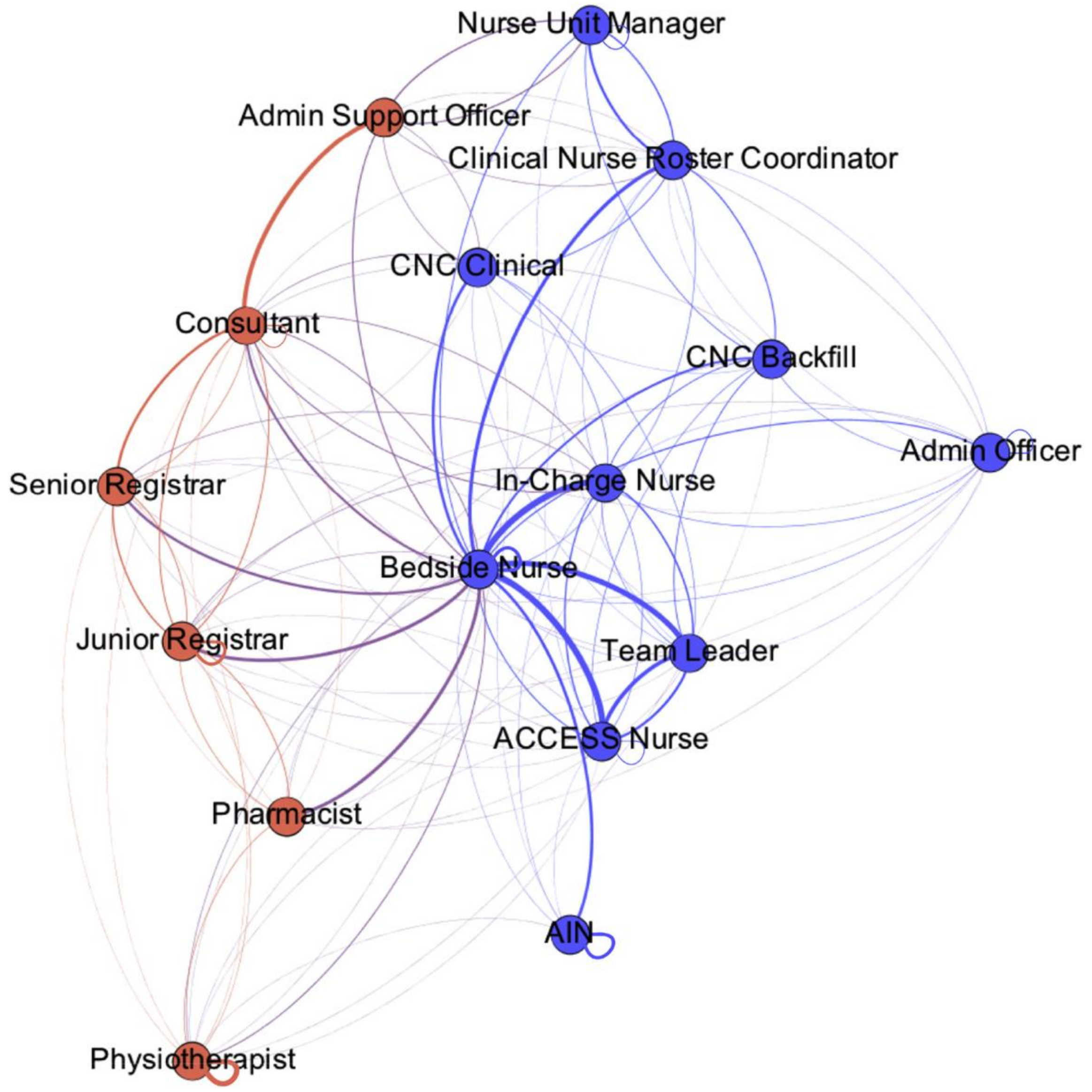
Dual Perspectives Method - Observing clinical interruptions



McCurdie T, et.al. Two sides to every story: The Dual Perspectives Method for examining interruptions in healthcare. Applied ergonomics. 2017 Jan 1;58:102-9.

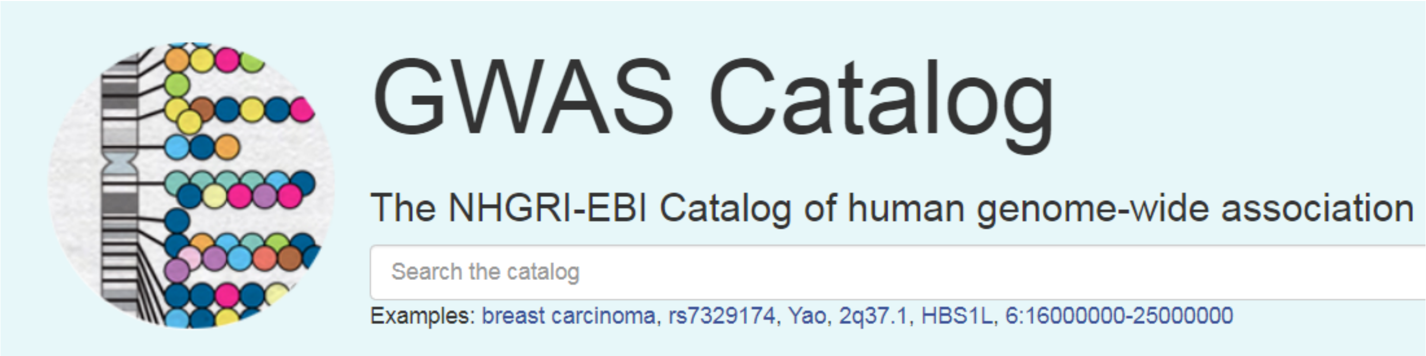
Inter-disciplinary interruption network observed in the ICU

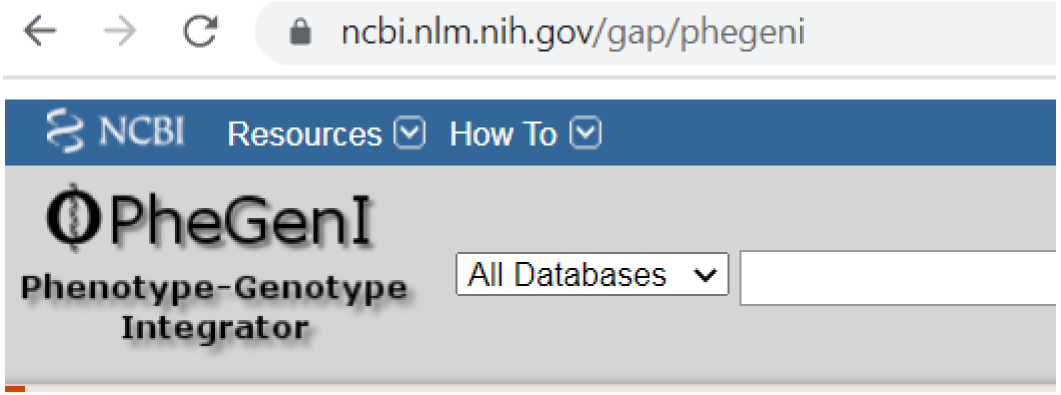
- The relationships are based on interruptions, upon which care work can continue
- Dependency between the Bedside Nurse, Team Leader, ACCESS (Assistance, Coordination, Contingency, Education, Supervision, Support) Nurse, and In-Charge Nurse roles
- These dependencies are potential focal points for intervention



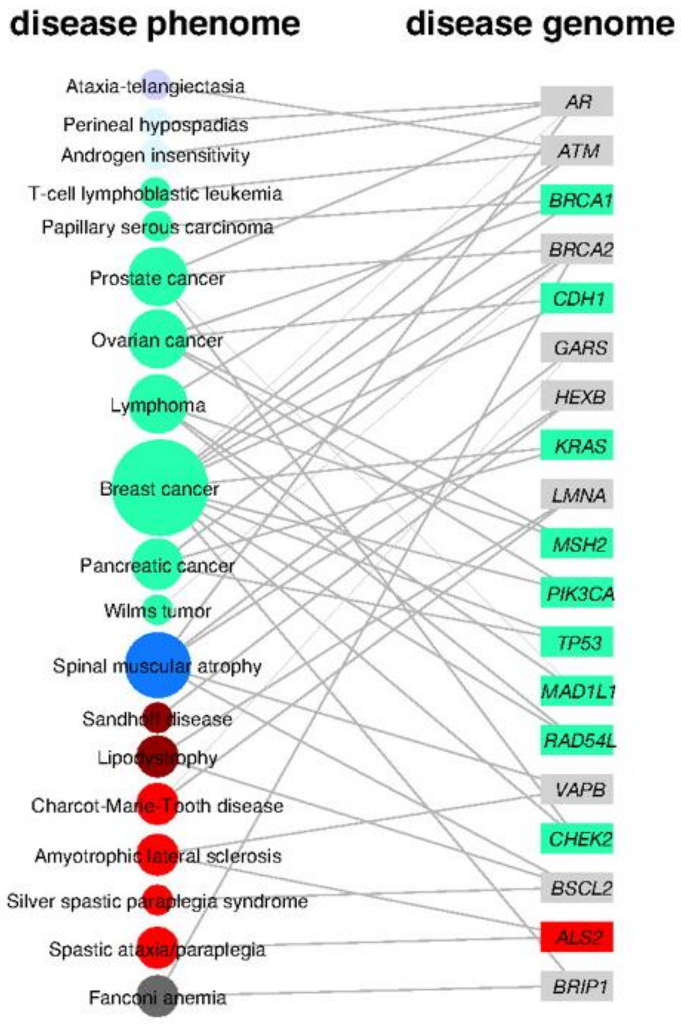
McCurdie T, et.al. Applying social network analysis to the examination of interruptions in healthcare. Applied ergonomics. 2018 Feb 1;67:50-60.

