Fundamentals of network science

Vinko Zlatić

Theoretical Physics Division
Institute Ruđer Bošković

September 6, 2020

Table of Contents

Table of Contents

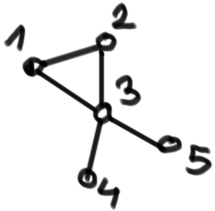
Network is a graph

- It is a collection of vertices and edges

$$G = (V, E), \ V = \{v_1, v_2, ..., v_n\}$$
$$E = \{e_{12}, e_{1,4}, e_{2,3} ...\}$$

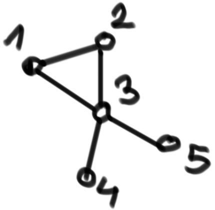
Network is a graph

- It is a collection of vertices and edges
- Graphical representation



Network is a graph

- It is a collection of vertices and edges
- Graphical representation
- Adjacency matrix



$$\widehat{\mathbf{A}} = \begin{bmatrix} 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix}$$

Degree, Clustering

We can compute some variables

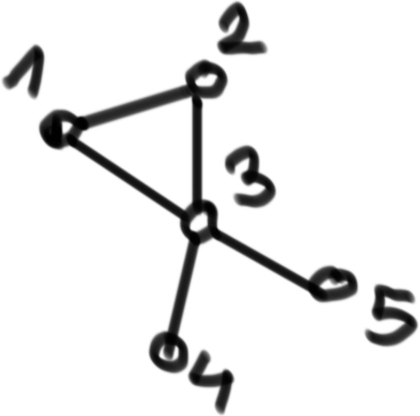
Degree, Clustering

We can compute some variables
Degree of vertex i is a number of neighbors a vertex i has -

$$k_i = \sum_j a_{ij}$$

Clustering coefficient of vertex i is a fraction of edges that exist in its neighborhood

$$C_i = \frac{2\sum_{j,k} a_{ij} a_{ik} a_{jk}}{k_i(k_i - 1)}$$



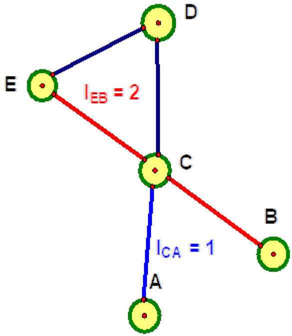
Paths

Graphs can be large

Paths

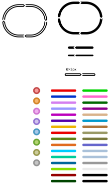
Graphs can be large

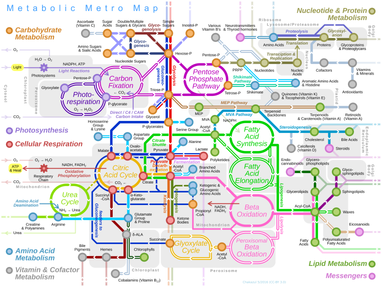
Shortest path is the shortest number of edges that connects to vertices



Graphs in reality

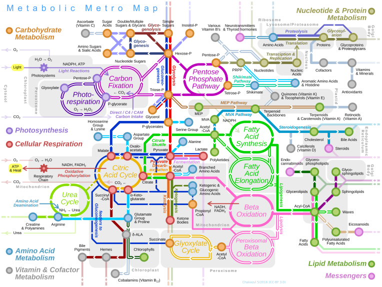
- Complex systems consist of many components interacting with each other





Graphs in reality

- Complex systems consist of many components interacting with each other
- Graph is a natural representation of such interactions



Food Webs

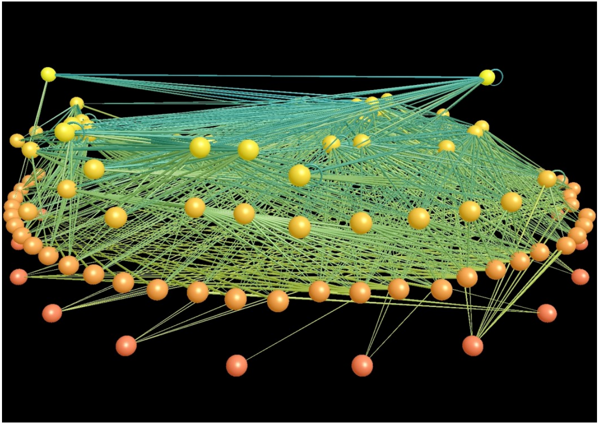


Figure: Little rock food web, courtesy of N. Martinez

Connections represent flow of energy (biomass) between different species.

Politics

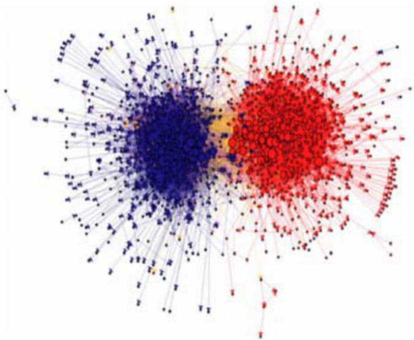


Figure: The network structure of political blogs prior to the 2004 U.S. Presidential election reveals two natural and well-separated clusters (Adamic and Glance, 2005)

Courtesy of Lada Adamic and Natalie Glance. Used with permission.

