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

NATIONAL RESEARCH UNIVERSITY

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

Linear Algebra

- Vectors, matrices
- Sparse matrices
- Eigenvalues and eigenvectors

Graph concepts

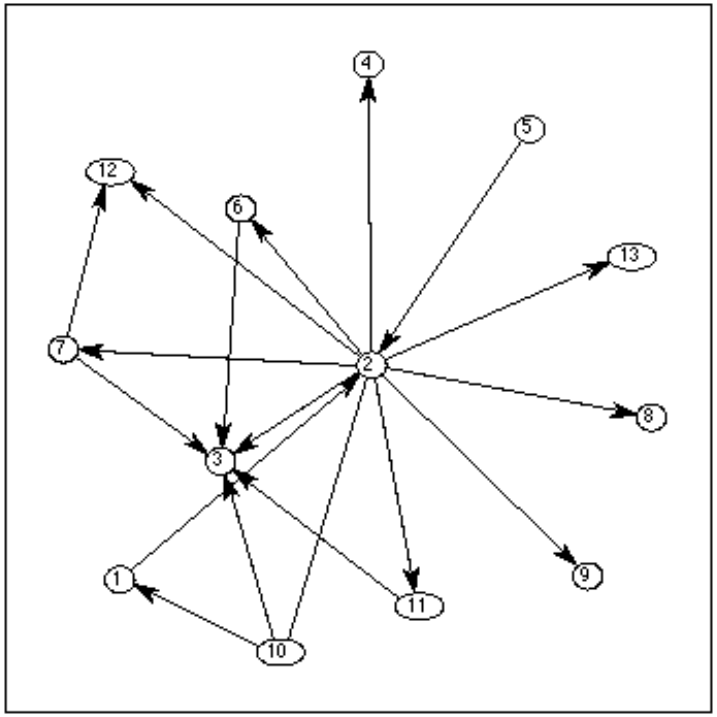
- Graph $G(V, E)$, V - vertices, E - edges
- Directed, undirected, simple, weighted
- Connected and strongly connected graph
- Connected component
- Path, cycles
- Path length, shortest path
- Tree

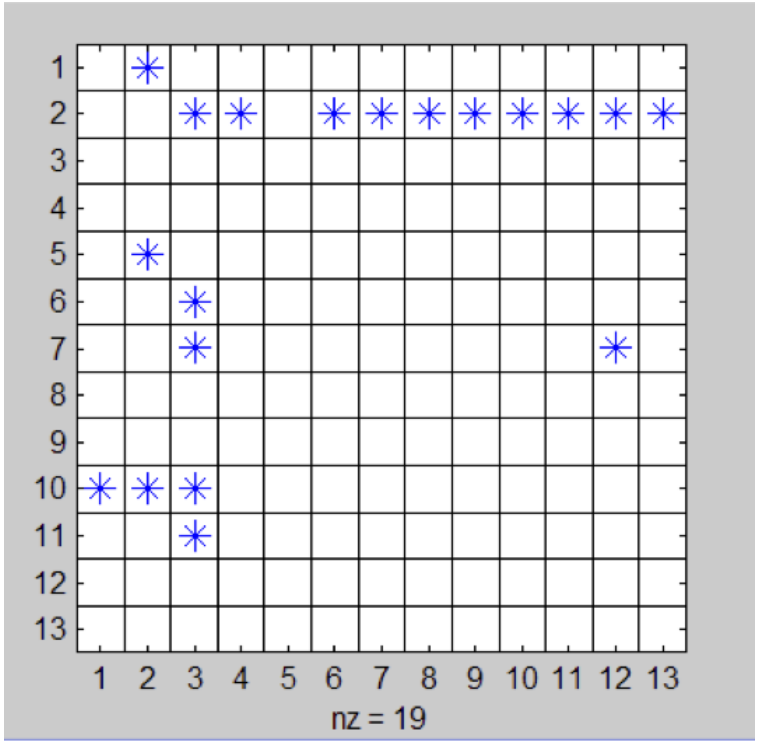
Graph Algorithms

- BFS - breadth first search
- DFS - depth first search
- Connected components
- Shortest paths

Graphs and matrices

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





Storage:

- adjacency list
- sparse matrix

Textbooks

- "Networks: An Introduction". Mark Newman. Oxford University Press, 2010.
- "Networks, Crowds, and Markets: Reasoning About a Highly Connected World". David Easley and John Kleinberg, Cambridge University Press 2010.
- "Social and Economic Networks". Matthew O. Jackson. Princeton University Press, 2010.
- "Social Network Analysis. Methods and Applications". Stanley Wasserman and Katherine Faust, Cambridge University Press, 1994
- "Network Analysis. Methodological Foundations". Eds. Ulrik Brandes and Thomas Erlebach. Springer, 2004

Module 3 lectures

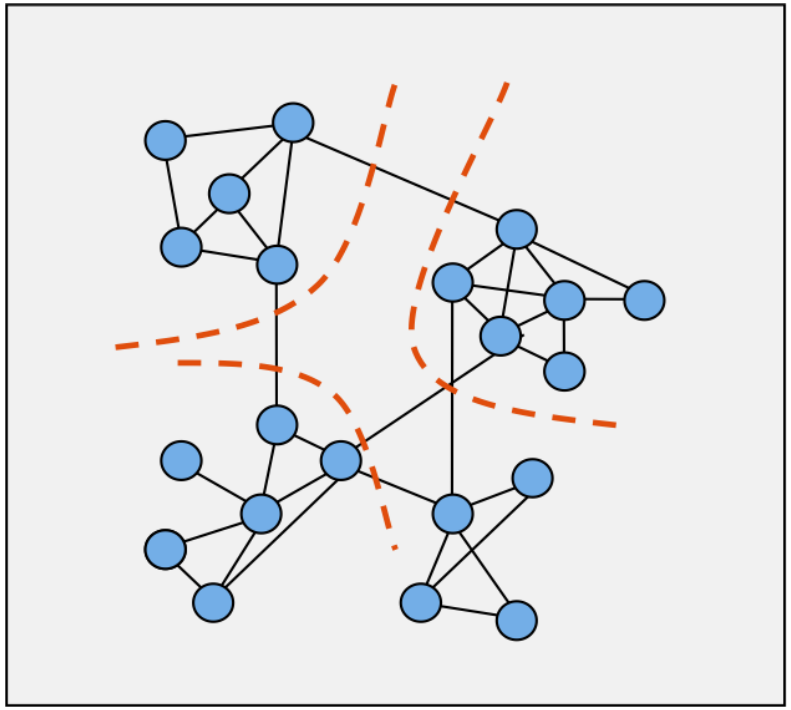
1. Introduction to network science
2. Power laws
3. Random graphs
4. Small world and dynamical growth models
5. Node centrality metrics
6. Link analysis and network structure
7. Structural equivalence and similarity
8. Network communities
9. Special classes of networks
10. Network visualization