Data resources to create human disease network





GWAS and PheWAS study findings

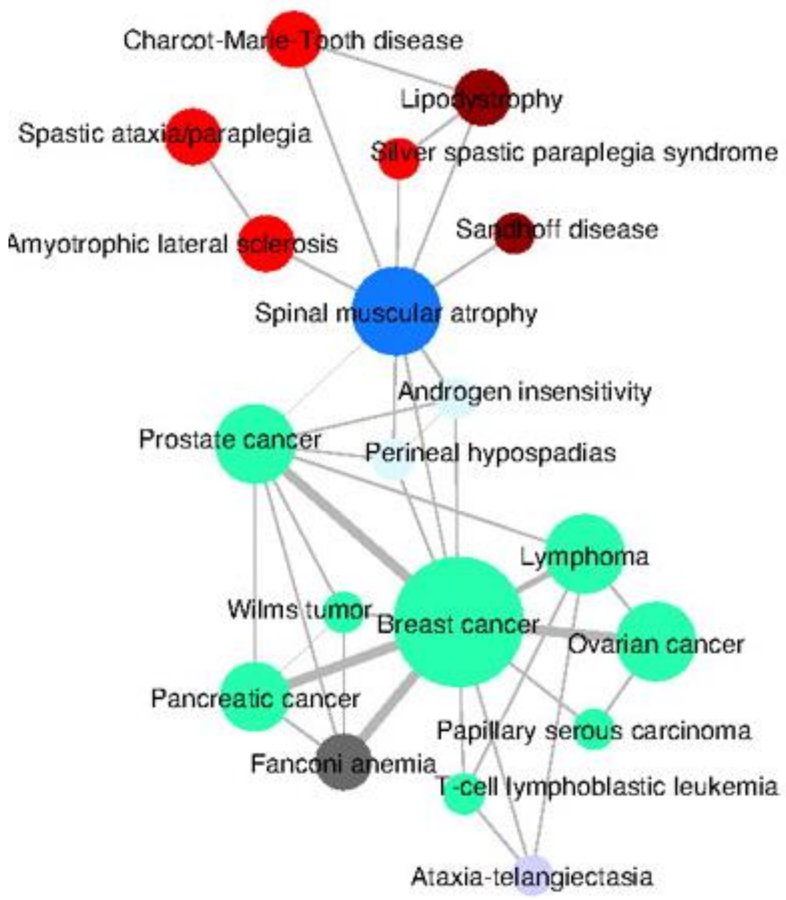


- **Circles:** disease phenome (disorder)
- **Rectangles:** disease genes
- **Links:** a link is placed between a disorder and a disease gene if mutations in that gene lead to the specific disorder
- **The size of a circle** is proportional to the number of genes participating in the corresponding disorder
- **The color** corresponds to the disorder class to which the disease belongs

Goh KI, et.al. The human disease network. Proceedings of the National Academy of Sciences. 2007 May 22;104(21):8685-90.

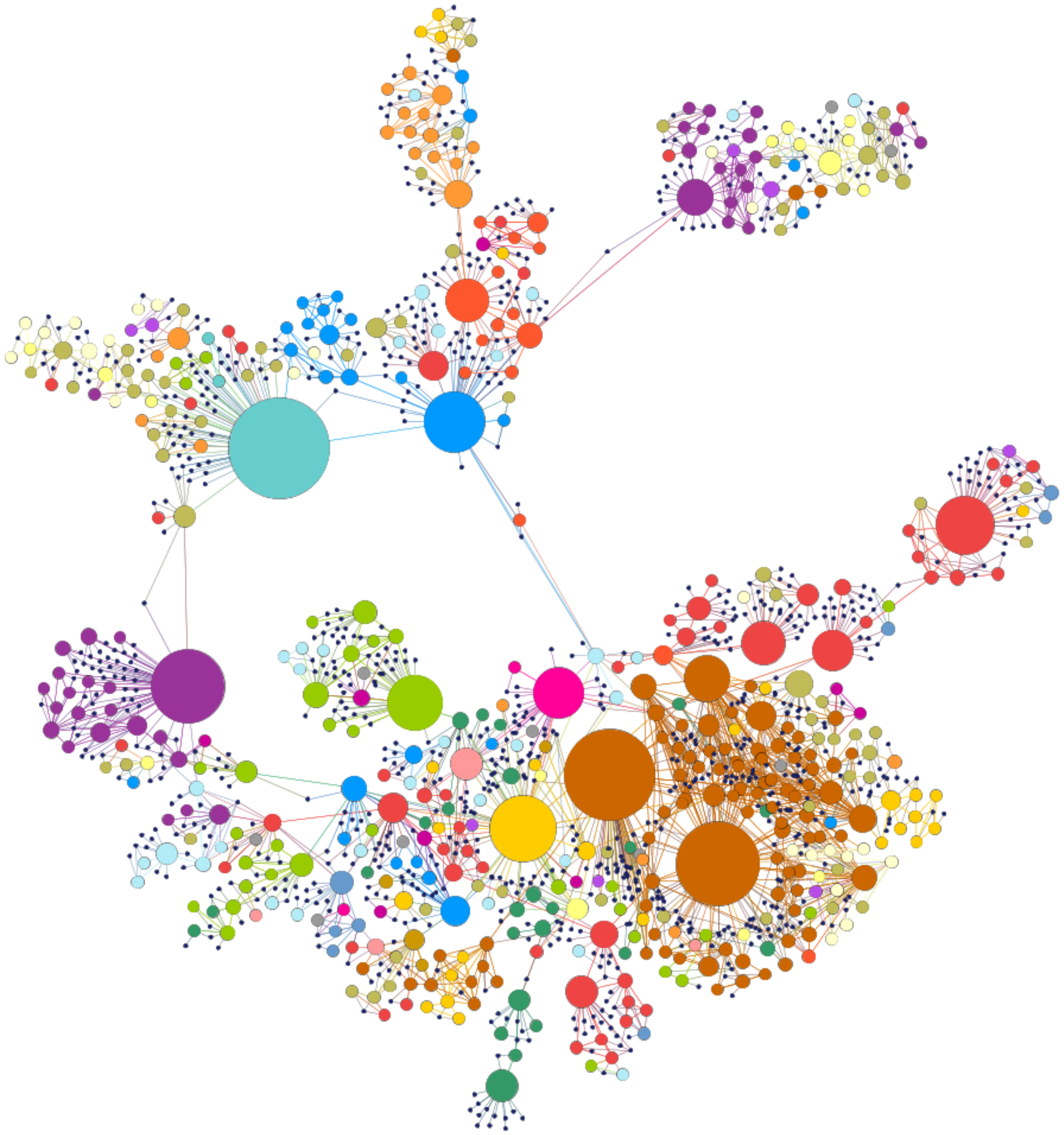
Projection of the bipartite graph of diseases and genes

- If there is a gene that is implicated in two disorders, then the two are connected.
- The width of a link is proportional to the number of genes that are implicated in both diseases.
- Three genes are implicated in both breast cancer and prostate cancer, resulting in a link of weight three between them



Human disease network

- Nodes represent diseases and two diseases are connected to each other if they share at least one gene in which mutations are associated with both diseases
- Each disease class is represented by a different color; the diseases include Bone, cancer, cardiovascular, skeletal, or metabolic diseases
- The size of a node is proportional to the number of genes participating in the corresponding disease



Goh KI, et.al. The human disease network. Proceedings of the National Academy of Sciences. 2007 May 22;104(21):8685-90.

PageRank method - a major ingredient of Google search engine

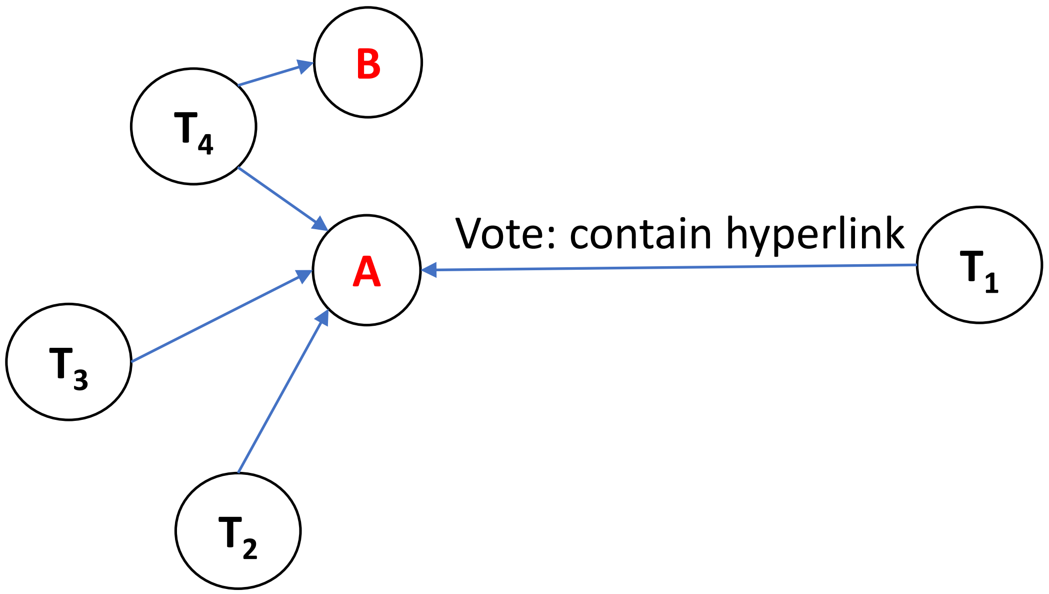
PR(A) = (1-d) + d (PR(T1)/C(T1) + ... + PR(Tn)/C(Tn))

d: a damping factor – 0.85

PR(Ti): PageRank score of Ti

C(Ti): The number of edges going out of Ti

PR(Ti)/C(Ti): the share of the vote



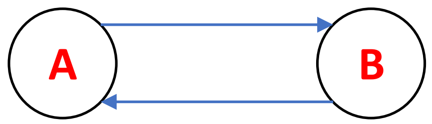
If there’s no link pointing to a page (no vote), then the default PageRank score would be 0.15