Connections represent hyperlinks between blogs

Social networks

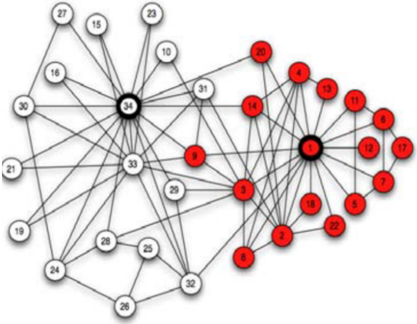
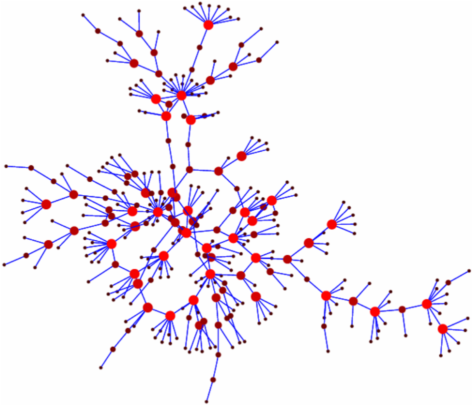


Figure: Girvan, M.; Newman, M. E. J. (2002). "Community structure in social and biological networks". PNAS. 99: 7821–7826

Connections represent friends in a karate club

Medical networks



Connections represent sexual relationships in a cluster of hiv-patients

technological networks

Connections represent flights between airports

Figure: Sun, Xiaoqian, Volker Gollnick, and Sebastian Wandelt. "Robustness analysis metrics for worldwide airport network: A comprehensive study." CJoAe 30.2 (2017): 500-512.

Economic networks

Connections represent debts between legal entities

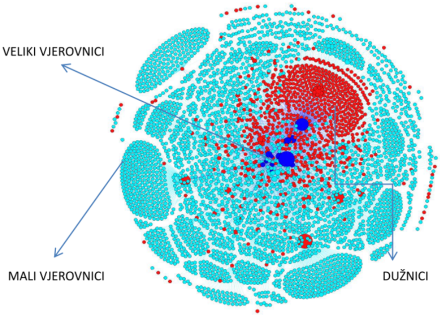


Figure: Croatian network of pre-bankruptcy settlements, V. Zlatić

Context

- All the graphs are taken from the data.

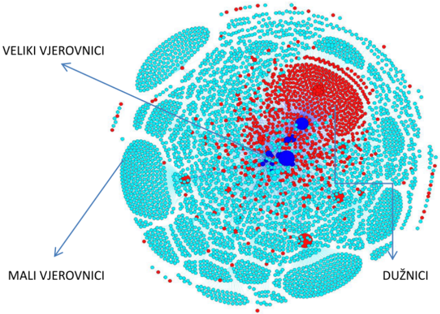
	$\ell_{mjereni}$	$\ell_{slucajni}$	$C_{mjereni}$	$C_{slucajni}$
actors	3.65	2.99	0.79	0.00027
Elec. grid	18.7	12.4	0.080	0.005
C. elegans	2.65	2.25	0.28	0.05

Context

- All the graphs are taken from the data.
- Context is important!!!

Context

- All the graphs are taken from the data.
- Context is important!!!
- Each vertex has a context (Number of Assets, liabilities, type of business). Each edge has a context (which type of vertices it connects, size of debt, maturity).

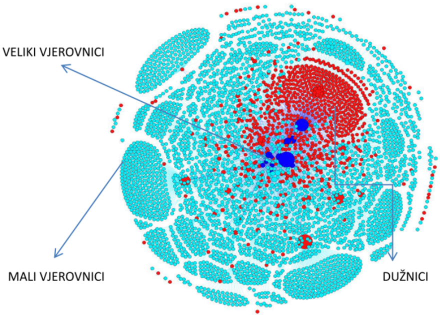


Data

- One can study all the data with connections

Data

- One can study all the data with connections
- Real connections (hyperlinks, roads, flights,)



Data

- One can study all the data with connections
- Real connections (hyperlinks, roads, flights,)
- Inferred connections (correlations, friendships...)