Network science

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

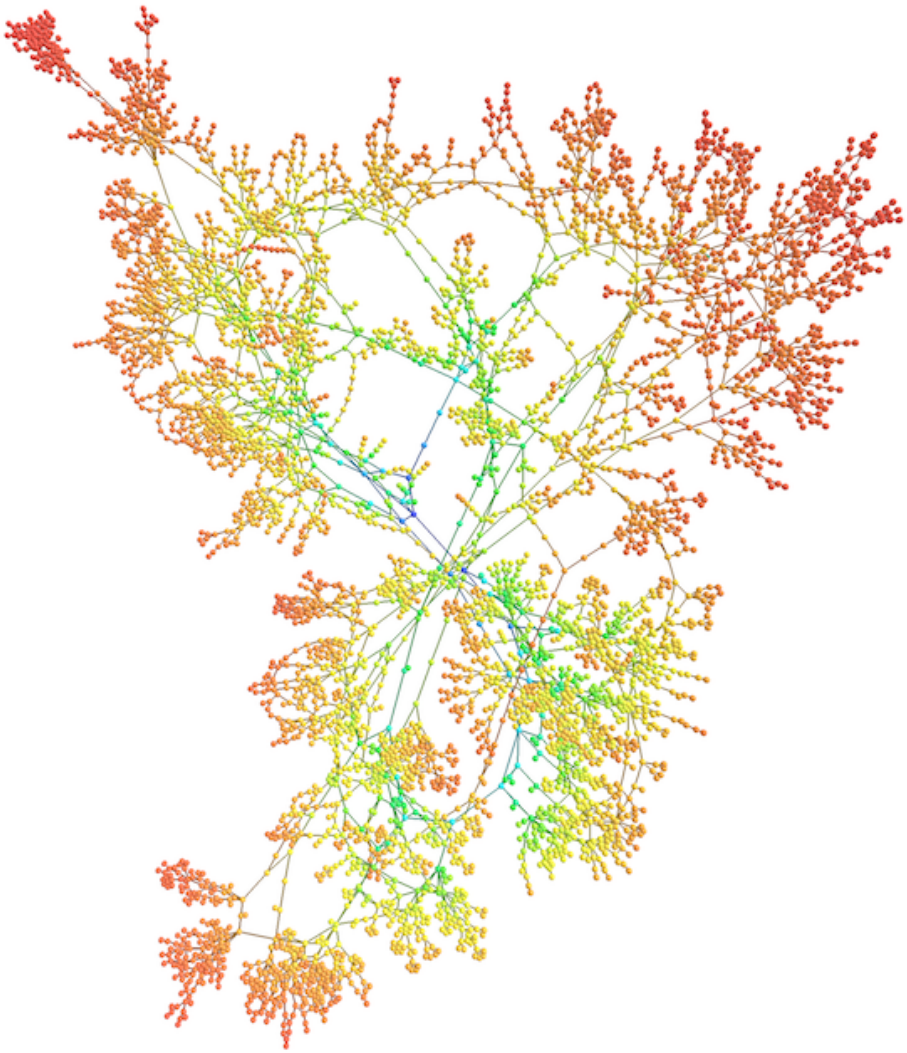
Terminology

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

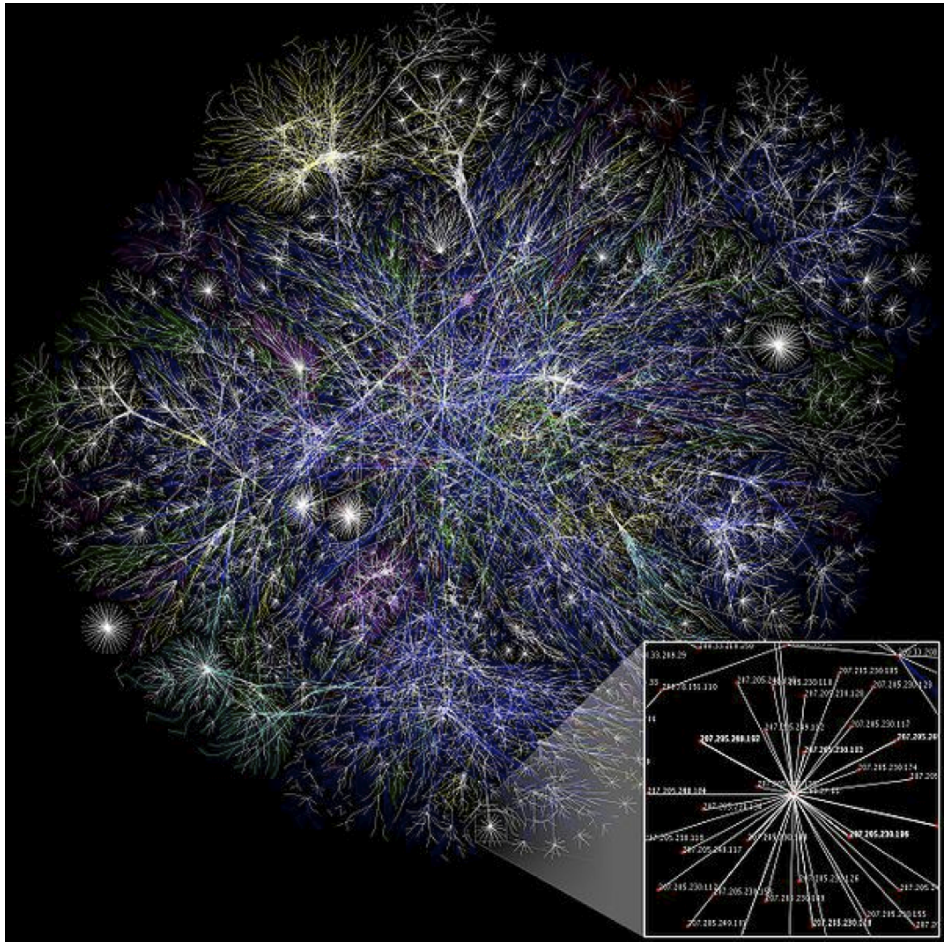