An example to illustrate the process of PageRank calculation

PR(A) = (1-d) + d (PR(T1)/C(T1) + ... + PR(Tn)/C(Tn))

Guess: PR(A) = 40; PR(B) = 40

First calculation
PR(A) = 0.15 + 0.85 * 40 = 34.15
PR(B) = 0.15 + 0.85 * 34.15 = 29.1775



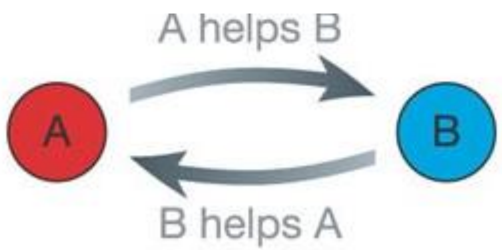
And again
PR(A) = 0.15 + 0.85 * 29.1775 = 24.950875
PR(B) = 0.15 + 0.85 * 24.950875 = 21.35824375

... repeat the calculations lots of times until the numbers stop changing much

PR(A) = 1.0

PR(B) = 1.0

Reciprocity



- In network science, reciprocity is a measure of the likelihood of vertices in a directed network to be mutually linked
- A common interest in looking at directed dyadic relationships is the extent to which ties are reciprocated
- There is an equilibrium tendency toward dyadic relationships to be either null or reciprocated (stable)
- The useful information from reciprocity is not the value itself, but whether mutual links occur more or less often than expected by chance

Reciprocity in healthcare

- The ***what*** of reciprocity refers to equivalence reciprocity, which is defined as the extent to which what is exchanged is directly comparable to what was received.
- The ***when*** of reciprocity refers to immediacy reciprocity, which is defined as the length of time between an initial action and its response
- Assigning and completing a clinical task
- Patient safety- be awareness of adverse events and report to teammates

Calculation – network level

For the Web, the reciprocity is about 57%, meaning that more than half of the links link back. For the network of who has whom in the email address book the reciprocity was found about 23%.

$$\rho \equiv \frac{\sum_{i\neq j}(a_{ij}-\bar{a})(a_{ji}-\bar{a})}{\sum_{i\neq j}(a_{ij}-\bar{a})^2}$$

$$\bar{a} \equiv \frac{\sum_{i\neq j} a_{ij}}{N(N-1)} = \frac{L}{N(N-1)}$$

measures the ratio of observed to possible directed links (N(N-1)) (link density), and self-linking loops are now excluded from L because of i not equal to j

$$r = \frac{L^{<->}}{L}$$

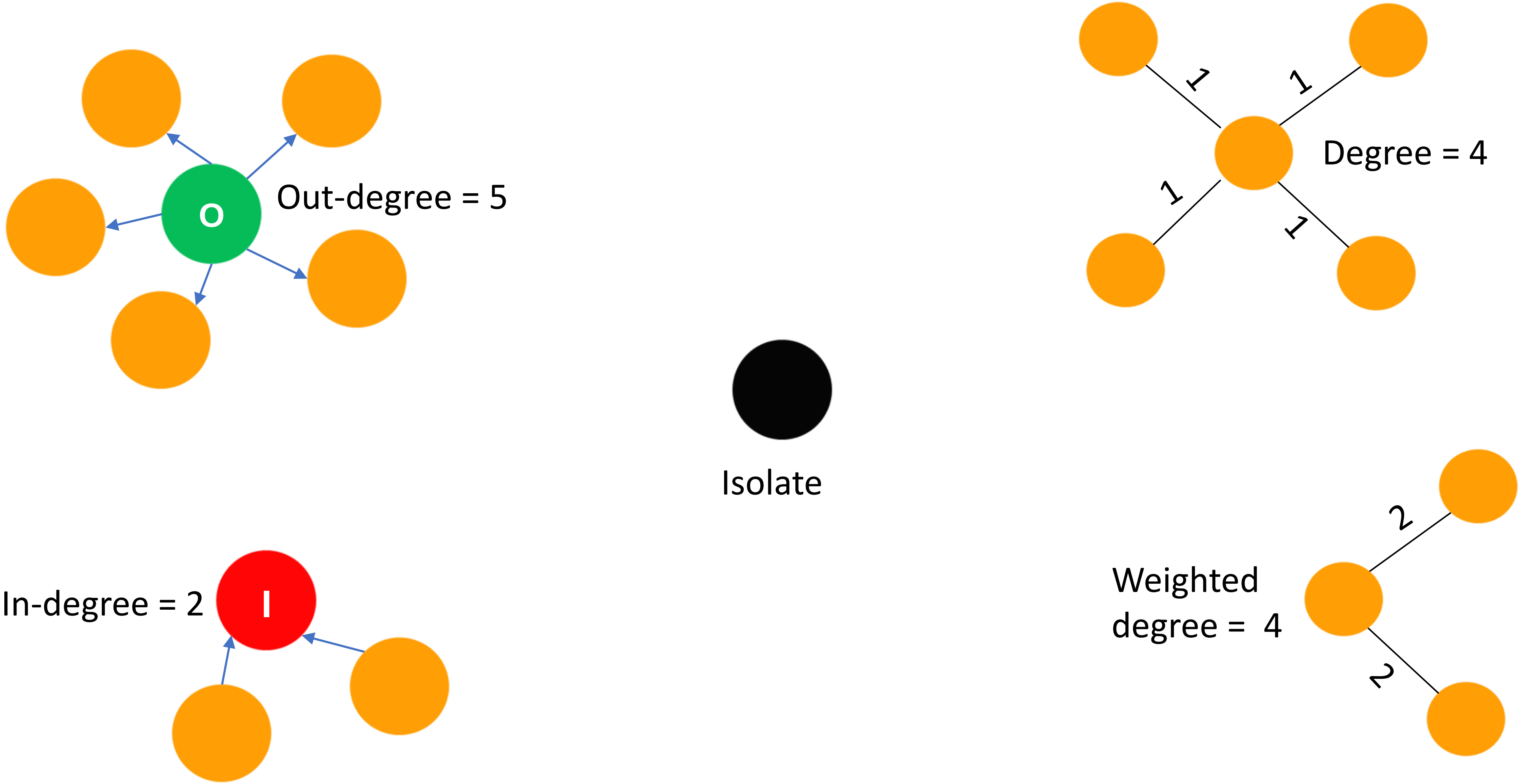
the ratio of the number of links pointing in both directions to the total number of links

$$\rho = \frac{r-\bar{a}}{1-\bar{a}}$$

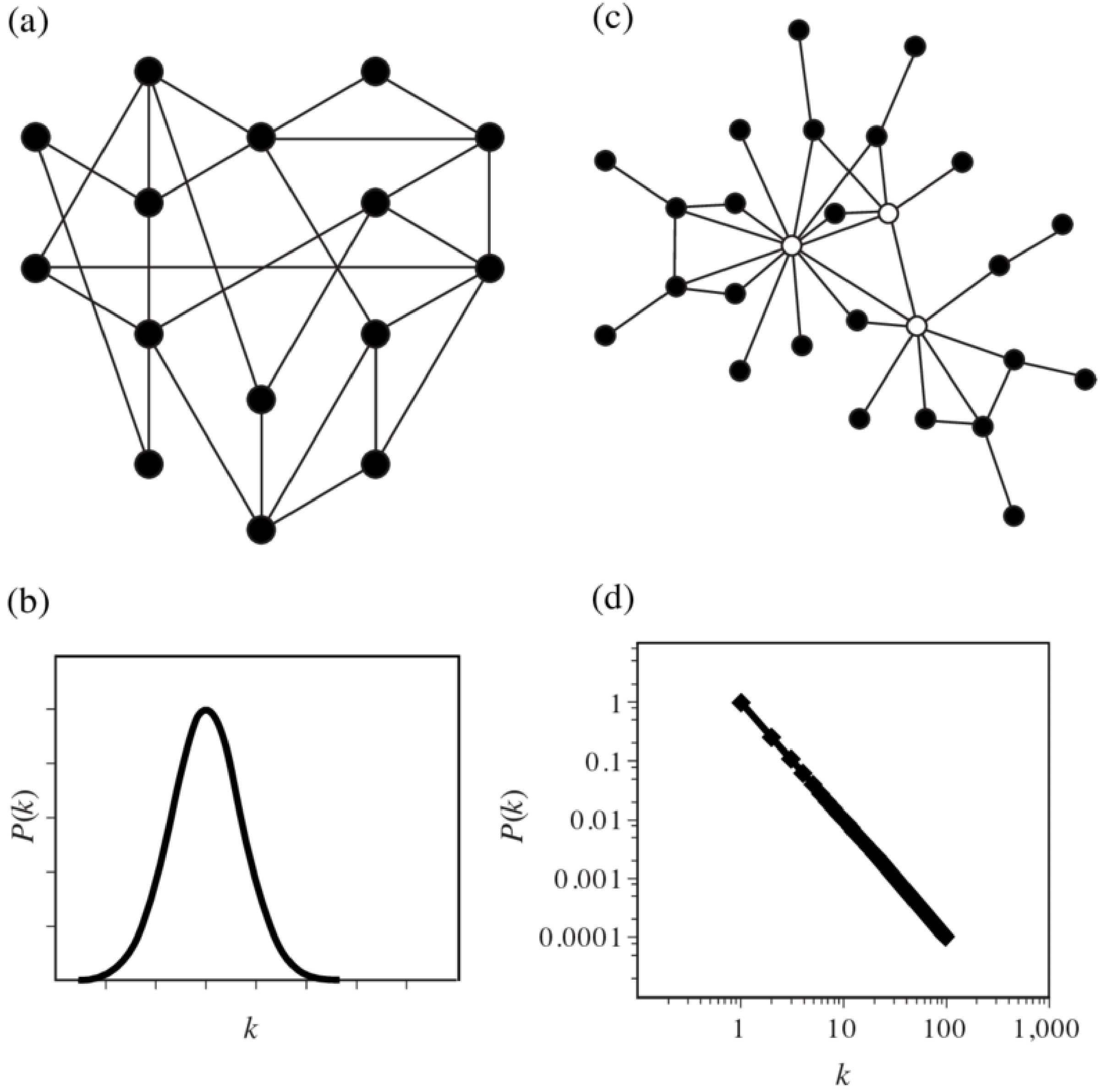
If r=0, then $\rho = \rho_{min}$

If all links occur in reciprocal pairs, r=1

Degree


O
Out-degree = 5
1
1
1
1
Degree = 4
Isolate
In-degree = 2
I
2
2
Weighted
degree = 4

