Introduction to Network Science

Leonid E. Zhukov

School of Data Analysis and Artificial Intelligence
Department of Computer Science
National Research University Higher School of Economics

Structural Analysis and Visualization of Networks
Class Technicalities

- Instructor: Leonid Zhukov
- T.A.: Andrey Shestakov
- Course length: 2 modules
- Module 3: 10 lectures, 10 labs
- Schedule: Thursdays, 18.10-21.00
- Website: http://www.leonidzhukov.net/hse/2015/networks
- Emails: lzhukov@hse.ru, shestakoffandrey@gmail.com
- Programming: Python, iPython
- Python libraries: NetworkX
- Visualization: yEd, Gephi
- Software for online lectures: Fuze meeting (www.fuze.com)
Prerequisites

- Discrete Mathematics
- Linear Algebra
- Algorithms and Data Structures
- Probability Theory
- Differential Equations
- Programming in Python
Vectors, matrices
Sparse matrices
Eigenvalues and eigenvectors
Graph concepts

- Graph $G(V, E)$, $V$ - vertices, $E$ - edges
- Directed, undirected, simple, weighted
- Connected and strongly connected graph
- Connected component
- Path, cycles
- Path length, shortest path
- Tree
Graph Algorithms

- BFS - breadth first search
- DFS - depth first search
- Connected components
- Shortest paths
Graphs and matrices

Graph $G(n, m)$, adjacency matrix $A_{ij}$, edge $i \rightarrow j$

Storage:
- adjacency list
- sparse matrix
Textbooks

Module 3 lectures

1. Introduction to network science
2. Power laws
3. Random graphs
4. Small world and dynamical growth models
5. Node centrality metrics
6. Link analysis and network structure
7. Structural equivalence and similarity
8. Network communities
9. Special classes of networks
10. Network visualization
Network science

- Sociology (SNA)
- Mathematics (Graphs)
- Computer Science (Graphs)
- Statistical Physics (Complex networks)
- Economics (Networks)
- Bioinformatics (Networks)
Terminology

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

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

Barrett Lyon, OPTE.org, class C networks
Examples: Political blogs
Examples: Twitter
Examples: Finance
Examples: Transportation
Examples: Biology

Protein-protein interactions
Examples: Organization
Examples: Facebook communities structure
Examples: Facebook

Social graph 500 mln people, Paul Butler, 2010
Complex network

1. Power law degree distribution
2. Small diameter and average path length: "small world"
3. High clustering coefficient
Power law
Power law
"Any two people are on average separated no more than by six intermediate connections"

- John Guare play (1991) and movie (1993), "Six Degrees of Separation"
"The small-world problem”. Stanley Milgram, 1967

HOW TO TAKE PART IN THIS STUDY

1. ADD YOUR NAME TO THE ROSTER AT THE BOTTOM OF THIS SHEET, so that the next person who receives this letter will know who it came from.

2. DETACH ONE POSTCARD. FILL IT OUT AND RETURN IT TO HARVARD UNIVERSITY. No stamp is needed. The postcard is very important. It allows us to keep track of the progress of the folder as it moves toward the target person.

3. IF YOU KNOW THE TARGET PERSON ON A PERSONAL BASIS, MAIL THIS FOLDER DIRECTLY TO HIM (HER). Do this only if you have previously met the target person and know each other on a first name basis.

4. IF YOU DO NOT KNOW THE TARGET PERSON ON A PERSONAL BASIS, DO NOT TRY TO CONTACT HIM DIRECTLY. INSTEAD, MAIL THIS FOLDER (POSTCARDS AND ALL) TO A PERSONAL ACQUAINTANCE WHO IS MORE LIKELY THAN YOU TO KNOW THE TARGET PERSON. You may send the folder
Stanley Milgram’s 1969 experiment

- 296 volunteers, 217 sent
- 196 in Nebraska (1300 miles)
- 100 in Boston (25 miles)
- Target in Boston
- Name, address, occupation, job, hometown
Stanley Milgram’s 1969 experiment

- Reached the target $N = 64$, 29%
- Average chain length $<L> = 5.2$
- Channeles: hometown $<L> = 6.1$
- Business contacts $<L> = 4.6$
- From Boston $<L> = 4.4$
- From Nebraska $<l> = 5.7$
Email graph:
D. Watts (2001), 48,000 senders, $< L > \approx 6$

MSN Messenger graph:
J. Lescovec et al (2007), 240mln users, $< L > \approx 6.6$

Facebook graph:
L. Backstrom et al (2012), 721 mln users, $< L > \approx 4.74$
Simple model

Cayley tree (Bethe lattice) - graph where every node connects to \( z \) neighbors

An estimate:

\[
z^d = N, \quad d = \log N / \log z
\]

\( N \approx 6.7 \text{ bln}, \quad z = 50 \text{ friends, } d \approx 5.8.\)
Reviews

- Statistical mechanics of complex networks, R. Albert and A-L. Barabasi, Rev. Mod. Phys. 74, 47, 2002
References