Complex networks

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

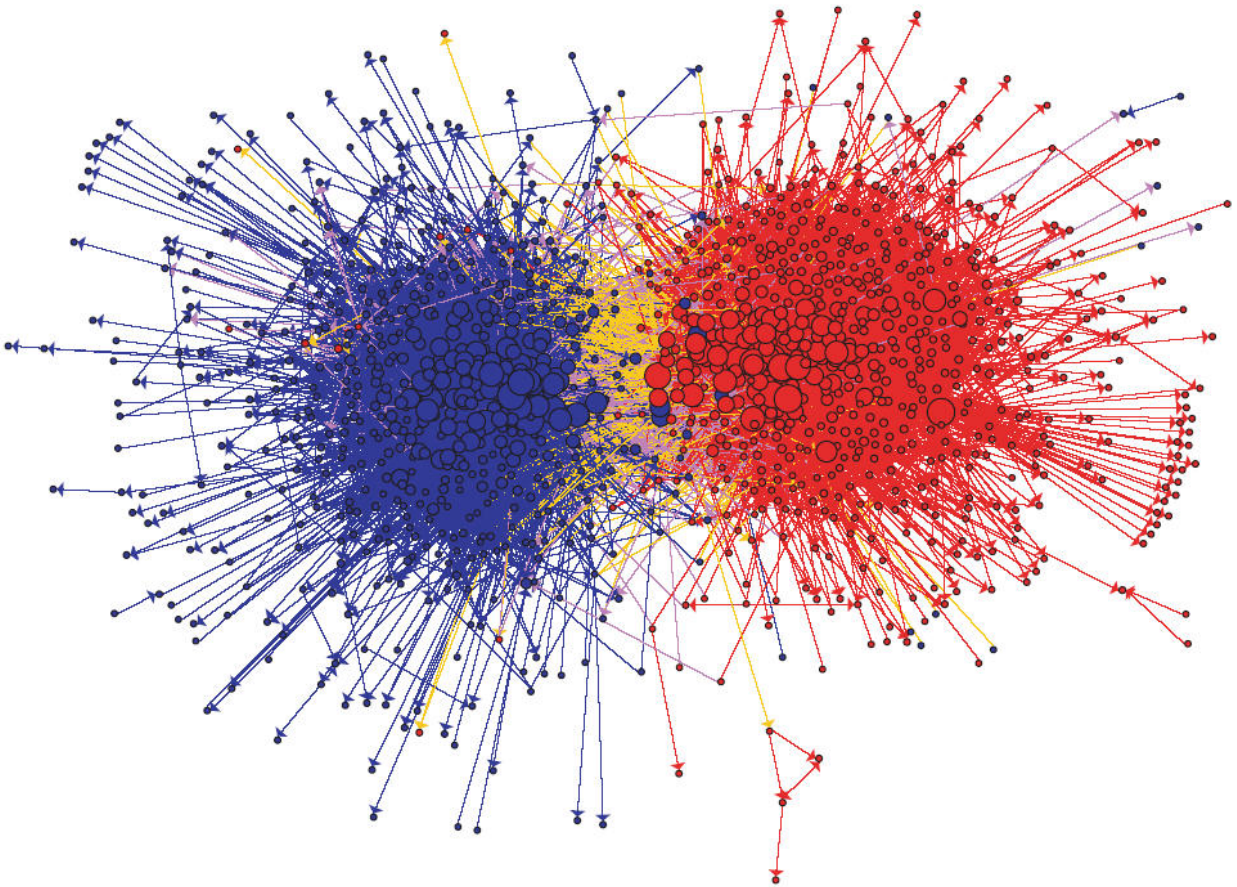


Examples: Internet

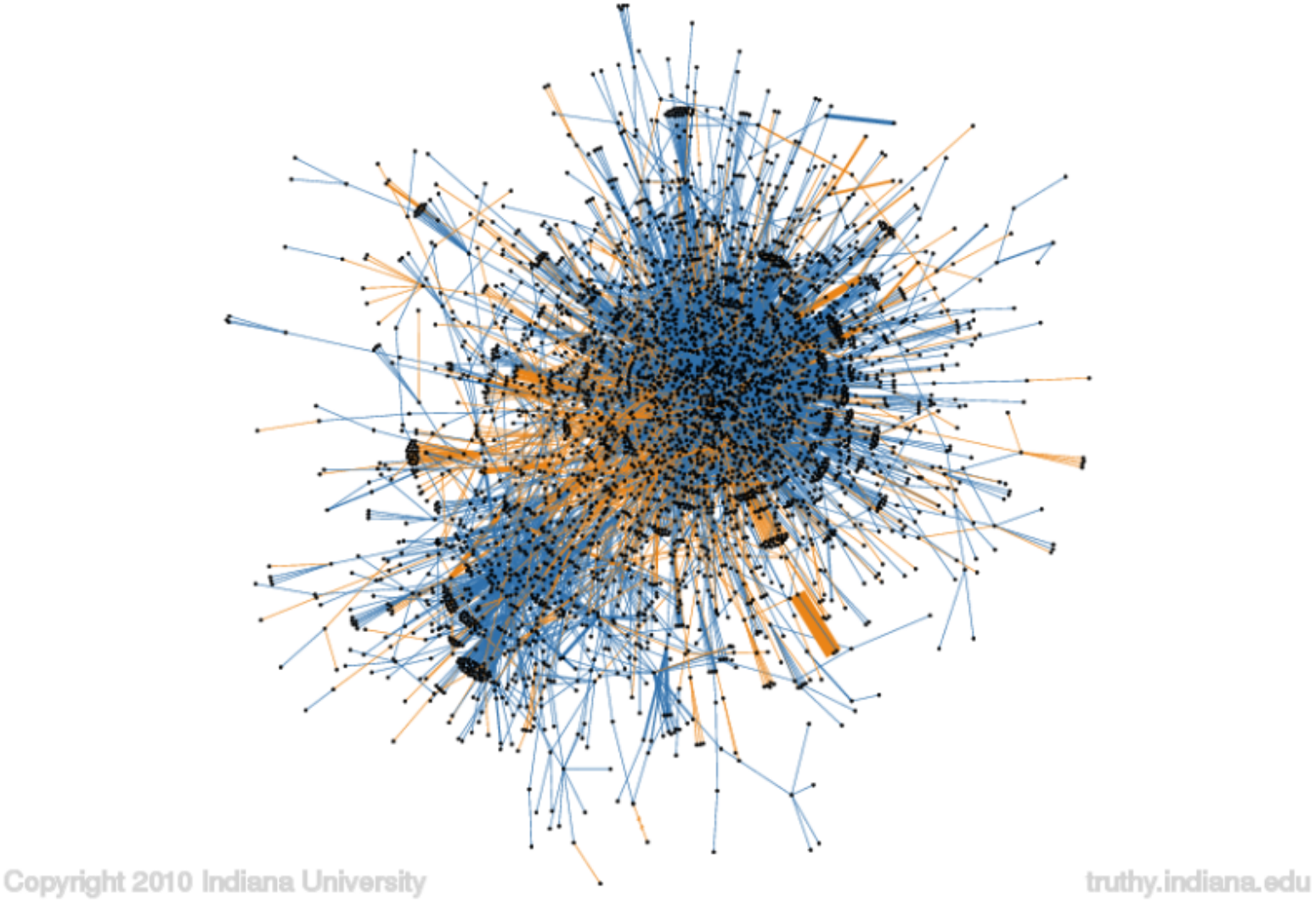


