Introduction to Network Science

Structure of Graphs

Natalia Semenova

Higher School of Economics

nasemenova_2@edu.hse.ru

Structural Analysis and Visualization of Networks
16.01.2024

Class Details

- Instructor: Natalia Semenova
- Tutor: Vasilii Latonov, Pavel Stepachev
- Course duration: Module 3
- Website: link to page
- Contacts: course chat
- Programming: Python, iPython notebooks (Anaconda)
- Python libraries: NetworKit, NetworkX, Pytorch Geometric
- Visualization: yEd, Gephy

Prerequisites

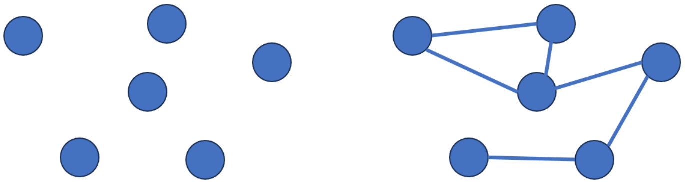
- Discrete Mathematics
- Linear Algebra
- Algorithms and Data Structures
- Probability Theory
- Differential Equations
- Programming in Python

Approximate Course Outline

Date	Topic
16.01	Introduction to Network Science; Structure of Graphs
23.01	Properties of Networks; Random Graphs Models
30.01	Subgraphs, Motifs and Graphlets
06.02	Community Structure in Networks
13.02	Spectral Clustering
20.02	Link Analysis; Page Rank
...	TBA

Why Networks?

Graphs are a general language for describing and analyzing entities with relations/interactions



Many types of data are graphs

Image credit: Medium

Social Networks

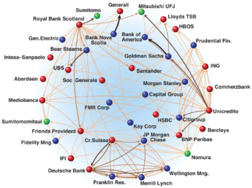


Image credit: Science

Economic Networks

Image credit: Lumen Learning

Communication Networks

Citation Networks

Image credit: Missoula Current News

Internet

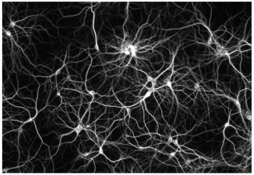
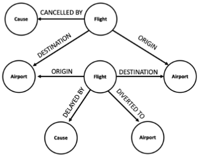


Image credit: The Conversation

Networks of Neurons

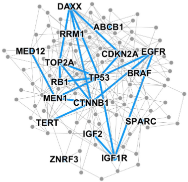
Many types of data are graphs



Event Graphs

Image credit: SalientNetworks

Computer Networks



Disease Pathways

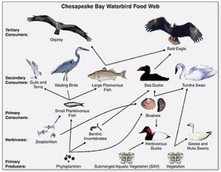


Image credit: Wikipedia

Food Webs

Image credit: Pinterest

Particle Networks

Image credit: visitlondon.com

Underground Networks

Many types of data are graphs

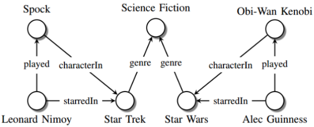


Image credit: Maximilian Nickel et al

Knowledge Graphs

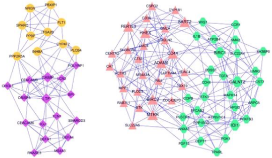


Image credit: ese.wustl.edu

Regulatory Networks

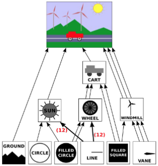


Image credit: math.hws.edu

Scene Graphs

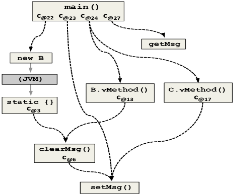


Image credit: ResearchGate

Code Graphs

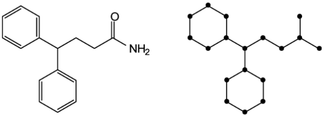


Image credit: MDPI

Molecules

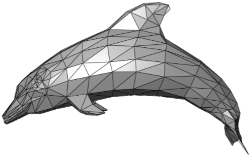


Image credit: Wikipedia

3D Shapes

Structure of Networks?