Degree distribution

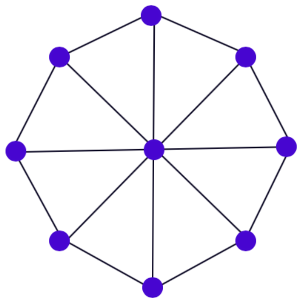


Random / Erdös-Rényi (ER) network

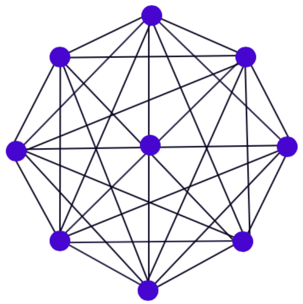
Power-law / Scale-free network

Density

- Number of ties expressed as percentage of number of expected pairs
- Actual connections/maximum possible connections



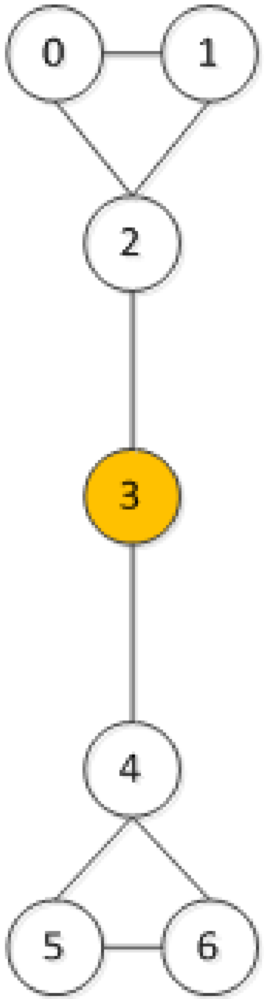
Low Density



High Density

$$D = \frac{2|E|}{|V|\,(|V|-1)}$$

Betweenness



Betweenness of node 2:
Node 3→ 0: 1
Node 3→ 1: 1
Node 4→ 0:1
Node 4→1:1
Node 5→0:1
Node 5→1:1
Node 6→0:1
Node 6→1:1
Betweenness Centrality of node 2:8

Betweenness of node 3:
Node 4→ 0: 1
Node 4→ 1: 1
Node 4→ 2:1
Node 5→0:1
Node 5→1:1
Node 5→2:1
Node 6→0:1
Node 6→1:1
Node 6→2:1
Betweenness Centrality of node 3:9

Betweenness of node 4:
Node 5→0:1
Node 5→1:1
Node 5→2:1
Node 5→3:1
Node 6→0:1
Node 6→1:1
Node 6→2:1
Node 6→3:1
Betweenness Centrality of node 4:8

Betweenness Centrality of node 0: 0
Betweenness Centrality of node 1: 0
Betweenness Centrality of node 5: 0
Betweenness Centrality of node 6: 0

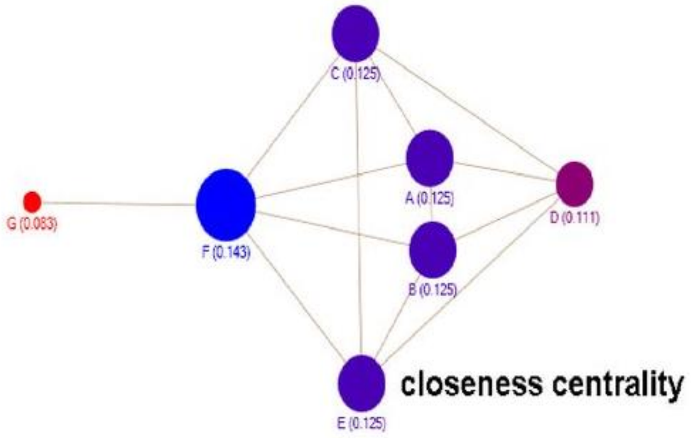
Closeness

$$C_i = 1 \Big/ \sum_{j=1}^{n} d(i,j)$$

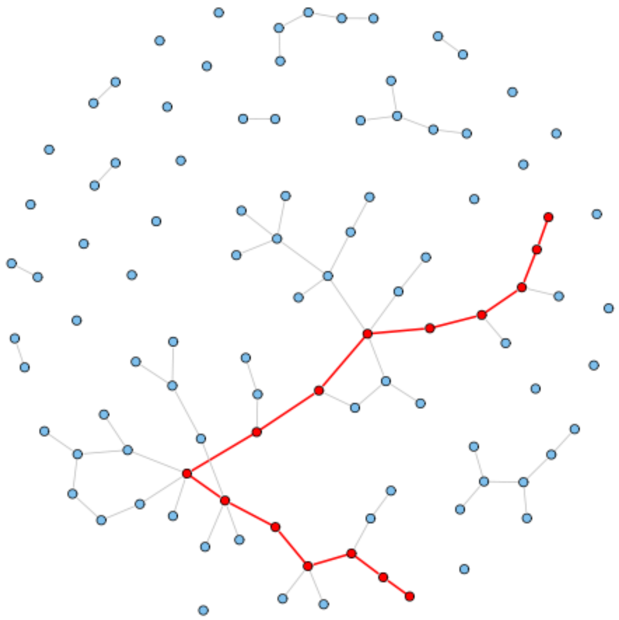
$$C_A = \frac{1}{d(AB) + d(AC) + d(AD) + d(AE) + d(AF) + d(AG)}$$