Barrett Lyon, OPTE.org, class C networks

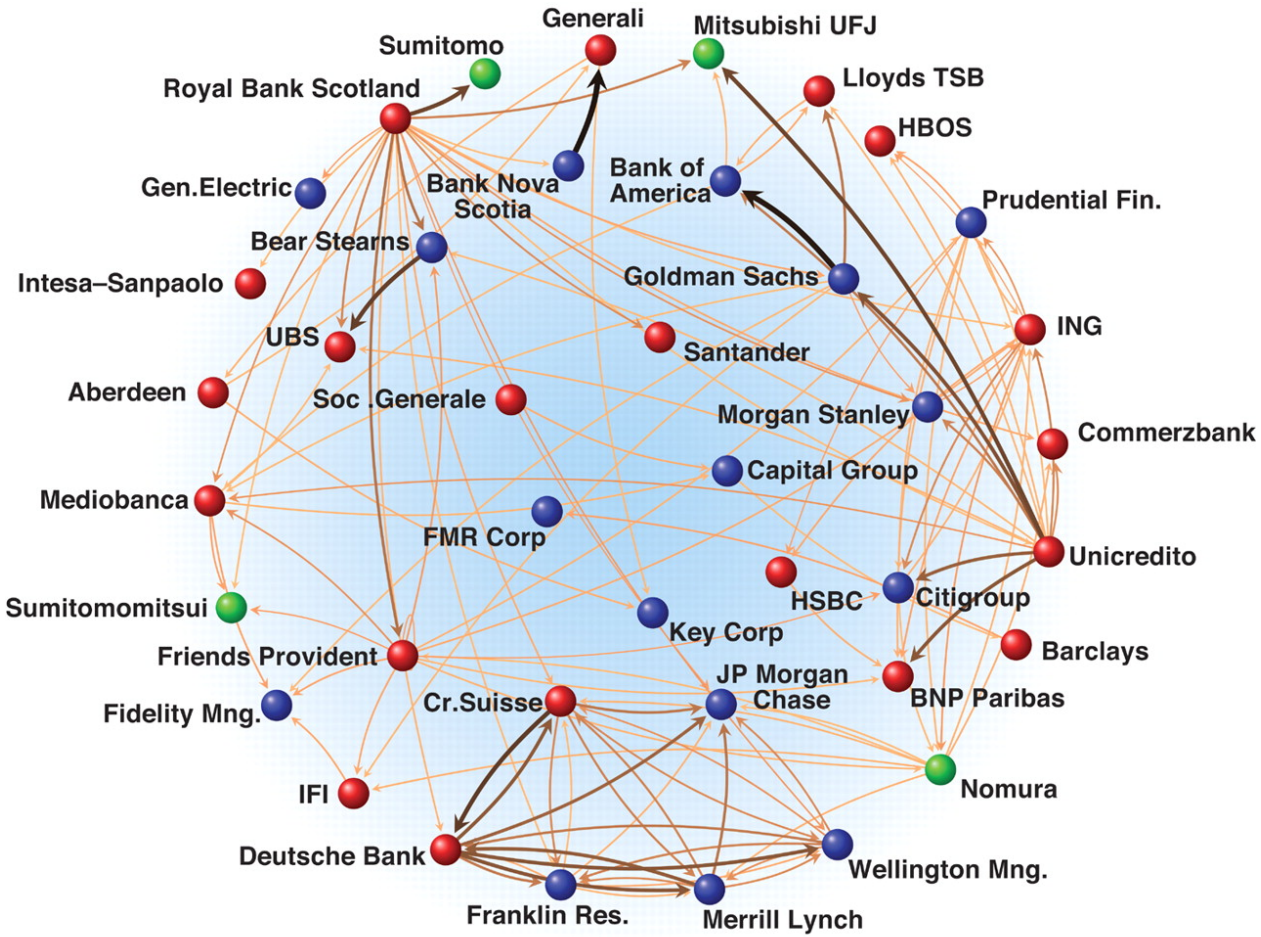
Examples: Political blogs



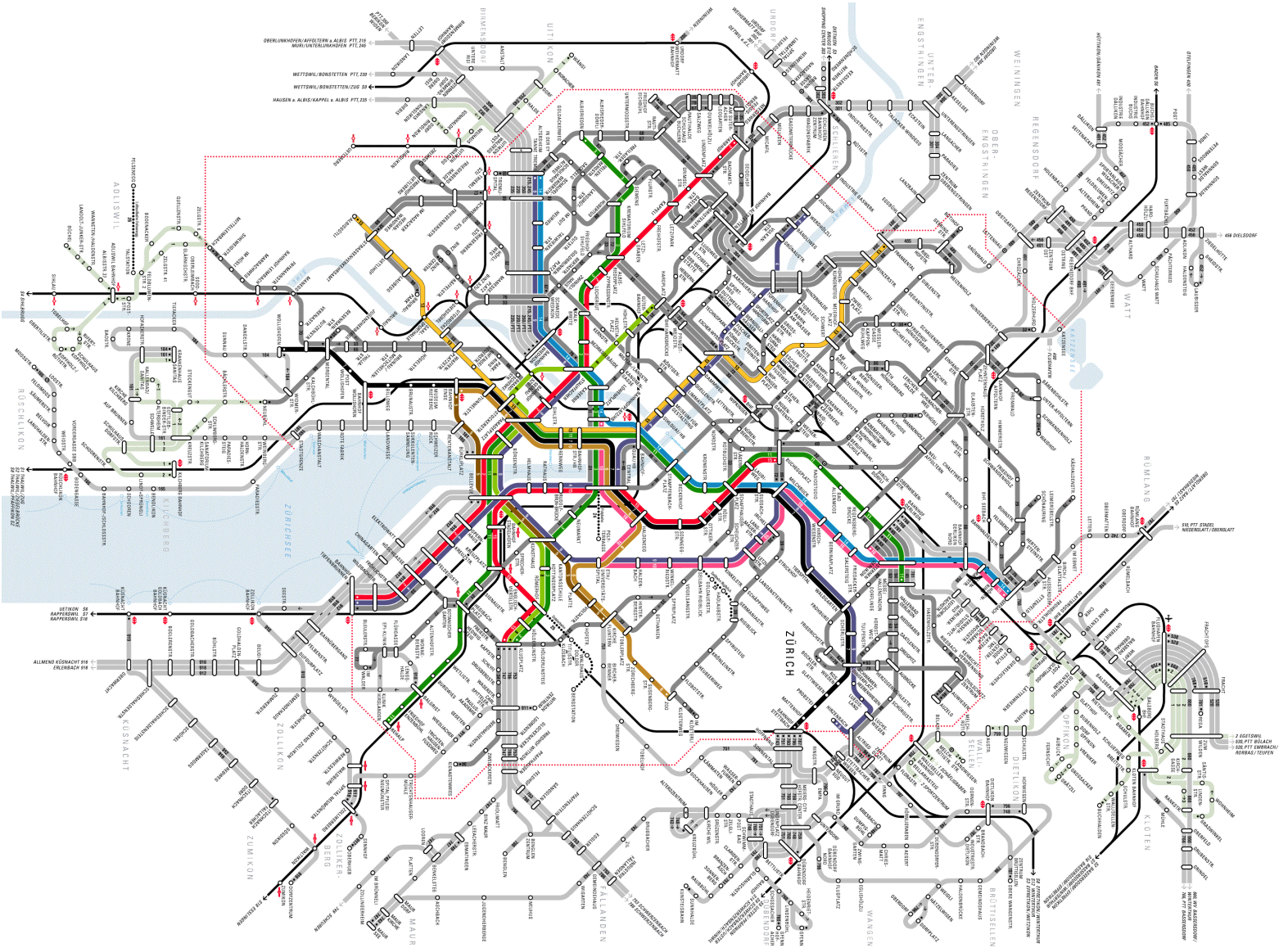
Examples: Twitter



Examples: Finance