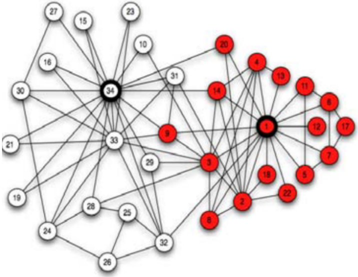
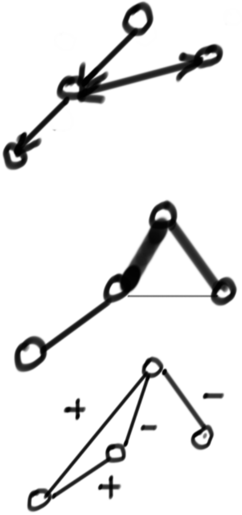


Table of Contents

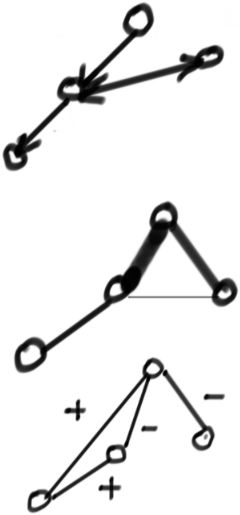
classical

- Directed networks - Vertices can be connected from one to another.



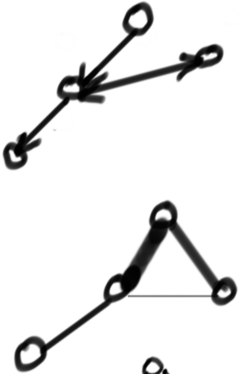
classical

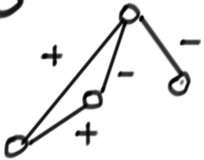
- Directed networks - Vertices can be connected from one to another.
- Weighted networks - edges can have weights signifying strength of interaction, capacity of flow, correlation ...



classical

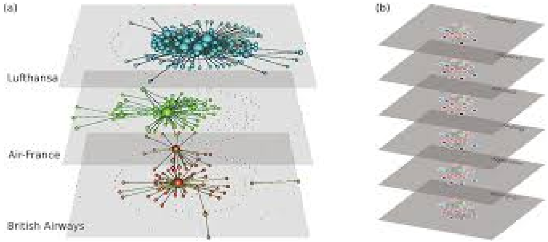
- Directed networks - Vertices can be connected from one to another.
- Weighted networks - edges can have weights signifying strength of interaction, capacity of flow, correlation ...
- signed networks - edges (or vertices) can have different roles in a network.





Multi layered networks

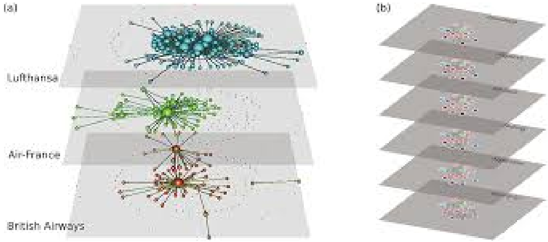
- Vertices can be connected in different ways.



M.A. Porter Notices of the AMS 65 (11)

Multi layered networks

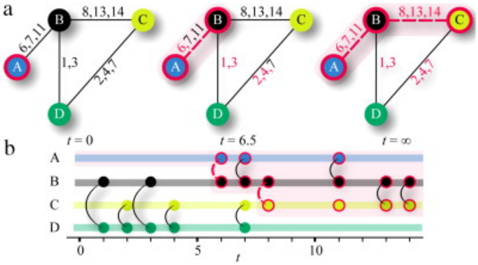
- Vertices can be connected in different ways.
- Multi layered networks!



M.A. Porter Notices of the AMS 65 (11)

Temporal networks

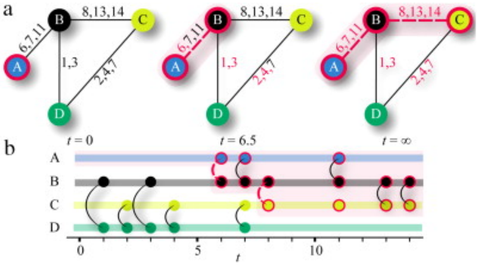
- Networks are often not static



Holme, Petter, and Jari Saramäki. "Temporal networks." Physics reports 519.3 (2012): 97-125.

Temporal networks

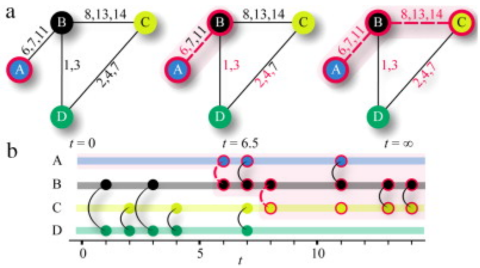
- Networks are often not static
- Temporal networks!



Holme, Petter, and Jari Saramäki. "Temporal networks." Physics reports 519.3 (2012): 97-125.

Temporal networks

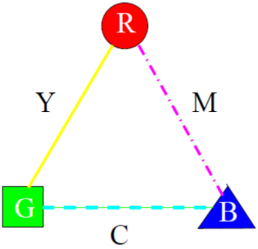
- Networks are often not static
- Temporal networks!
- Both vertices and edges can be created and broken.



Holme, Petter, and Jari Saramäki. "Temporal networks." Physics reports 519.3 (2012): 97-125.