A network is a collection of objects where some pairs of objects are connected by links. What is the structure of the network?

Networks or Graphs?

Network often refers to real systems

Web, Social network, Metabolic network Terms: Network, Node, Link

Graph is a mathematical representation of a network

Web graph, Social graph, Knowledge Graph Terms: Graph, Vertex, Edge

We will try to make this distinction whenever it is appropriate, but in most cases we will use the two terms as synonyms

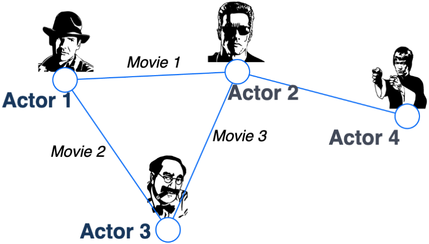
How do you define a network?

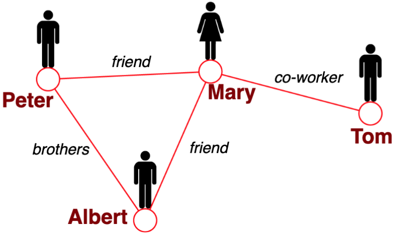
- How to build a graph:
  - What are nodes?
  - What are edges?

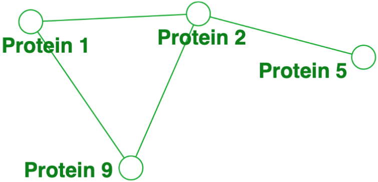
- Choice of the proper network representation of a given domain/problem determines our ability to use networks successfully:

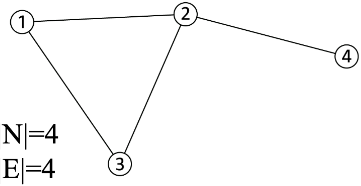
  - In some cases there is a unique, unambiguous representation
  - In other cases, the representation is by no means unique
  - The way you assign links will determine the nature of the question you can study

Networks: General Terms





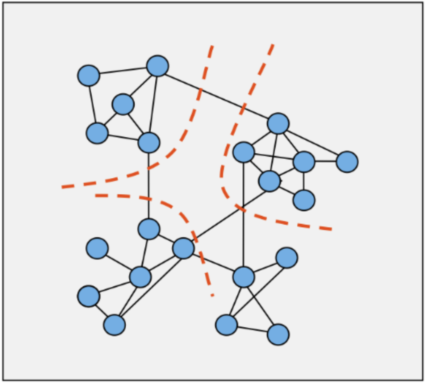




$|N|=4$
$|E|=4$

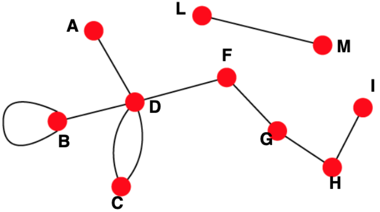
Terminology

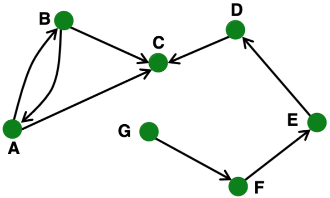
- network = graph
- nodes = vertices, actors
- links = edges, relations
- clusters = communities



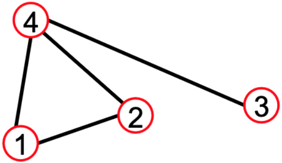
- **Network** is represented by a graph $G(V, E)$, comprising a set of vertices V and a set of edges E, connecting those vertices.
- **Graph** can be represented by an adjacency matrix A, where A_{ij} - availability of an edge between nodes i and j
- In an **unweighted** graph A_{ij} is binary $\{0, 1\}$, in a weighted graph an edge can carry a weight, A - non-binary.
- **Undirected** graph is a graph where edges have no orientation, edges are defined by unordered pairs of vertices, $A_{ij} = A_{ji}$
- **Directed** graph is a graph where edges have a direction associated with them, edges are defined by ordered pairs of vertices, $A_{ij} \neq A_{ji}$