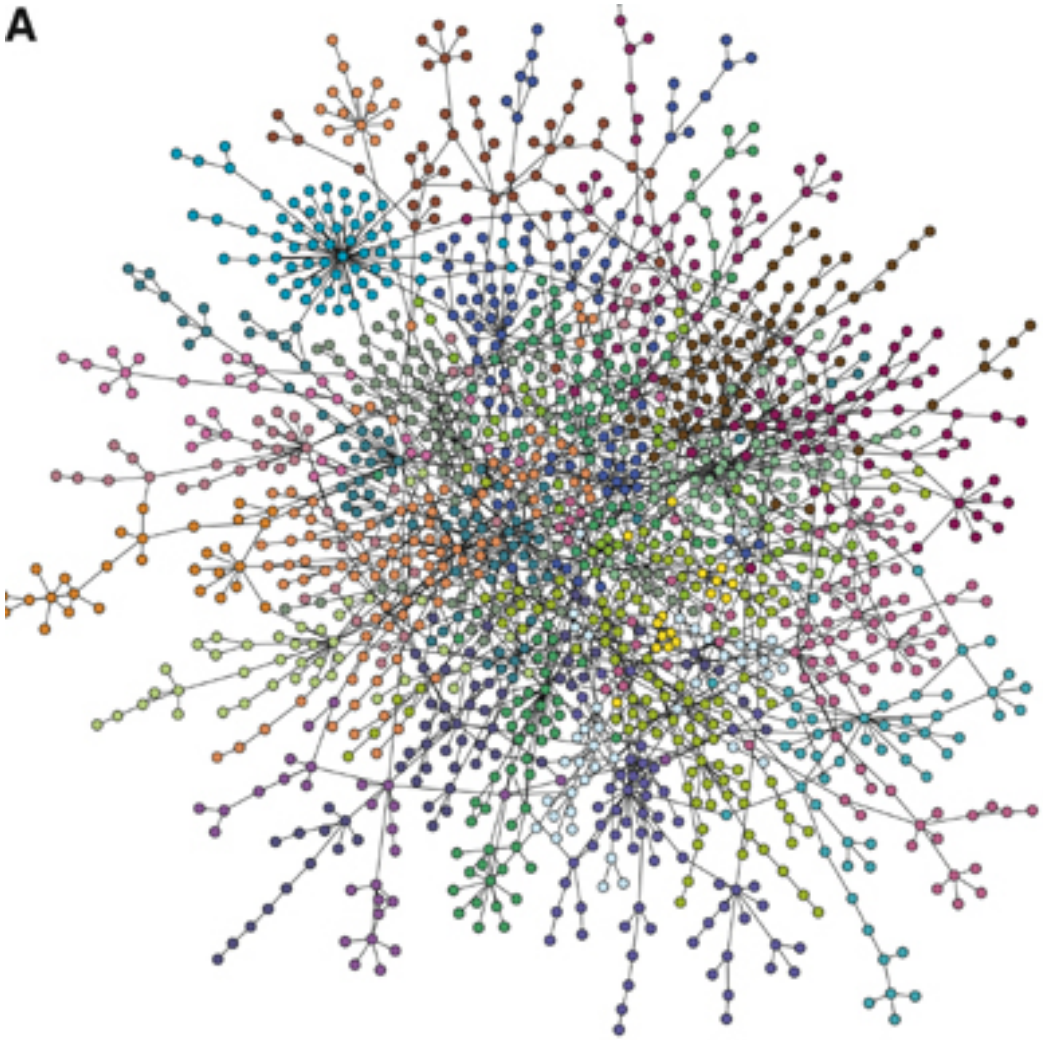


Examples: Transportation

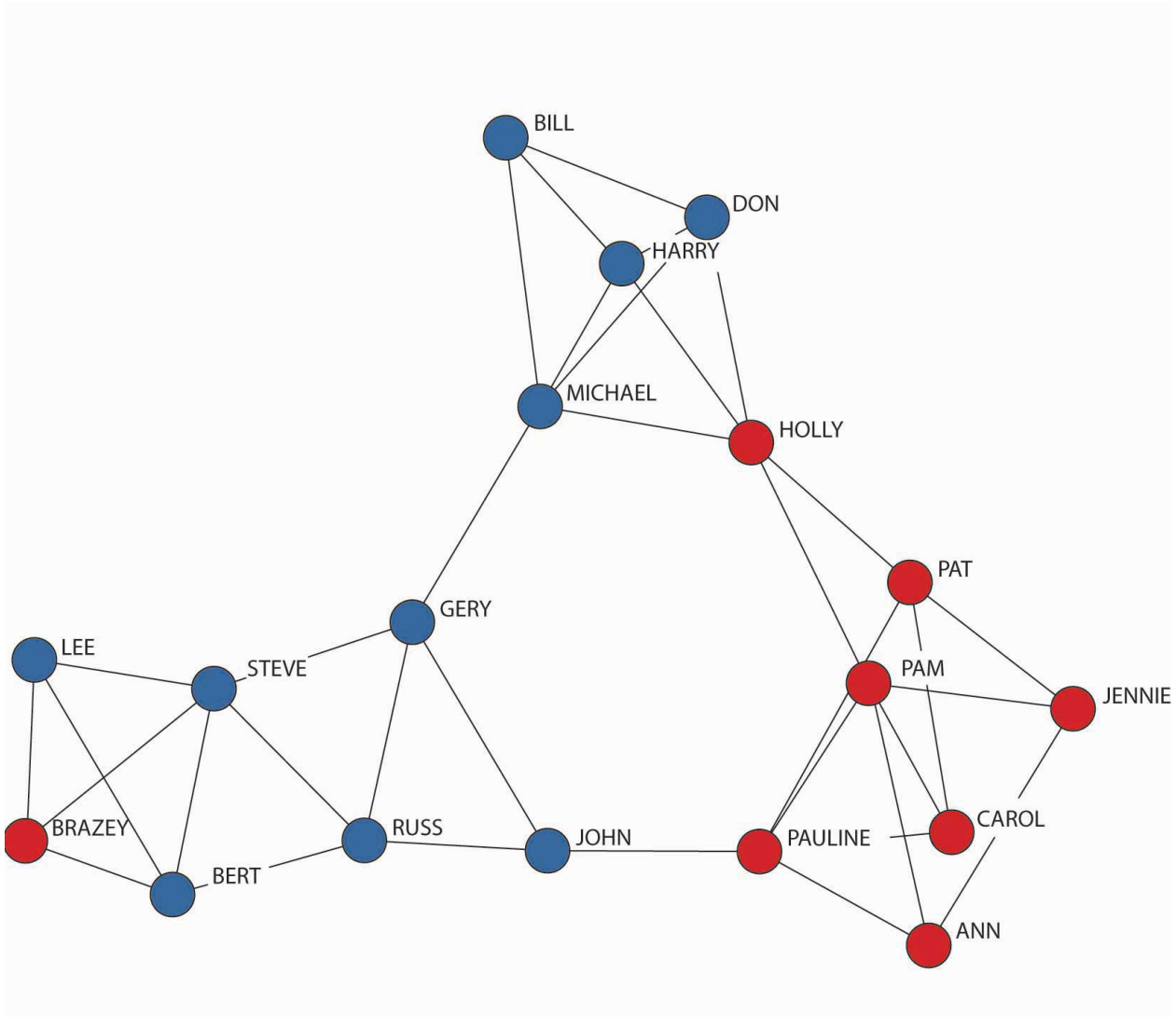


Examples: Biology

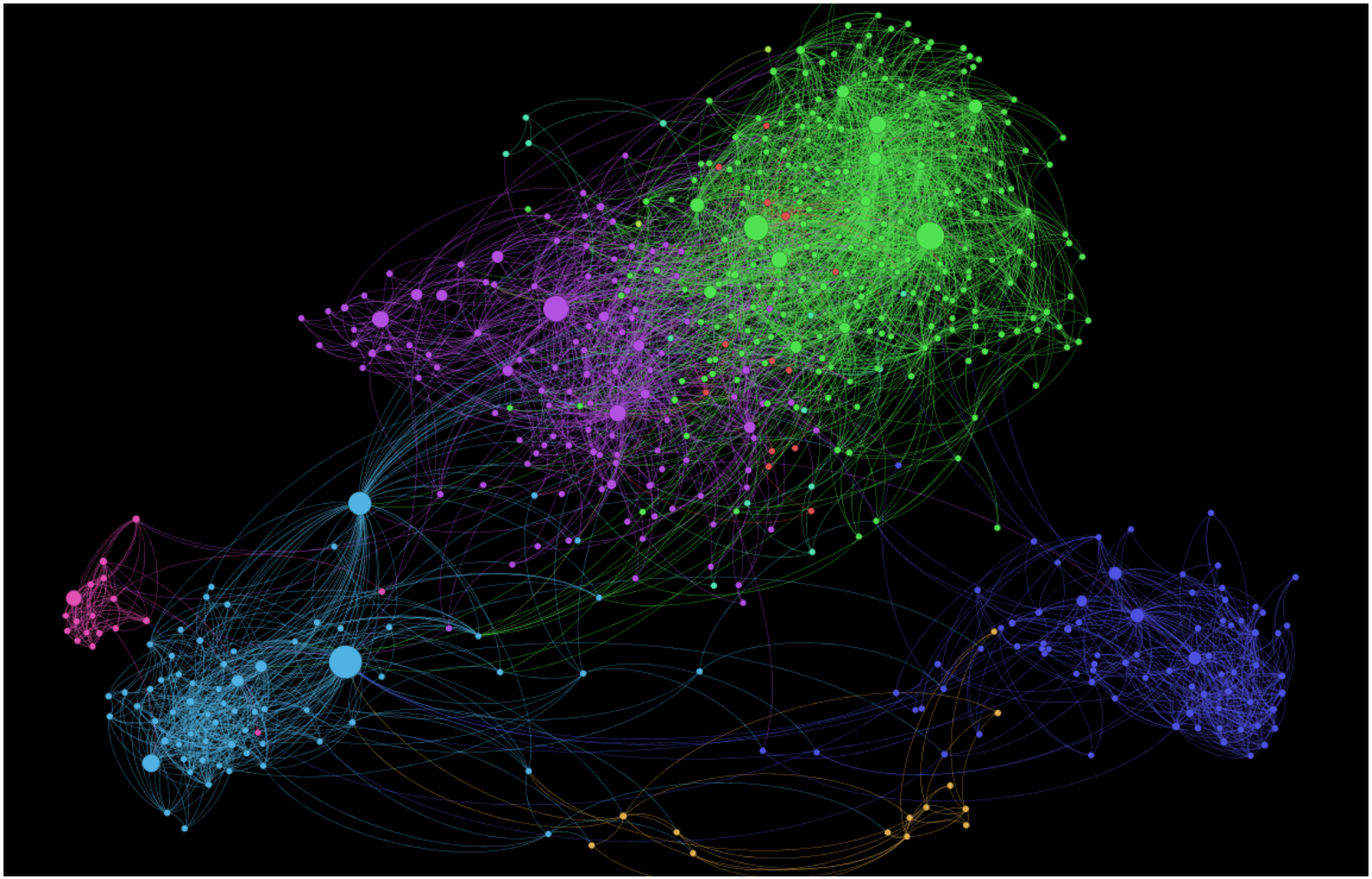
Protein-protein interactions