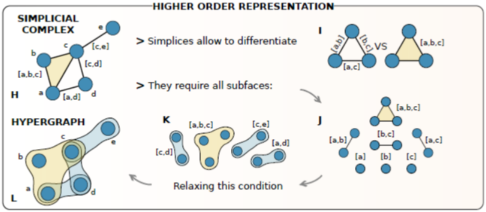
Higher order networks

- More than two vertices in interaction



Higher order networks

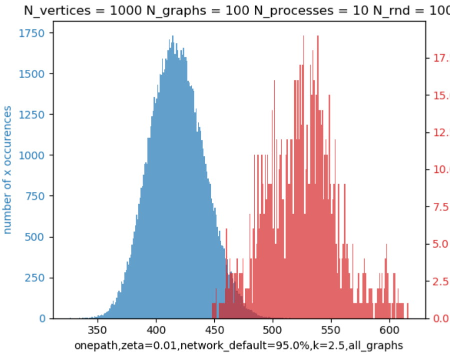
- More than two vertices in interaction
- Hypergraphs
- Simplicial complexes



Battiston, Federico, et al. "Networks beyond pairwise interactions: structure and dynamics." Physics Reports (2020).

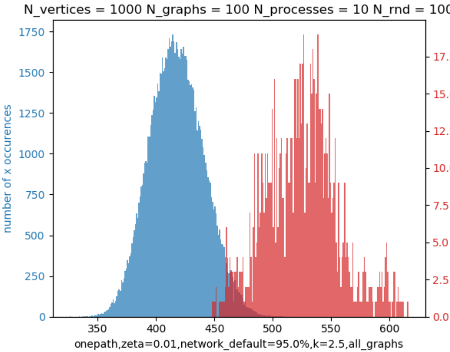
Models

- Understanding data



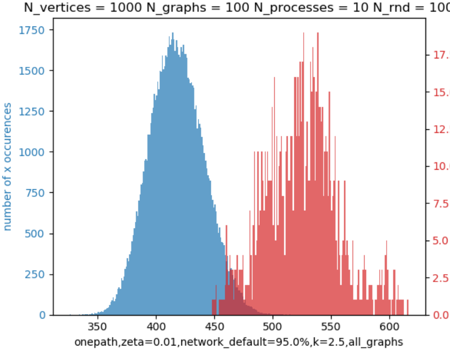
Models

- Understanding data
- What does it tell us?
  Compare it with a model.



Models

- Understanding data
- What does it tell us?
  Compare it with a model.
- Null-model stochastic graphs



$$\langle X \rangle = \sum P(G)X(G)$$

Erdős-Rényi

- Erdős-Rényi model - all the edges can happen with the same probability (maximally random model).

Erdős-Rényi

- Erdős-Rényi model - all the edges can happen with the same probability (maximally random model).
- Fixing average degree will produce an ensemble of networks with the same or similar degree.

$$< k >= \frac{\sum_i k_i}{N}$$

Erdős-Rényi

- Erdős-Rényi model - all the edges can happen with the same probability (maximally random model).
- Fixing average degree will produce an ensemble of networks with the same or similar degree.
- Same if we preserve all the edges and just randomly spread them on network.

$$E = const$$

Microcanonical

Erdős-Rényi

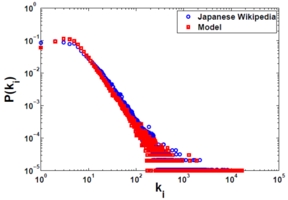
- Erdős-Rényi model - all the edges can happen with the same probability (maximally random model).
- Fixing average degree will produce an ensemble of networks with the same or similar degree.
- Same if we preserve all the edges and just randomly spread them on network.
- Similar if the number of edges is used to compute probability of creating an edge

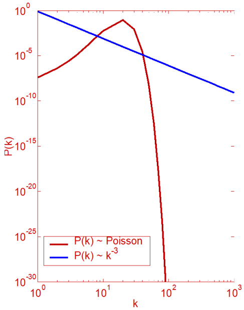
$$p = \frac{2E}{N(N-1)}$$

Canonical

Other models

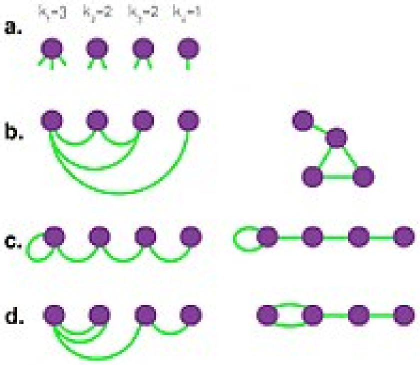
- Distribution of degrees is usually very different from Erdős-Rényi model.





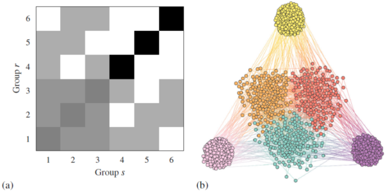
Other models

- Distribution of degrees is usually very different from Erdős-Rényi model.
- Fixing degree distribution - configuration model.