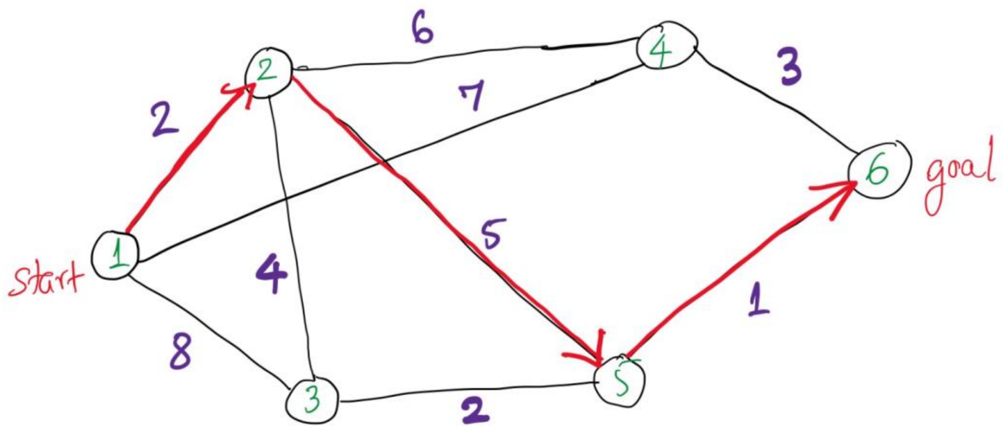
$$C_A = \frac{1}{1+1+1+2+1+2}$$

$$C_A = \frac{1}{8} = 0.125$$



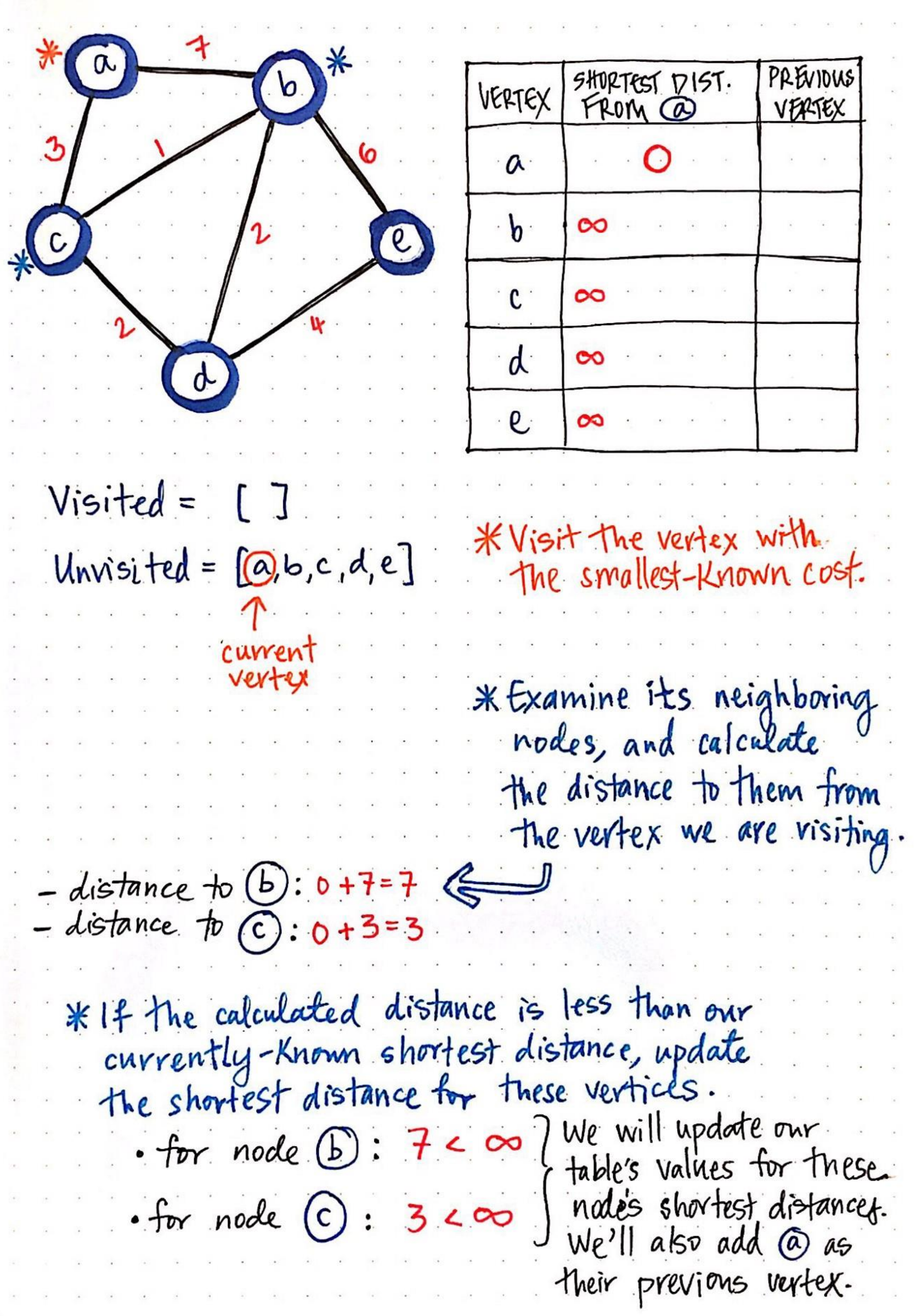
Shortest path - Dijkstra's shortest path algorithm



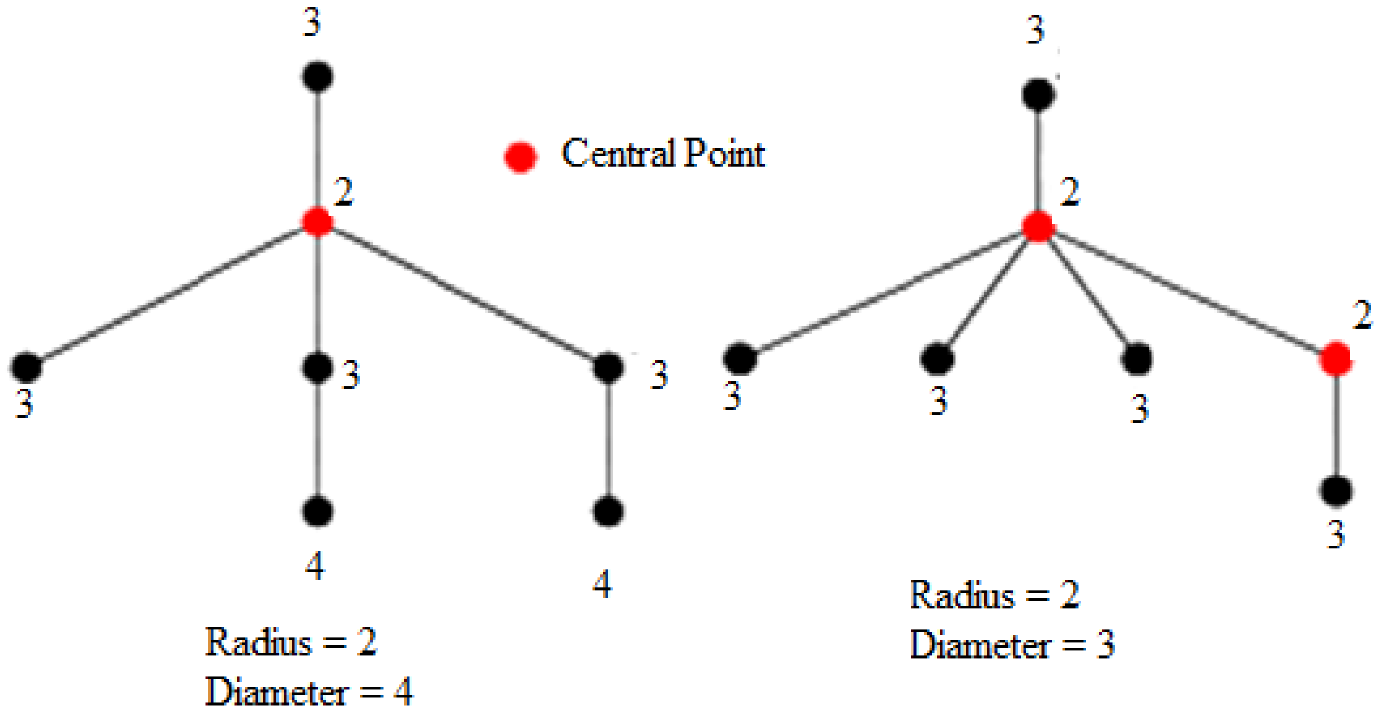


Dijkstra's shortest path algorithm rules

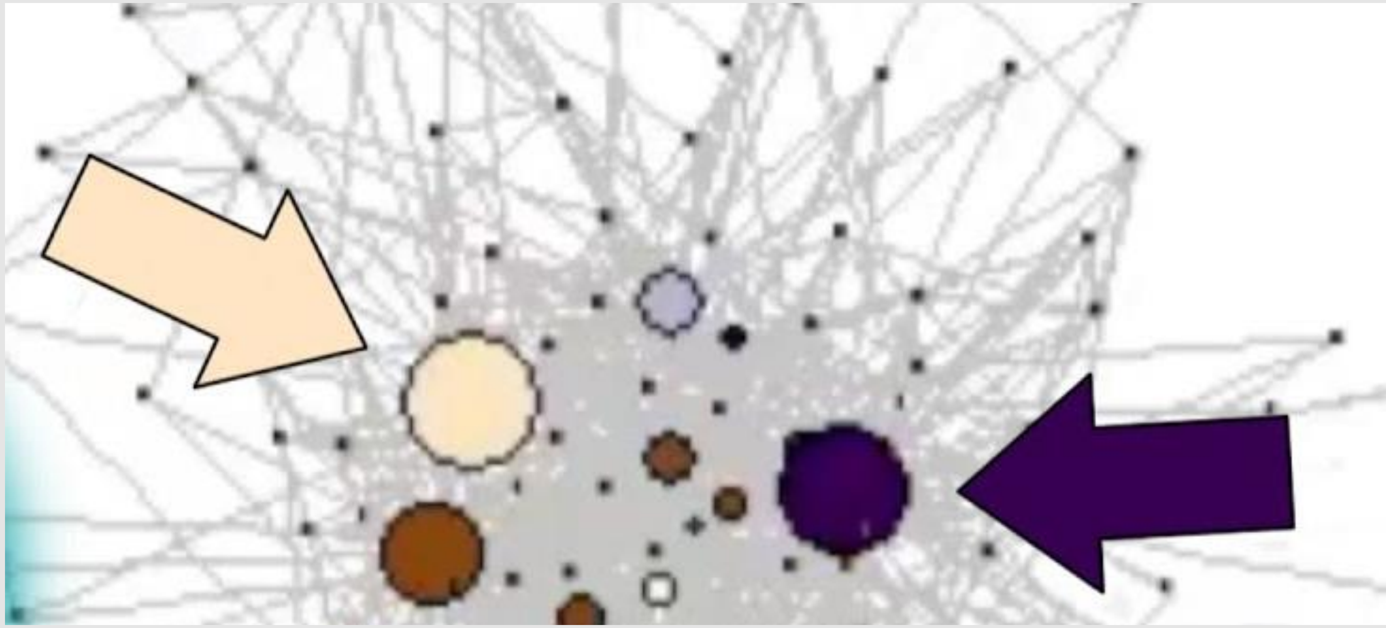
- There is no negative edges
- Set distance to source vertex as 0, and set all other distances to infinity



Diameter and radius



Eigenvector centrality



- Eigenvector centrality is extensively used in complex network theory to assess the significance of nodes in a network based on the eigenvector of the network adjacency matrix

Eigenvalue and Eigenvector

- A is a square matrix
- A scalar λ is called an Eigenvalue of A if there is a nonzero vector X such that AX = λX. Such a vector X is called an Eigenvector of A corresponding to λ.

$$\begin{pmatrix}3 & 2\\ 3 & -2\end{pmatrix}\begin{pmatrix}2\\ 1\end{pmatrix} \stackrel{?}{=} 4\begin{pmatrix}2\\ 1\end{pmatrix}$$

$$\begin{pmatrix}3\cdot 2 + 2\cdot 1\\ 3\cdot 2 + (-2)\cdot 1\end{pmatrix} \stackrel{?}{=} \begin{pmatrix}8\\ 4\end{pmatrix}$$

$$\begin{pmatrix}8\\ 4\end{pmatrix} \stackrel{\checkmark}{=} \begin{pmatrix}8\\ 4\end{pmatrix}$$

Finding eigenvalues and eigenvectors

$$A = \begin{bmatrix} 7 & 3 \\ 3 & -1 \end{bmatrix}$$

① $$\lambda I = \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix}$$

② $$A - \lambda I = \begin{bmatrix} 7 & 3 \\ 3 & -1 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix} = \begin{bmatrix} 7-\lambda & 3 \\ 3 & -1-\lambda \end{bmatrix}$$