Examples: Organization



Examples: Facebook communities tructure



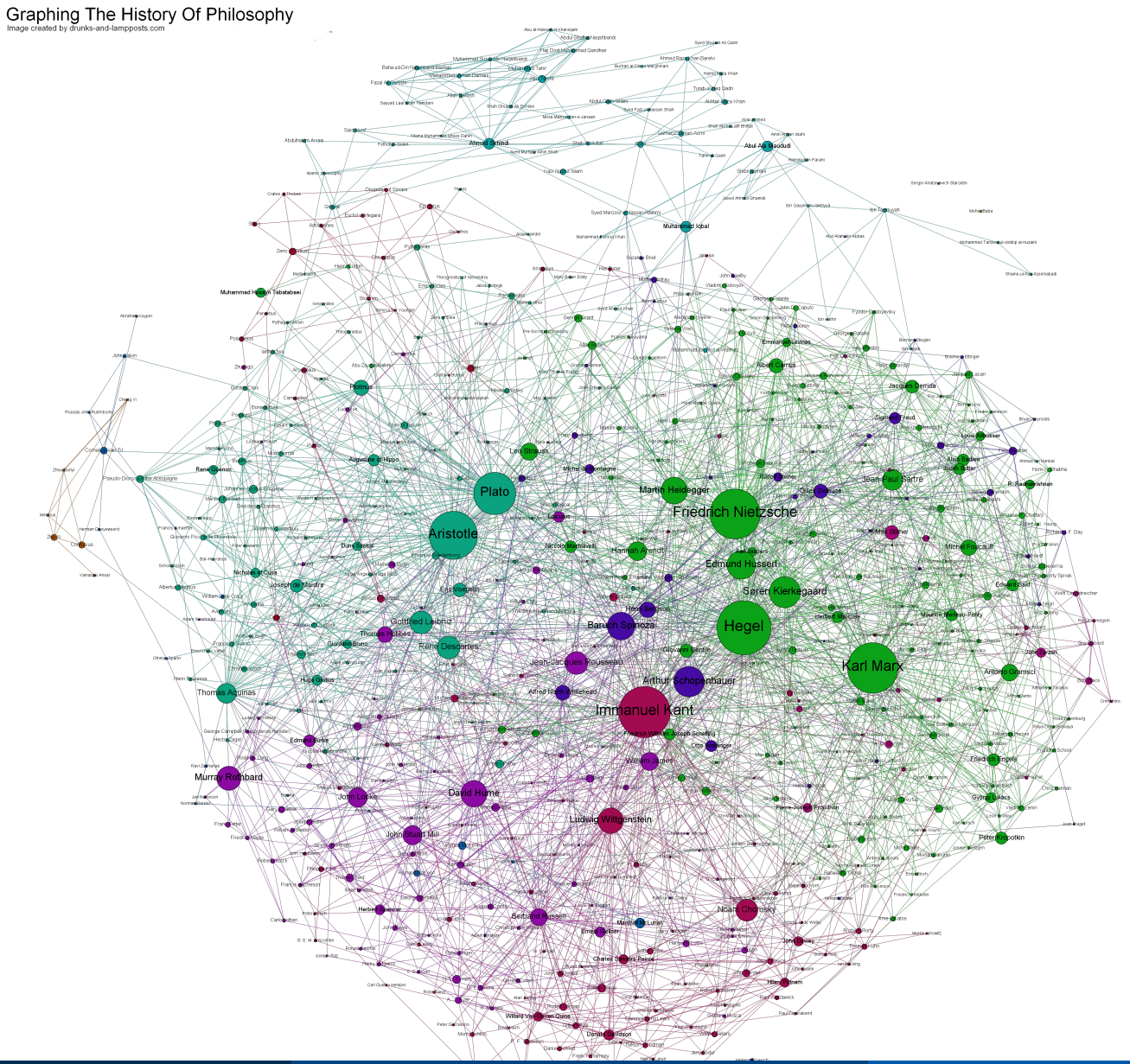
Examples: Facebook

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. Hight clustering coefficient