Other models

- Distribution of degrees is usually very different from Erdős-Rényi model.
- Fixing degree distribution - configuration model.
- Fixing probabilities of edge occurrence in subsets of vertices and between them - stochastic block models.



Peixoto, Tiago P. "Bayesian stochastic blockmodeling." Advances in network clustering and blockmodeling (2019): 289-332.

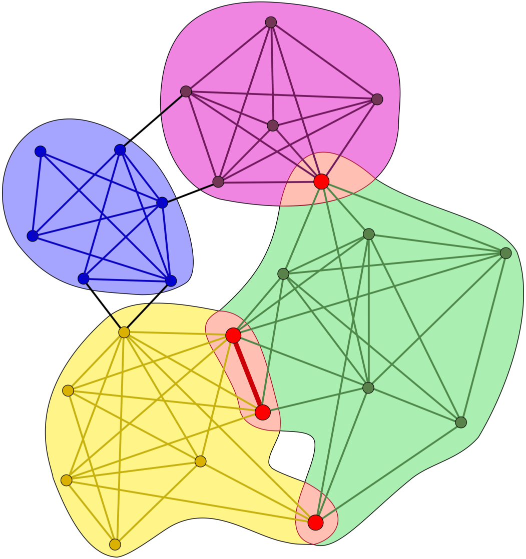
Other models

- Distribution of degrees is usually very different from Erdős-Rényi model.
- Fixing degree distribution - configuration model.
- Fixing probabilities of edge occurrence in subsets of vertices and between them - stochastic block models.
- Computing "hidden variables" and probabilities of edge occurrence as a function - hidden variable models

$$p_{ij} = \frac{x_i x_j}{1 - x_i x_j} \tag{1}$$

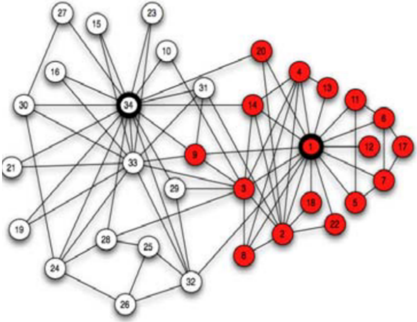
What can be interesting

- Are there sets of vertices which are better connected?



What can be interesting

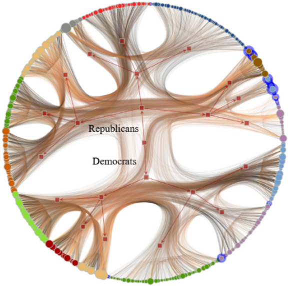
- Are there sets of vertices which are better connected?
- Community detection



$$Q = \frac{1}{4E} \sum_{ij} \left(a_{ij} - \frac{k_i k_j}{2E} \right) \delta(c_i, c_j)$$

What can be interesting

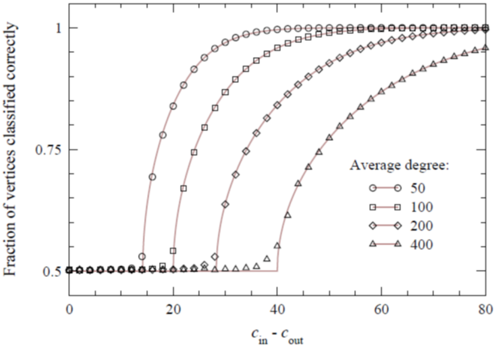
- Are there sets of vertices which are better connected?
- Community detection
- Model dependent



Peixoto, Tiago P. "Hierarchical block structures and high-resolution model selection in large networks." Physical Review X 4.1 (2014): 011047.

What can be interesting

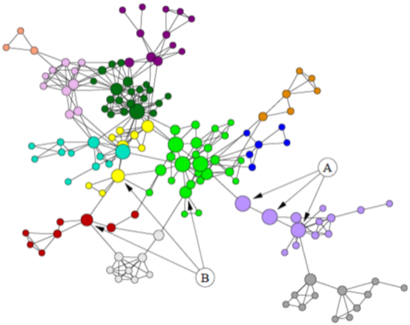
- Are there sets of vertices which are better connected?
- Community detection
- Model dependent
- Resolution limit



Nadakuditi, Raj Rao, and Mark EJ Newman. "Graph spectra and the detectability of community structure in networks." Physical review letters 108.18 (2012): 188701.

What can be interesting 2

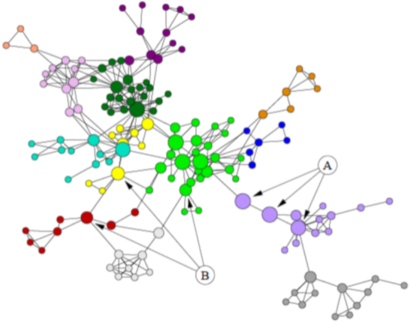
- Which vertex is the most "useful" in a network? Can we rank them?



Newman, Mark EJ. "A measure of betweenness centrality based on random walks." Social networks 27.1 (2005): 39-54.