③ $$\det \begin{bmatrix} 7-\lambda & 3 \\ 3 & -1-\lambda \end{bmatrix} = (7-\lambda)(-1-\lambda) - (3)(3) = -7 - 7\lambda + \lambda + \lambda^2 - 9 = \lambda^2 - 6\lambda - 16$$

(4) Solving for λ:
$(\lambda - 8)(\lambda + 2) = 0$
$\lambda = 8$ and $\lambda = -2$ are the Eigen values

(5) Consider $A - \lambda I$

⑤ $$\begin{bmatrix} 7-\lambda & 3 \\ 3 & -1-\lambda \end{bmatrix}$$

$\lambda = 8$:

$$\begin{bmatrix} 7-8 & 3 \\ 3 & -1-8 \end{bmatrix} = \begin{bmatrix} -1 & 3 \\ 3 & -9 \end{bmatrix} = B$$

Solve B X = 0

(A-λI)X = 0

$$\begin{pmatrix} -1 & 3 \\ 3 & -9 \end{pmatrix} \begin{bmatrix} X1 \\ X2 \end{bmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

-X1 + 3X2 = 0 ⟶ X1 = 3X2

3X1 – 9X2 = 0 ⟶ 3X1 = 9X2 ➔ X1 = 3X2

If X2 = 1;
X1 = 3

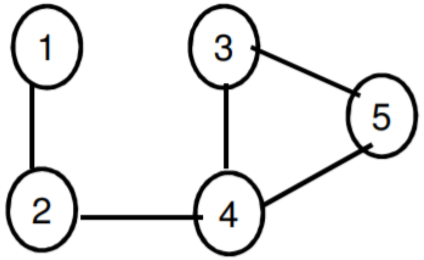
$$\begin{bmatrix} 3 \\ 1 \end{bmatrix}$$ is an eigenvector for **λ = 8**

Matrix determinant

$$\det(A) = \det\begin{bmatrix} \cancel{a} & \cancel{b} & c \\ d & e & \cancel{f} \\ g & h & \cancel{i} \end{bmatrix} = a \cdot \det\begin{bmatrix} e & f \\ h & i \end{bmatrix} - b \cdot \det\begin{bmatrix} d & f \\ g & i \end{bmatrix} + c \cdot \det\begin{bmatrix} d & e \\ g & h \end{bmatrix}$$

$$\det\begin{bmatrix} 2 & -3 & 1 \\ 2 & 0 & -1 \\ 1 & 4 & 5 \end{bmatrix} = 2 \bullet \det\begin{bmatrix} 0 & -1 \\ 4 & 5 \end{bmatrix} - (-3) \bullet \det\begin{bmatrix} 2 & -1 \\ 1 & 5 \end{bmatrix} + 1 \bullet \det\begin{bmatrix} 2 & 0 \\ 1 & 4 \end{bmatrix}$$

Eigenvector centrality



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

Let X0 = $\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$

<u>Iteration 1</u>

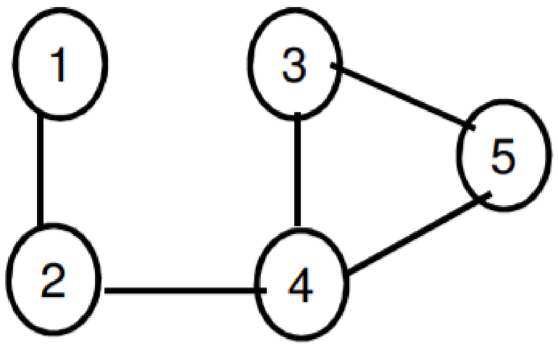
$$\begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 2 \\ 3 \\ 2 \end{bmatrix} \equiv \begin{bmatrix} 0.213 \\ 0.426 \\ 0.426 \\ 0.639 \\ 0.426 \end{bmatrix}$$

Normalized Value = 4.69

<u>Iteration 2</u>

$$\begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} 0.213 \\ 0.426 \\ 0.426 \\ 0.639 \\ 0.426 \end{bmatrix} = \begin{bmatrix} 0.426 \\ 0.852 \\ 1.065 \\ 1.278 \\ 1.065 \end{bmatrix} \equiv \begin{bmatrix} 0.195 \\ 0.389 \\ 0.486 \\ 0.584 \\ 0.486 \end{bmatrix}$$

Normalized Value = 2.19



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

Let X0 = $\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$

<u>Iteration 3</u>

$$\begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} 0.195 \\ 0.389 \\ 0.486 \\ 0.584 \\ 0.486 \end{bmatrix} = \begin{bmatrix} 0.389 \\ 0.779 \\ 1.07 \\ 1.361 \\ 1.07 \end{bmatrix} \equiv \begin{bmatrix} 0.176 \\ 0.352 \\ 0.484 \\ 0.616 \\ 0.484 \end{bmatrix}$$

Normalized Value = 2.21

<u>Iteration 4</u>

$$\begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} 0.176 \\ 0.352 \\ 0.484 \\ 0.616 \\ 0.484 \end{bmatrix} = \begin{bmatrix} 0.352 \\ 0.792 \\ 1.100 \\ 1.320 \\ 1.100 \end{bmatrix}$$

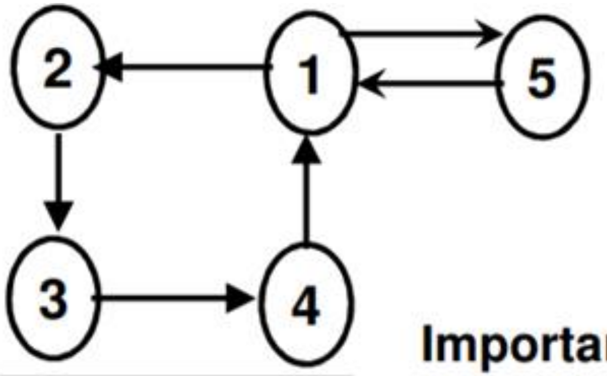
Normalized Value = 2.21 converges

<u>Eigen Vector Centrality</u>

$$\begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} \begin{bmatrix} 0.176 \\ 0.352 \\ 0.484 \\ 0.616 \\ 0.484 \end{bmatrix}$$

Eigenvector centrality in directed graph

- Importance (out-going links)– out-degree eigenvector
- Prestige (in-coming links) – in-degree eigenvector



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

Out-going links based Adj. Matrix

Importance of Nodes (Out-deg. Centrality)

Node	Score
1	0.5919
4	0.4653
5	0.4653
3	0.3658
2	0.2876

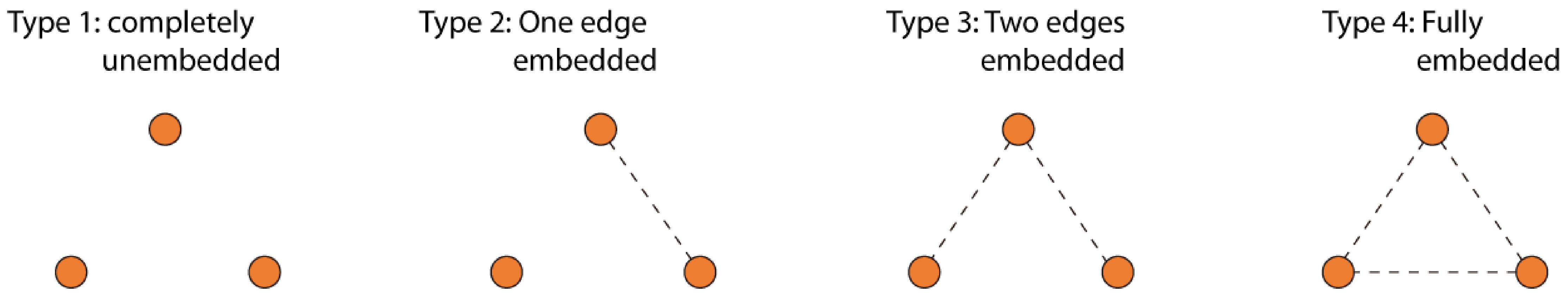
$$\begin{bmatrix} 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix}$$

In-coming links based Adj. Matrix

Prestige of Nodes (In-deg. Centrality)

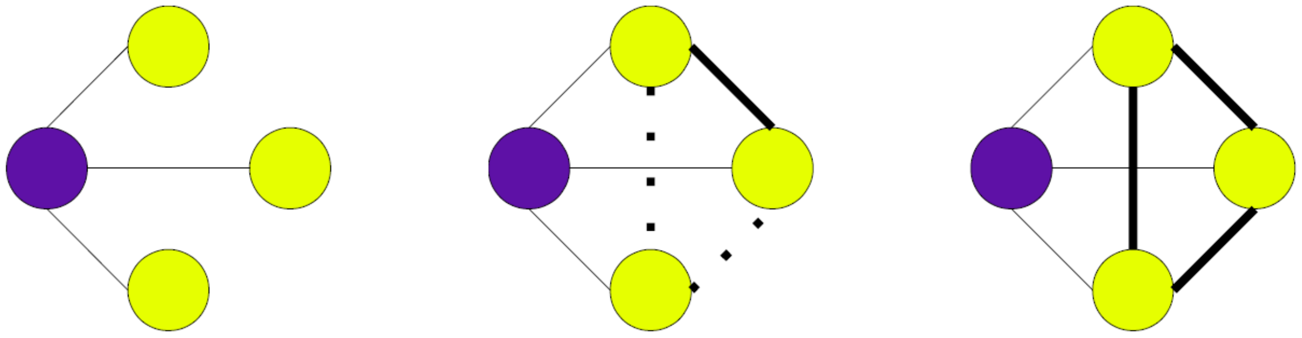
Node	Score
1	0.5919
2	0.4653
5	0.4653
3	0.3658
4	0.2876