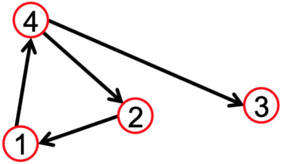
Undirected vs Directed Graphs





Adjacency Matrix



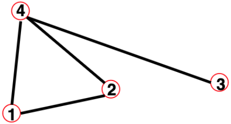


$A_{ij} = 1$ if there is a link from node i to node j

$A_{ij} = 0$ otherwise

$$A = \begin{pmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix} \qquad A = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \end{pmatrix}$$

Adjacency Matrix



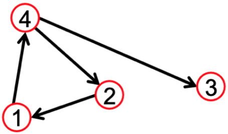
$$A_{ij} = \begin{pmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix}$$

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

$$k_i = \sum_{j=1}^{N} A_{ij}$$

$$k_j = \sum_{i=1}^{N} A_{ij}$$

$$L = \frac{1}{2}\sum_{i=1}^{N} k_i = \frac{1}{2}\sum_{ij}^{N} A_{ij}$$



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

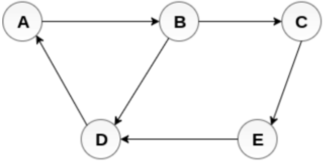
$$A_{ij} \neq A_{ji}$$
$$A_{ii} = 0$$

$$k_i^{out} = \sum_{j=1}^{N} A_{ij}$$

$$k_j^{in} = \sum_{i=1}^{N} A_{ij}$$

$$L = \sum_{i=1}^{N} k_i^{in} = \sum_{j=1}^{N} k_j^{out} = \sum_{i,j}^{N} A_{ij}$$

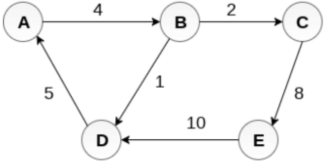
Unweighted vs Weighted Graphs



Directed Graph

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

Adjacency Matrix



Weighted Directed Graph

	A	B	C	D	E
A	0	4	0	0	0
B	0	0	2	1	0
C	0	0	0	0	8
D	5	0	0	0	0
E	0	0	0	10	0

Adjacency Matrix

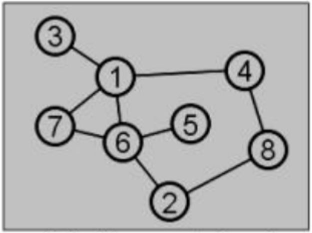
Graph Theory

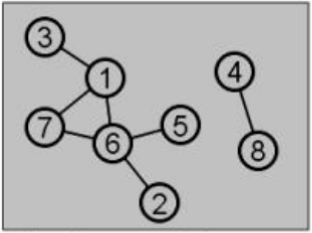
- A **path** between nodes *i* and *j* is a sequence of edges connecting vertices, starting at *i* and ending at *j*, where every vertex in the sequence is distinct
- **Distance** between two vertices in a graph is the number of edges in a shortest path (graph geodesic) connecting them.
- The **diameter** of a network is the largest shortest paths (distance between any two nodes) in the network
- **Average path length** is bounded from above by the diameter; in some cases, it can be much shorter than the diameter

Graph Theory

- A graph is **connected** when there is a path between every pair of vertices.
- A **connected component** is a maximal connected subgraph of the graph. Each vertex belongs to exactly one connected component, as does each edge.
- A directed graph is called **weakly connected** if replacing all of its directed edges with undirected edges produces a connected (undirected) graph.
- A directed graph is **strongly connected** if it contains a directed path between every pair of vertices. A directed graph can be connected but not strongly connected.