What can be interesting 2

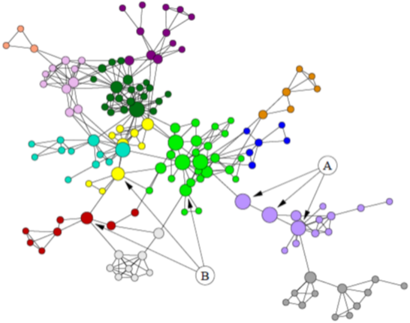
- Which vertex is the most "useful" in a network? Can we rank them?
- Centrality measures



Newman, Mark EJ. "A measure of betweenness centrality based on random walks." Social networks 27.1 (2005): 39-54.

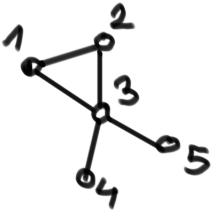
What can be interesting 2

- Which vertex is the most "useful" in a network? Can we rank them?
- Centrality measures
- Appropriate centrality measure depends on context.

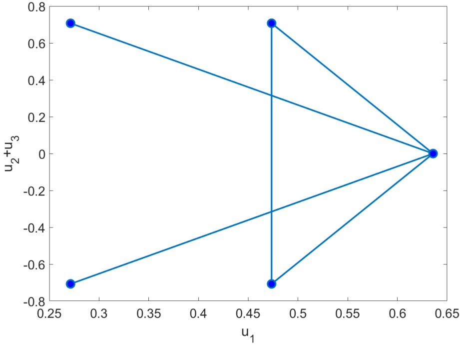


Newman, Mark EJ. "A measure of betweenness centrality based on random walks." Social networks 27.1 (2005): 39-54.

Centralities

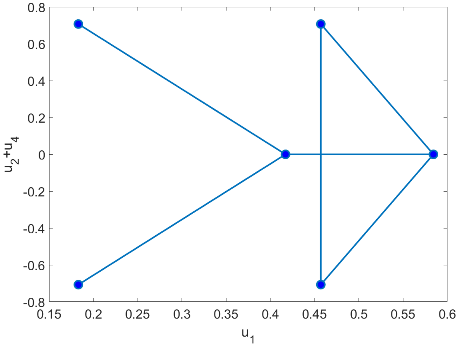


- Eigenvector centrality - ordering of vertices by degree



Centralities

- Eigenvector centrality - ordering of vertices by degree



Centralities

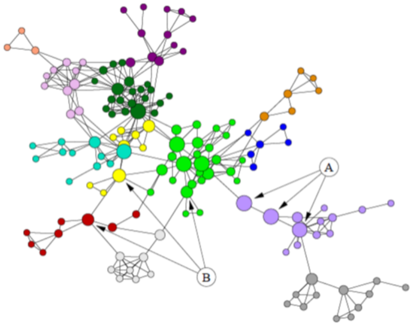
- Eigenvector centrality - ordering of vertices by degree
- Katz and Page Rank

$$x_i = \alpha \sum_j a_{ij} x_j + \beta$$

$$x_i = \alpha \sum_j \frac{a_{ij}}{k_{o,j}} x_j + \beta$$

Centralities

- Eigenvector centrality - ordering of vertices by degree
- Katz and Page Rank
- Betweenness centrality

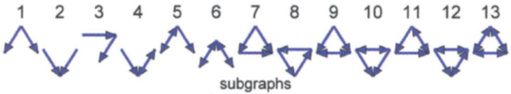


Newman, Mark EJ. "A measure of betweenness centrality based on random walks." Social networks 27.1 (2005): 39-54.

$$b_i = \frac{2\sum_{s<t} g_i(s-t)/n(s-t)}{N(N-1)}$$

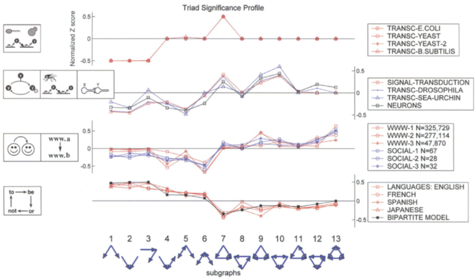
Motifs

- Small subgraphs



Motifs

- Small subgraphs
- computational elements



Milo, Ron, et al. "Superfamilies of evolved and designed networks." Science 303.5663 (2004): 1538-1542.