Embeddedness



Cluster coefficient and clique

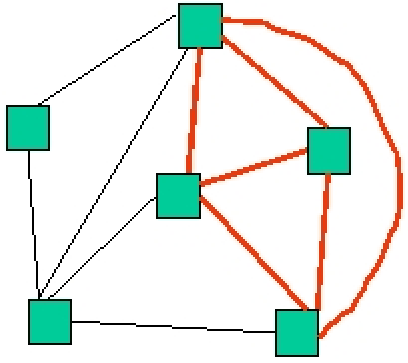
4 clique



(a) No pairs formed among neighbors: $C = 0$

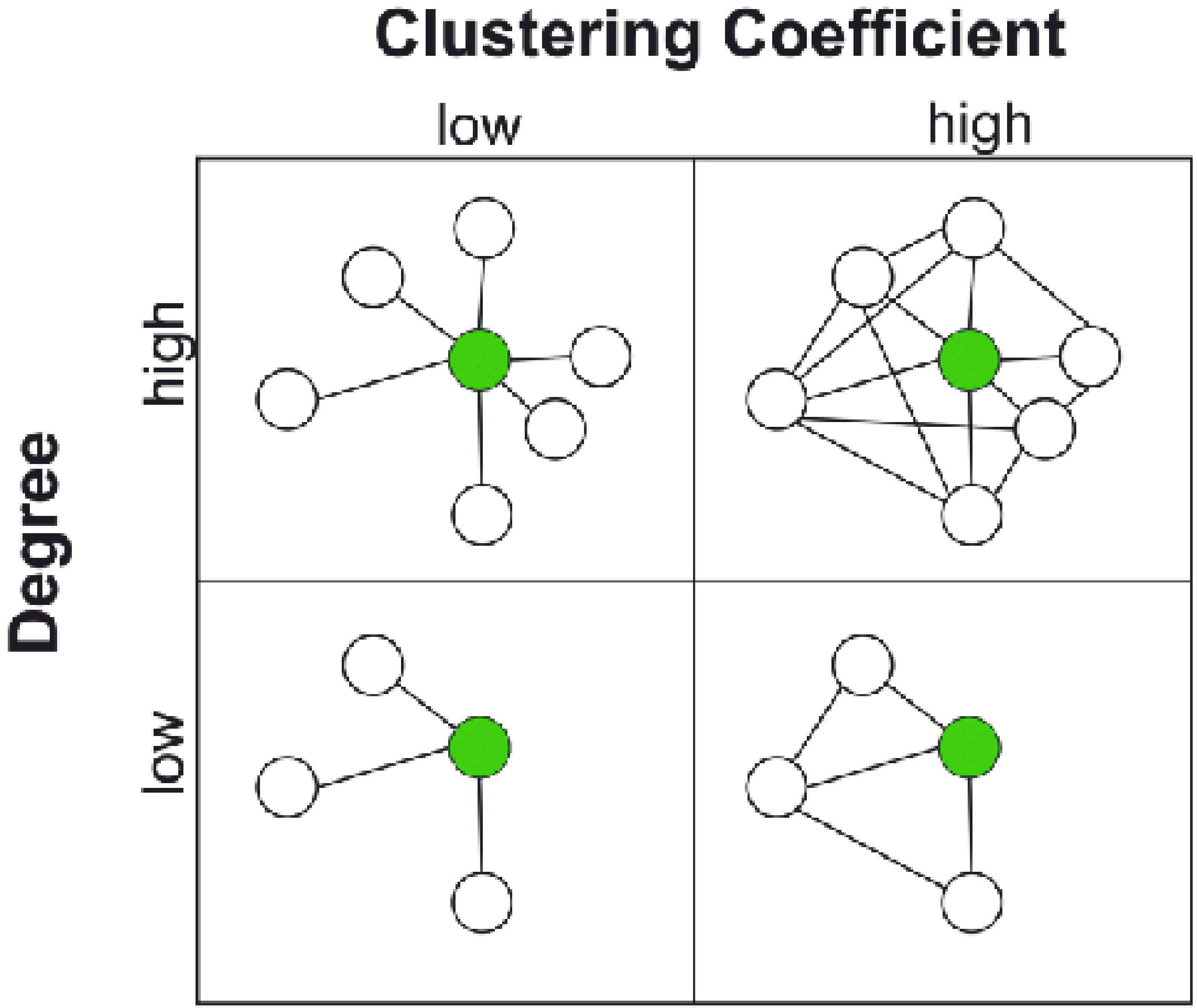
(b) One pair formed among neighbors: $C = 1/3$

(c) Three pairs formed among neighbors: $C = 3/3$



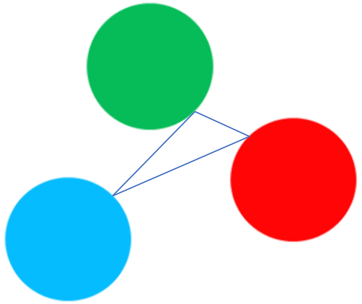
4-clique

Degree and clustering coefficient

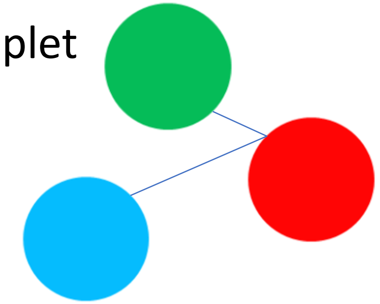


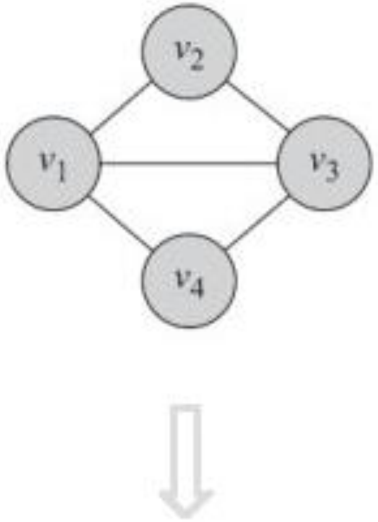
Cluster coefficient – network level

Three closed Triplets



One open Triplet

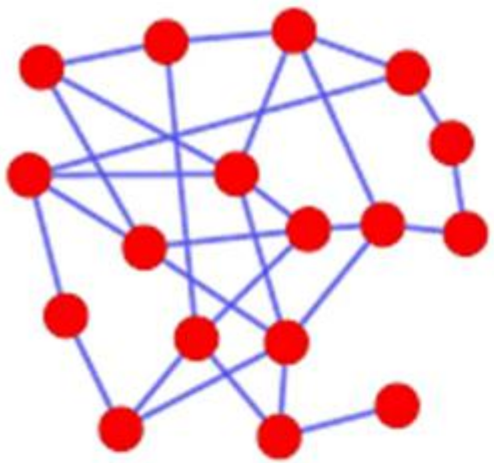




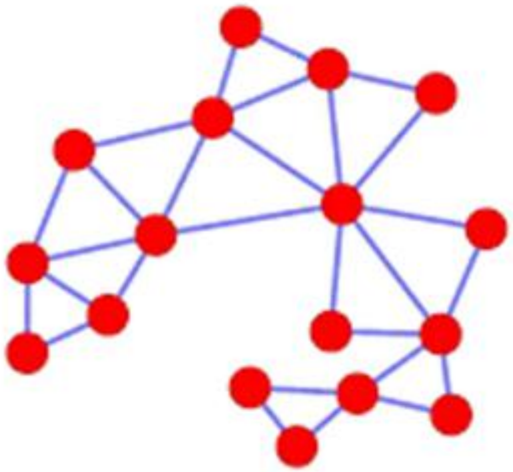
$$C = \frac{(\textit{Number of Triangles}) \times 3}{\textit{Number of Connected Triples of Nodes}} = \frac{2 \times 3}{2 \times 3 + \underbrace{2}_{v_2v_1v_4,\ v_2v_3v_4}} = 0.75.$$