Graph Connectivity





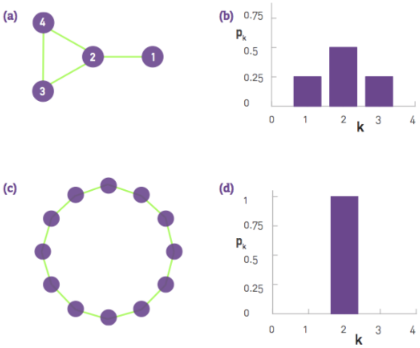
Nodes Degree

- The **degree** $\langle k \rangle$ a vertex of a graph is the number of edges incident to the vertex
- A vertex with degree 0 is called an **isolated** vertex.
- In a directed graph the number of head ends adjacent to a vertex is called the **in-degree** of the vertex and the number of tail ends adjacent to a vertex is its **out-degree**
- A vertex with in-degree$=$0 is called a **source** vertex, with out-degree$=$0 is a **sync** vertex

Degree distribution

- k_i - node degree, $k_i = 1, 2, \ldots k_{\max}$
- n_k - number of nodes with degree k, total nodes $n = \sum_k n_k$
- Degree distribution is a fraction of the nodes with degree k

$$P(k_i = k) = P(k) = P_k = \frac{n_k}{\sum_k n_k} = \frac{n_k}{n}$$



Directedness and Average Degree

NETWORK	NODES	LINKS	DIRECTED UNDIRECTED	N	L	⟨k⟩
Internet	Routers	Internet connections	Undirected	192,244	609,066	6.33
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	Directed	36,595	91,826	2.51
Email	Email addresses	Emails	Directed	57,194	103,731	1.81
Science Collaboration	Scientists	Co-authorship	Undirected	23,133	93,439	8.08
Actor Network	Actors	Co-acting	Undirected	702,388	29,397,908	83.71
Citation Network	Paper	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

Networks are Sparse Graphs

Most real-world networks are sparse $E \ll E_{max}$ (or $k \ll N - 1$)

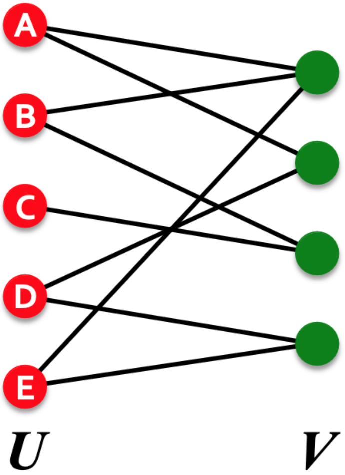
WWW (Stanford-Berkeley):	N=319,717	$\langle k \rangle$=9.65
Social networks (LinkedIn):	N=6,946,668	$\langle k \rangle$=8.87
Communication (MSN IM):	N=242,720,596	$\langle k \rangle$=11.1
Coauthorships (DBLP):	N=317,080	$\langle k \rangle$=6.62
Internet (AS-Skitter):	N=1,719,037	$\langle k \rangle$=14.91
Roads (California):	N=1,957,027	$\langle k \rangle$=2.82
Proteins (S. Cerevisiae):	N=1,870	$\langle k \rangle$=2.39

(Source: *Leskovec et al., Internet Mathematics, 2009*)

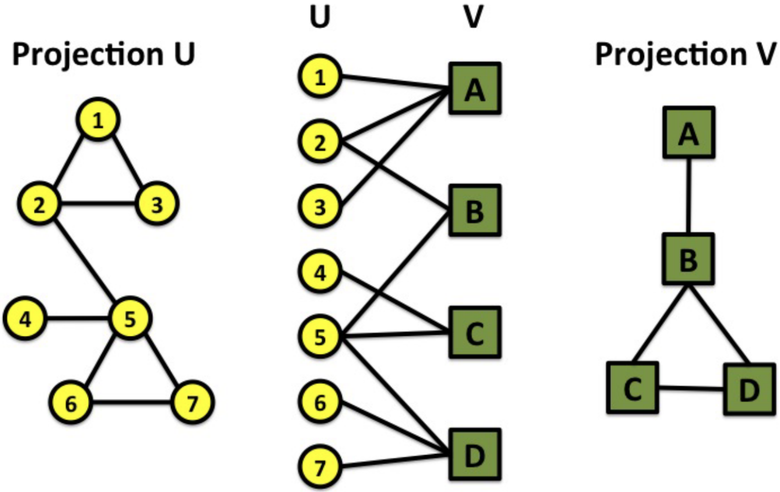
Consequence: Adjacency matrix is filled with zeros!