Motifs

- Small subgraphs
- computational elements
- Easily extended concept

Causal Motif	Probability	Causal Motif	Probability
$t_1 < t_2$	$p(\mathcal{M})\frac{1}{2}\pi(T)^2$	$t_1 < t_2 < t_3;\ t_4 < t_3$	$p(\mathcal{M})\frac{1}{8}\pi(T)^4$
$t_1 < t_2 < t_3$	$p(\mathcal{M})\frac{1}{6}\pi(T)^3$	$t_1 < t_2 < t_3;\ t_1 < t_4$	$p(\mathcal{M})\frac{1}{8}\pi(T)^4$
$t_1 < t_2, t_3$	$p(\mathcal{M})\frac{1}{3}\pi(T)^3$	$t_1 < t_2, t_3, t_4$	$p(\mathcal{M})\frac{1}{4}\pi(T)^4$
$t_1, t_2 < t_3$	$p(\mathcal{M})\frac{1}{3}\pi(T)^3$	$t_1, t_2, t_3 < t_4$	$p(\mathcal{M})\frac{1}{4}\pi(T)^4$
$t_1 < t_2 < t_3 < t_4$	$p(\mathcal{M})\frac{1}{24}\pi(T)^4$	$t_1, t_2 < t_3;\ t_2 < t_4$	$p(\mathcal{M})\frac{5}{24}\pi(T)^4$
$t_1 < t_2 < t_3, t_4$	$p(\mathcal{M})\frac{1}{12}\pi(T)^4$	$t_1 < t_2 < t_3$	$p(\mathcal{M})\frac{1}{6}\pi(T)^3$
$t_1, t_2 < t_3 < t_4$	$p(\mathcal{M})\frac{1}{12}\pi(T)^4$		

Barjašić, I et al. in preparation

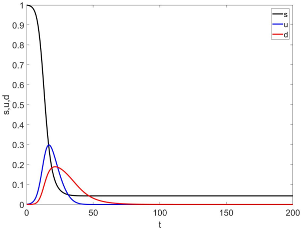
Table of Contents

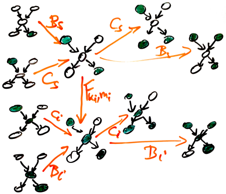
Dynamics on networks

- Random walker (central for Katz and Page Rank)

Dynamics on networks

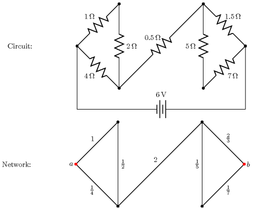
- Random walker (central for Katz and Page Rank)
- Epidemic dynamics





Dynamics on networks

- Random walker (central for Katz and Page Rank)
- Epidemic dynamics
- Minimum resistence flows



Random walker

- Where could I go?

Random walker

- Where could I go?
- Movement equally likely in all directions (no edge is more attractive than other)

Random walker

- Where could I go?
- Movement equally likely in all directions (no edge is more attractive than other)
- Can be different in weighted networks

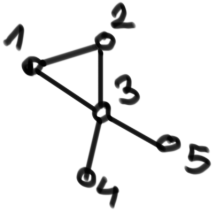
Stochastic matrix

- Markov process

$$p_i(t+1) = \sum_{path} Pr(i, t+1 | j, t; k, t-1; ...z, 0)$$
$$= \sum_{j} Pr(i, t+1 | j, t)$$

Stochastic matrix

- Markov process
- Stochastic matrix



$$\hat{\mathbf{T}} = \begin{bmatrix} 0 & \frac{1}{2} & \frac{1}{4} & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{4} & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 1 & 1 \\ 0 & 0 & \frac{1}{4} & 0 & 0 \\ 0 & 0 & \frac{1}{4} & 0 & 0 \end{bmatrix}$$

Stochastic matrix

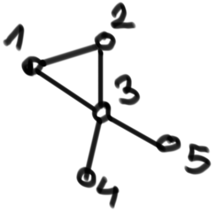
- Markov process
- Stochastic matrix
- Transition rule

$$\mathbf{p}(t+1) = \hat{\mathbf{T}}\mathbf{p}(t)$$

$$p_i(t+1) = \sum_j T_{ij} p_j(t)$$

Stochastic matrix

- Markov process
- Stochastic matrix
- Transition rule
- Not symmetric



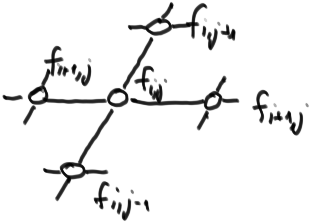
$$\hat{\mathbf{T}} = \begin{bmatrix} 0 & \frac{1}{2} & \frac{1}{4} & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{4} & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 1 & 1 \\ 0 & 0 & \frac{1}{4} & 0 & 0 \\ 0 & 0 & \frac{1}{4} & 0 & 0 \end{bmatrix}$$

Laplacian and graph laplacian

- Laplacian

$$\bigtriangledown^2 f = \sum_{i=1}^{n} \frac{\partial^2 f}{\partial x_i^2}$$

Laplacian and graph laplacian

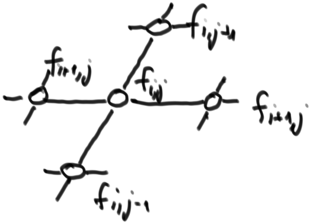


- Laplacian
- Discretized laplacian

$$\nabla^2 f \approx f_{i+1,j} + f_{i-1,j} + f_{i,j+1} + f_{i,j-1} - 4 f_{i,j}$$

Laplacian and graph laplacian

- Laplacian
- Discretized laplacian
- Direct connection



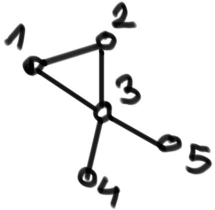
$$\hat{\mathsf{L}} = \hat{\mathsf{K}} - \hat{\mathsf{A}}$$