closed open

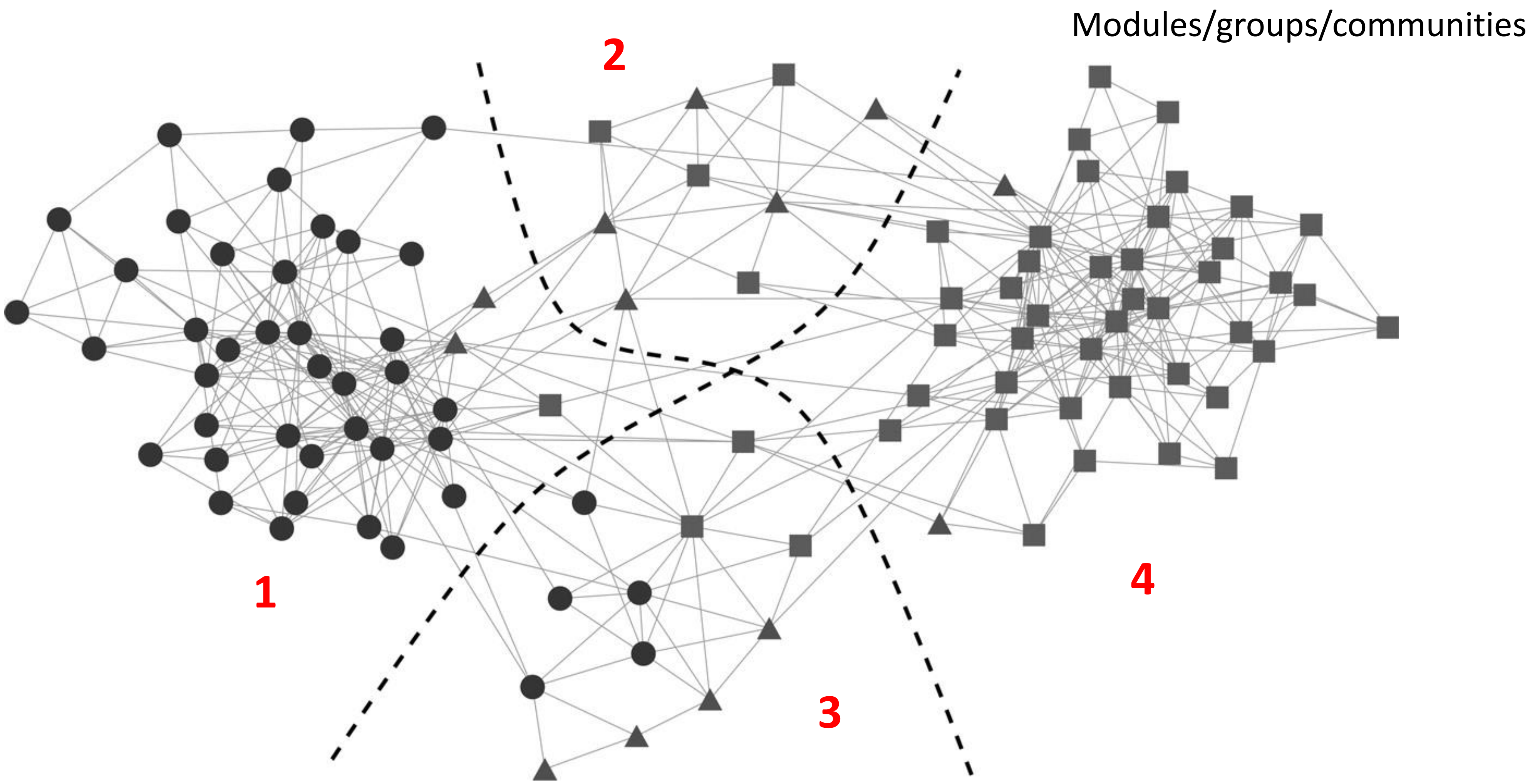
Clustering



>



Modularity – NP hard to optimize

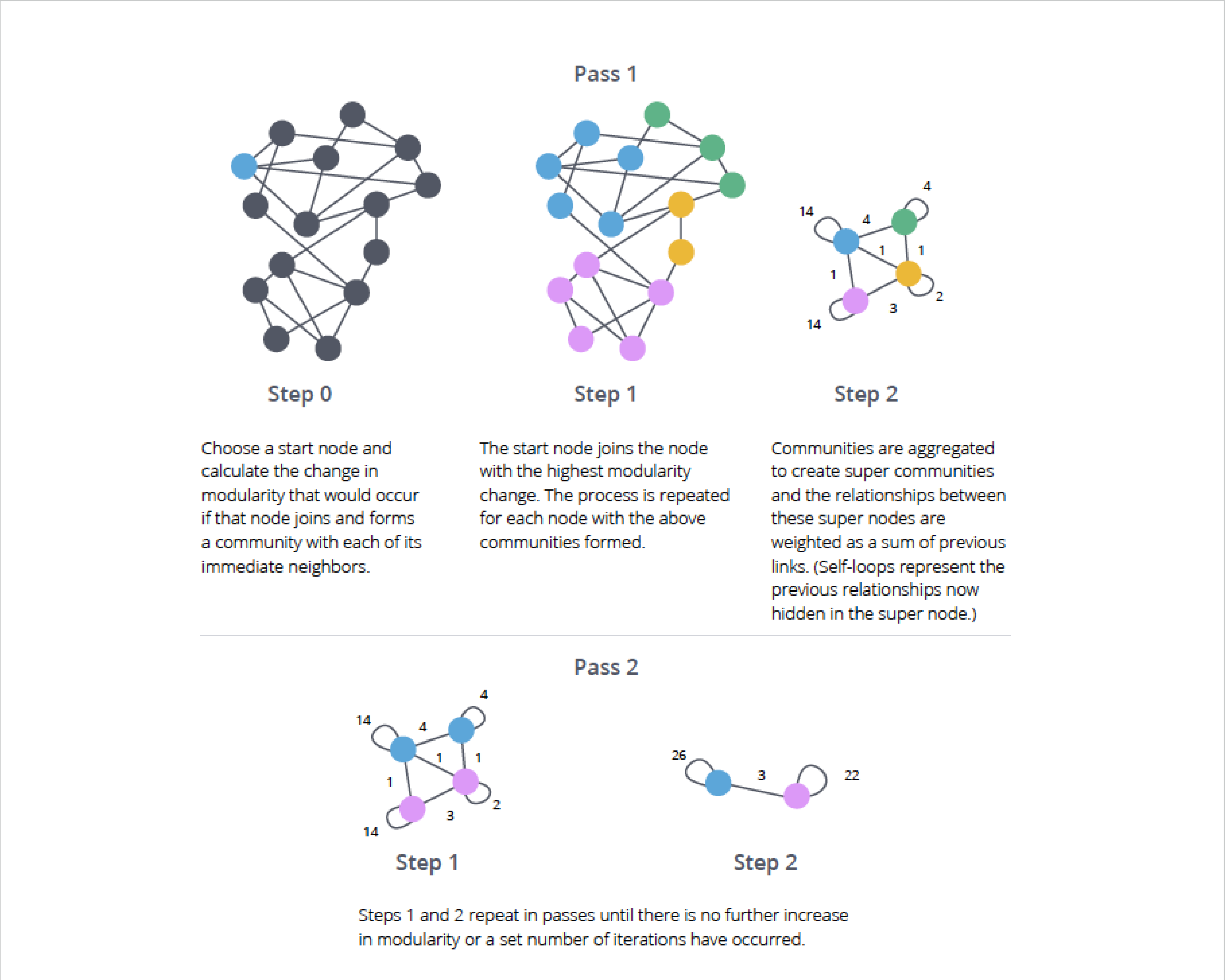


Brandes U, et.al. On modularity clustering. IEEE transactions on knowledge and data engineering. 2007 Dec 26;20(2):172-88.

Newman ME. Modularity and community structure in networks. Proceedings of the national academy of sciences. 2006 Jun 6;103(23):8577-82.

Greedy Heuristic

- Measures the relative density of edges inside communities with respect to edges outside communities
- Optimizing this value theoretically results in the best possible grouping of the nodes of a given network.
- Going through all possible iterations of the nodes into groups is impractical, heuristic algorithms are used



Louvain Modularity Algorithm

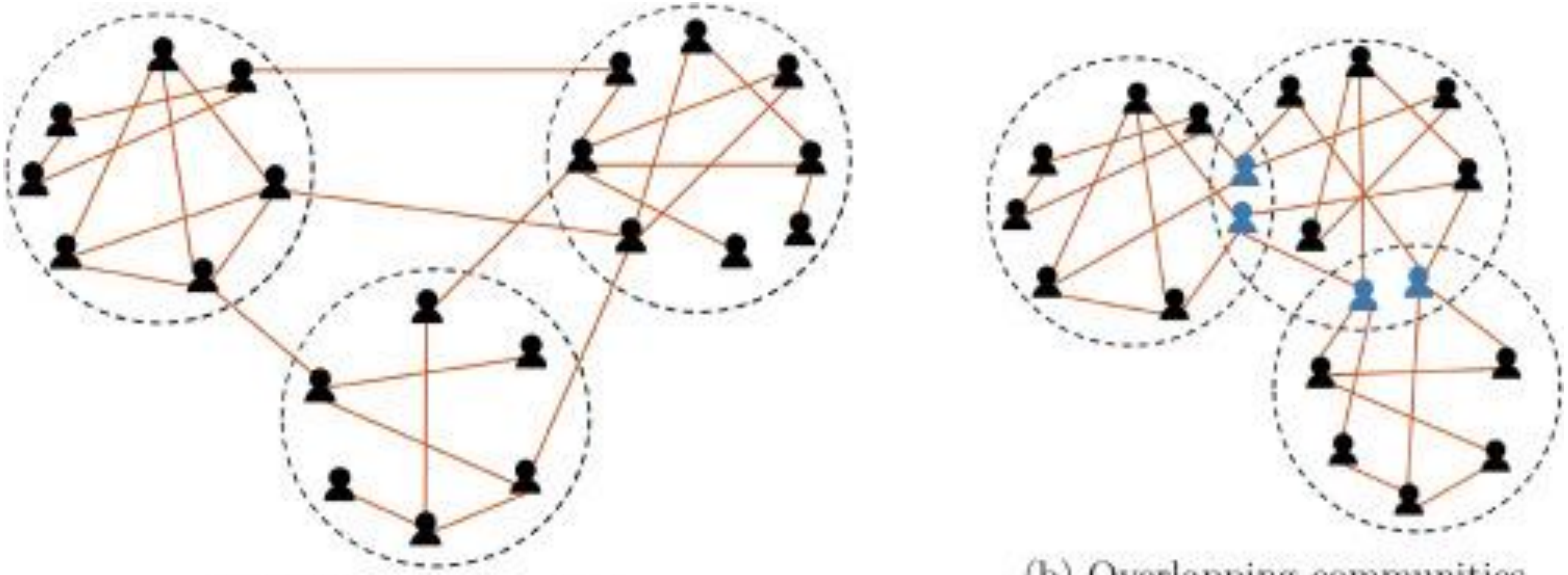
Blondel VD, et.al. Fast unfolding of communities in large networks. Journal of statistical mechanics: theory and experiment. 2008 Oct 9;2008(10):P10008.

Spectral Method for Modularity Maximization

-eigenvector indicating which groups a node belonging

https://www.youtube.com/watch?v=LnJk3LRx82U

Modularity based overlapping community detection algorithms



(a) Disjoint communities

(b) Overlapping communities