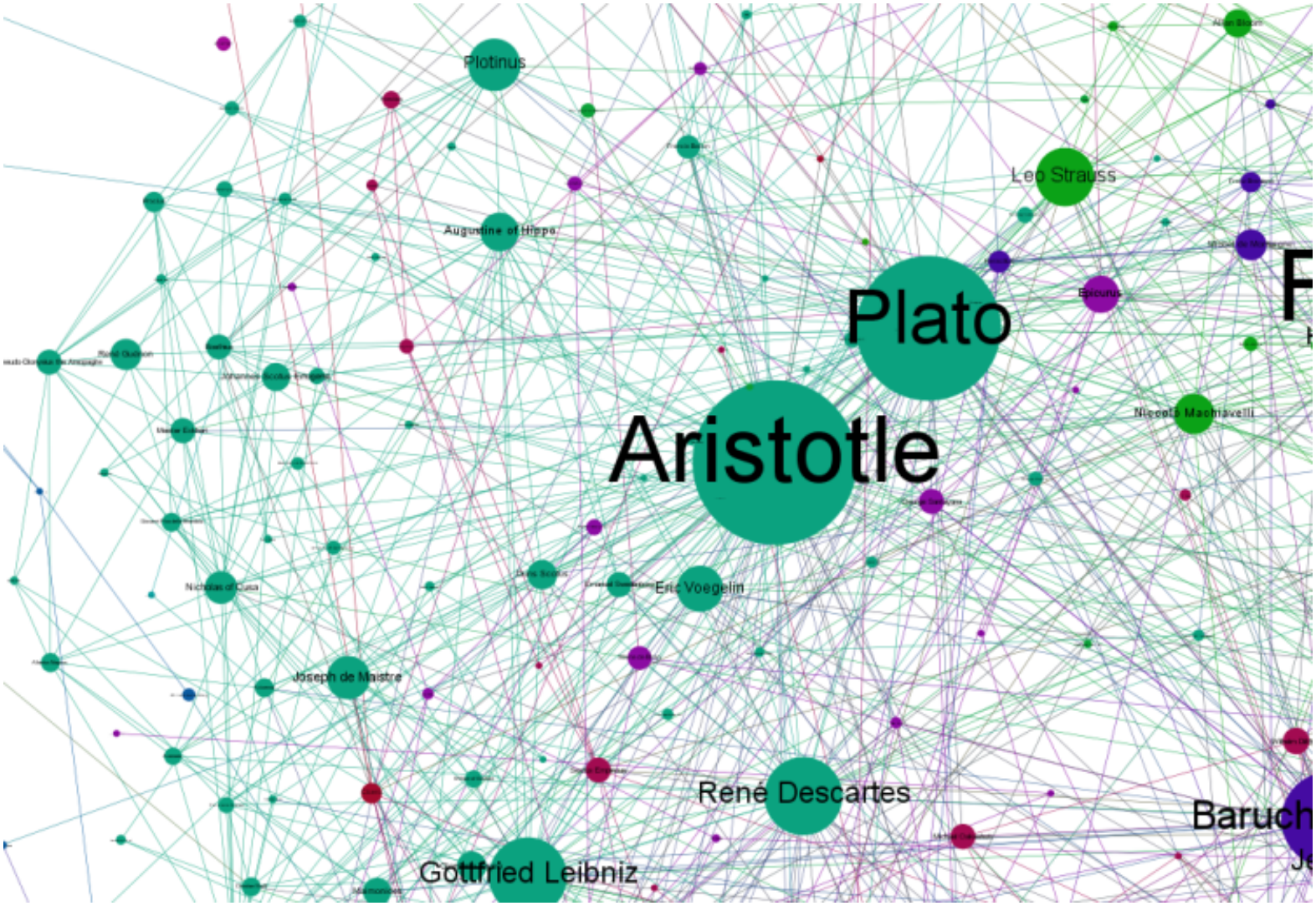
Power law



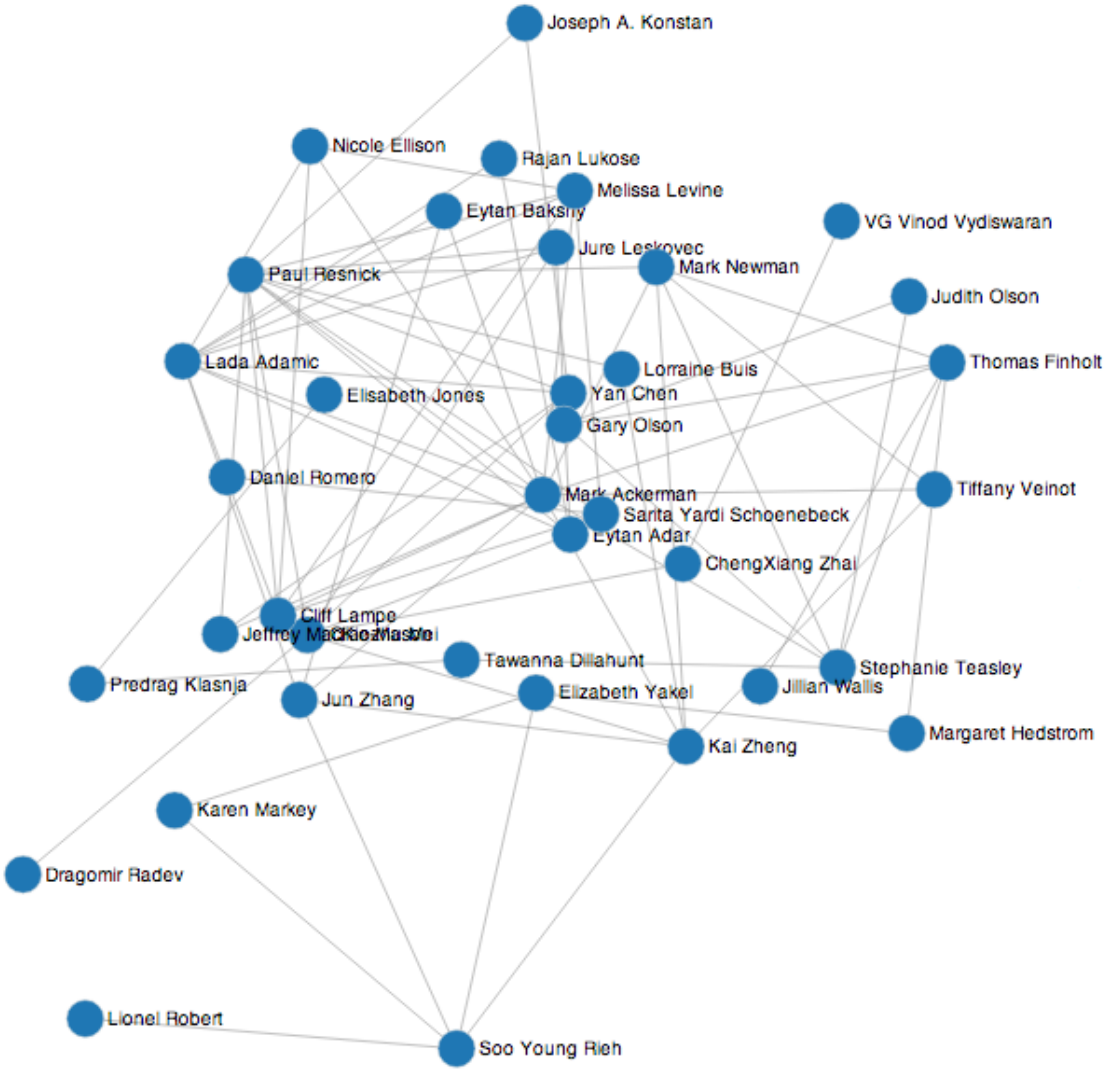
Graphing The History Of Philosophy

Image created by drunks-and-lampposts.com

Power law


Plotinus
Leo Strauss
Augustine of Hippo
Plato
Aristotle
Nicolò Machiavelli
Eric Voegelin
Joseph de Maistre
René Descartes
Gottfried Leibniz

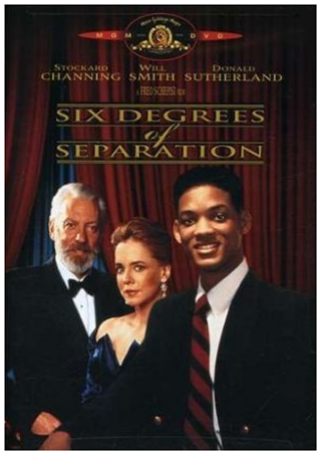
High clustering



Six degrees of separation

"Any two people are on average separated no more that by six intermediate connections"

- Frigyes Karinthy, short story "Chain-Links", 1929 book.
- John Guare play (1991) and movie (1993), "Six Degrees of Separation"



Small world


© Al Satterwhite

An Experimental Study of the Small World Problem*

JEFFREY TRAVERS
Harvard University
AND
STANLEY MILGRAM
The City University of New York

Arbitrarily selected individuals (N=296) in Nebraska and Boston are asked to generate acquaintance chains to a target person in Massachusetts, employing "the small world method" (Milgram, 1967). Sixty-four chains reach the target person. Within this group the mean number of intermediaries between starters and targets is 5.2. Boston starting chains reach the target person with fewer intermediaries than those starting in Nebraska; subpopulations in the Nebraska group do not differ among themselves. The funneling of chains through sociometric "stars" is noted, with 48 per cent of the chains passing through three persons before reaching the target. Applications of the method to studies of large scale social structure are discussed.

- "The small-world problem". Stanley Milgram, 1967
- "An experimental study of the small world problem", J. Travers, S. Milgram, 1969

Stanley Milgram's 1969 experiment

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.

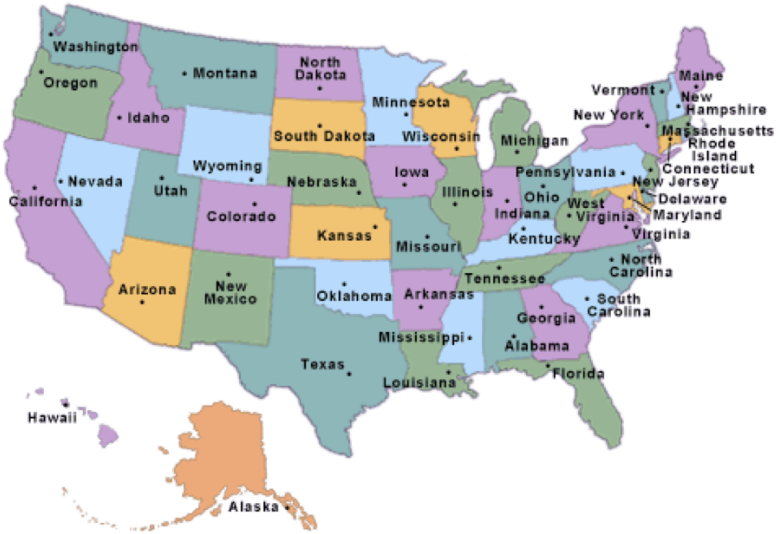
2. DETACH ONE POSTCARD. FILL IT OUT AND RETURN IT TO HARVARD UNIVERSITY. No stamp is needed. The postcard is very important. It allows us to keep track of the progress of the folder as it moves toward the target person.