Graph Laplacian

- Another type of process is diffusion

$$\dot{\phi} = \hat{\mathsf{L}}\phi$$

Graph Laplacian



- Another type of process is diffusion
- Symmetric

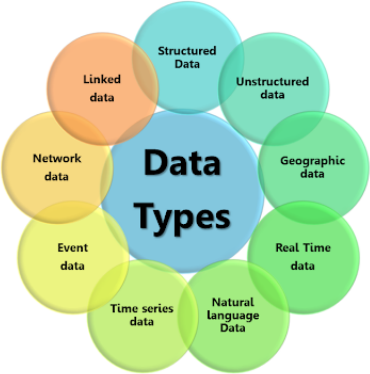
$$\hat{L} = \begin{bmatrix} 2 & -1 & -1 & 0 & 0 \\ -1 & 2 & -1 & 0 & 0 \\ -1 & -1 & 4 & -1 & -1 \\ 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 0 & 1 \end{bmatrix}$$

Graph Laplacian

- Another type of process is diffusion
- Symmetric
- Eigenvalues, eigenvectors

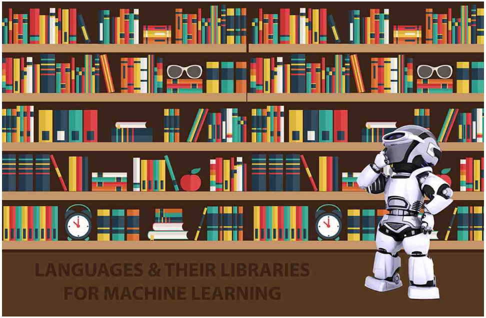
Network framework

- Capable of handling more and more data types



Network framework

- Capable of handling more and more data types
- Rich ecosystem of libraries, visualization tools etc



Network framework

Holloway, Todd, Božicevic, Miran and Börner, Katy. (2007) Analyzing and Visualizing the Semantic Coverage of Wikipedia and Its Authors. Complexity, Special issue on Understanding Complex Systems. Vol. 12(3), pp. 30-40. Also available as cs.IR/0512085.

- Capable of handling more and more data types
- Rich ecosystem of libraries, visualization tools etc
- Graphical representation of relations based on rigorous analysis

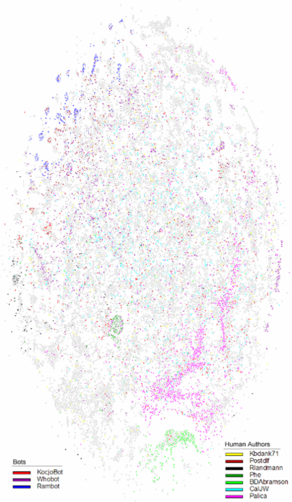


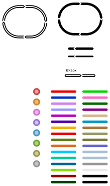
Figure 6: English Wikipedia category network colored by top ten most active authors.

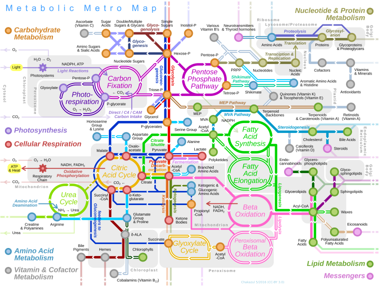
Challenges

- Reasons why certain networks look like they look

Challenges

- Reasons why certain networks look like they look
- More detailed simulations of complex systems





Challenges

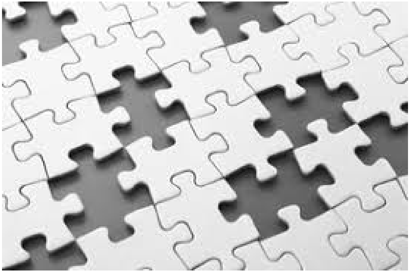
- Reasons why certain networks look like they look
- More detailed simulations of complex systems
- Data!!!

Data science challenges

- Handling large scale of data

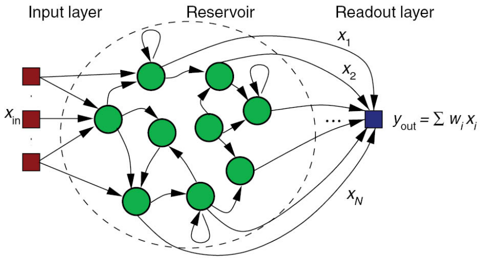
Data science challenges

- Handling large scale of data
- Data inference (handling missing data)



Data science challenges

- Handling large scale of data
- Data inference (handling missing data)
- AI



Thank you for your attention!