Bipartite Graph

- Bipartite graph is a graph whose nodes can be divided into two disjoint sets U and V such that every link connects a node in U to one in V; that is, U and V are **independent sets**

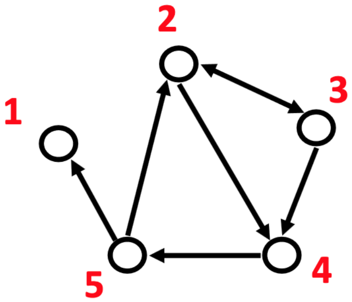


Bipartite Graph



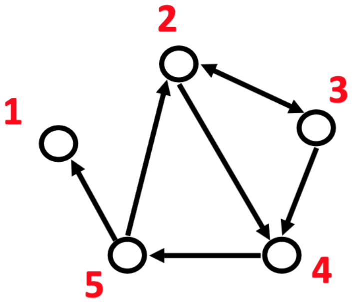
Representing Graphs: Edge List

- Represent graph as a set of edges:
  - (2, 3)
  - (2, 4)
  - (3, 2)
  - (3, 4)
  - (4, 5)
  - (5, 2)
  - (5, 1)



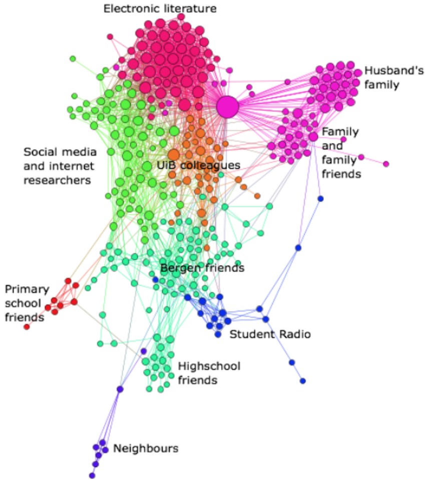
Representing Graphs: Adjacency list

- Easier to work with if network is:
  - Large
  - Sparse
- Allows us to quickly retrieve all neighbors of a given node
  - 1:
  - 2: 3, 4
  - 3: 2, 4
  - 4: 5
  - 5: 1, 2



Complex networks

- not regular, but not random
- non-trivial topology
- scale-free networks
- universal properties
- everywhere
- complex systems



Complex networks

- Power law node degree distribution: ”scale-free” networks
- Small diameter and average path length: ”small world” networks
- Hight clustering coefficient: transitivity

Textbooks

- ”Network Science”, Albert-Laszlo Barabasi, Cambridge University Press, 2016: networksciencebook.com
- ”Networks: An Introduction”. Mark Newman. Oxford University Press, 2010.
- ”Social Network Analysis. Methods and Applications”. Stanley Wasserman and Katherine Faust, Cambridge University Press, 1994.
- ”Networks, Crowds, and Markets: Reasoning About a Highly Connected World”. David Easley and John Kleinberg, Cambridge University Press 2010

References

- The Small-World Problem. Stanley Milgram. Psychology Today, Vol 1, No 1, pp 61-67, 1967
- An Experimental Study of the Small World Problem. J. Travers and S. Milgram. . Sociometry, vol 32, No 4, pp 425-433, 1969
- Planetary-Scale Views on a Large Instant-Messaging Network.
- Four Degrees of Separation. L. Backstrom, P. Boldi, M. Rosa, J.J. Leskovec and E. Horvitz. , Procs WWW 2008 Ugander, S. Vigna, WebSci '12 Procs. 4th ACM Web Science Conference, 2012 pp 33-42