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

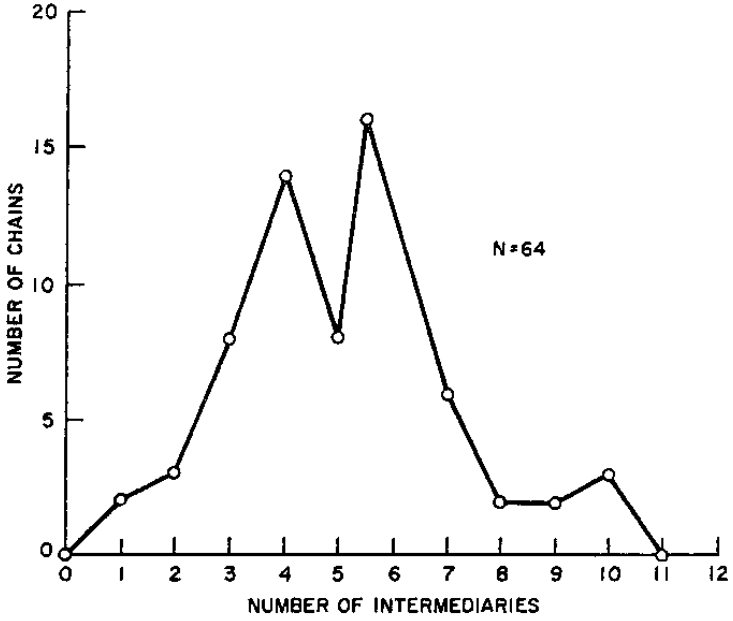
Stanley Milgram's 1969 experiment

- 296 volunteers, 217 sent
- 196 in Nebraska (1300 miles)
- 100 in Boston (25 miles)
- Target in Boston
- Name, address, occupation, job, hometown



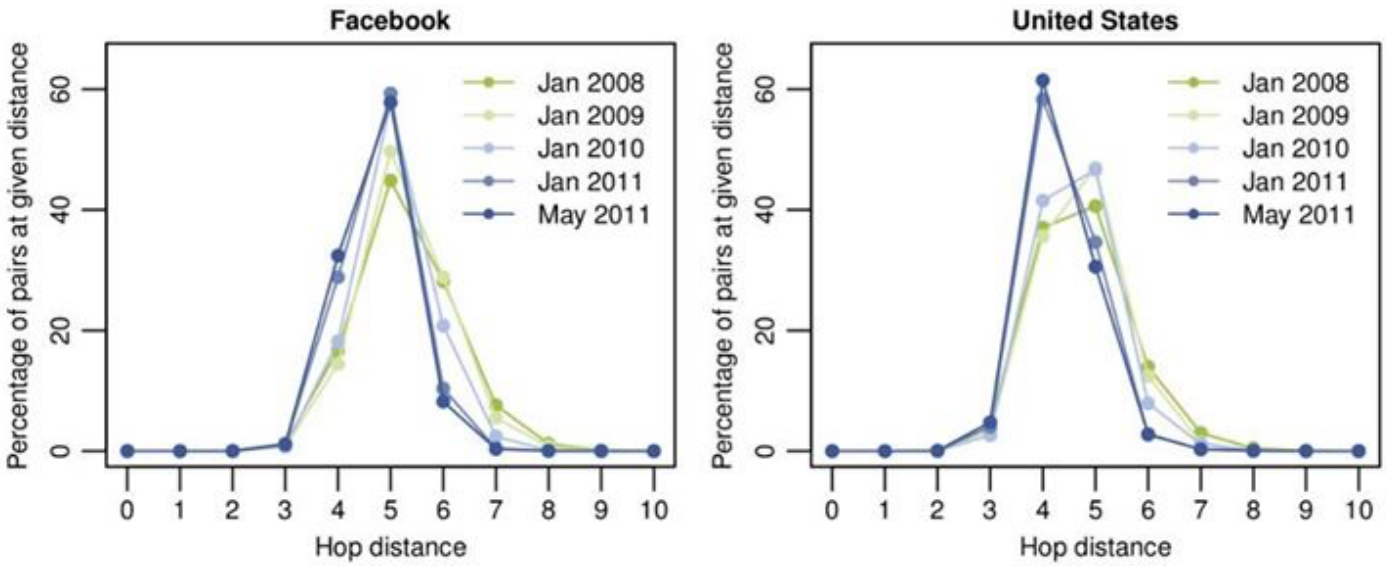
Stanley Milgram's 1969 experiment

- 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



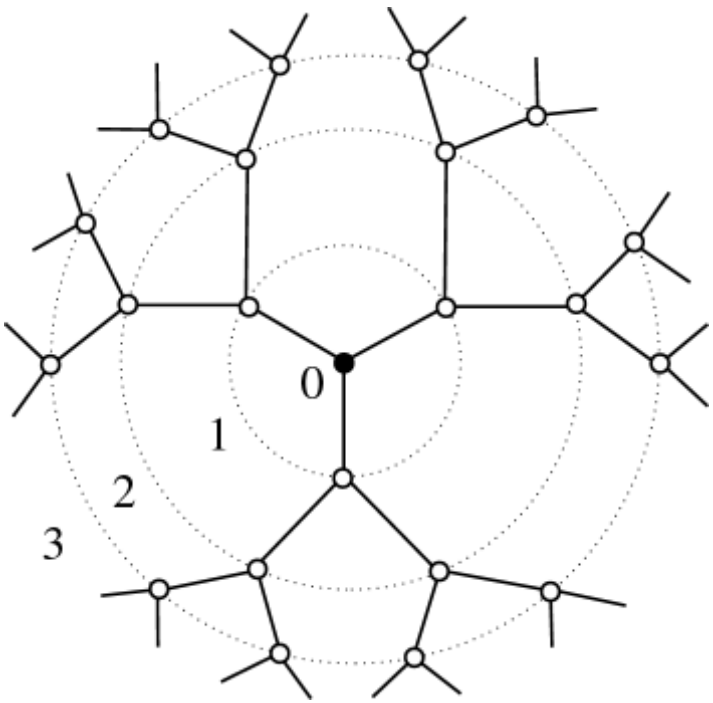
Small world

- 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 connect to z neighbors



An estimate: $z^d = N$, $d = \log N / \log z$
$N \approx 6.7$ bln, $z = 50$ friends, $d \approx 5.8$.

Reviews

- Statistical mechanics of complex networks, R. Albert and A-L. Barabasi, Rev. Mod. Phys. 74, 47, 2002
- The Structure and Function of Complex Networks, M. E. J. Newman, SIAM Review, 45, pp 167-256, 2003
- Evolution of Networks, S. N. Dorogovtsev and J. F. F. Mendes, Adv. Phys., pp 1079-1187, 2002
- Complex networks: Structure and dynamics, S. Boccaletti et al., Physics Reports, Volume 424, Issue 4-5, p. 175-308, 2006

References

- Scale free networks. A.-L. Barabasi, E. Bonabeau, Scientific American 288, 50-59 (2003)
- Scale-Free Networks: A Decade and Beyond. A.-L. Barabasi, Science 325, 412-413 (2009)
- The Physics of Networks. Mark Newman, Physics Today, November 2008, pp. 33–38.

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. J. Leskovec and E. Horvitz. , Procs WWW 2008
- Four Degrees of Separation. L. Backstrom, P. Boldi, M. Rosa, J. Ugander, S. Vigna, WebSci '12 Procs. 4th ACM Web Science Conference, 2012 pp 33-42