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

Network Science

NATIONAL RESEARCH
UNIVERSITY

Class Details

- Instructors: Leonid Zhukov, Ilya Makarov
- Teaching Assistant: Roman Griculyak
- Course length: Modules 3-4
- Module 3: 10 lectures, 5 labs, 4 homeworks
- Schedule:
  Lectures - Saturday, 16.40-19.30
  Seminars/labs - Tuesday every other week, 18.10-21.00
- Website: http://www.leonidzhukov.net/hse/2017/networkscience
- Emails: lzhukov@hse.ru, iamakarov@hse.ru, rgriculyak@hse.ru

- Programming: Python, iPython notebooks
- Python libraries: NetworkX
- Visualization: yEd, Gephi
- Software for online lectures: Zoom meeting (www.zoom.us)

Prerequisites

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

Textbooks

- "Network Science", Albert-Laszlo Barabasi, Cambridge University Press, 2016. http://barabasi.com/networksciencebook/
- "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

Topics

1. Statistical properties and network modeling
2. Network structure and dynamics
3. Processes on networks
4. Predictions on networks (ML)
5. Applications

Module 3 lectures

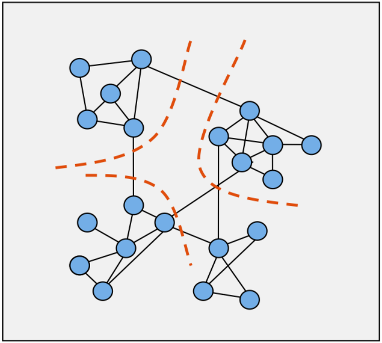
1. Introduction to network science
2. Power laws
3. Random graphs
4. Small world and dynamical growth models
5. Centrality measures
6. Link analysis
7. Structural equivalence
8. Network communities
9. Graph partitioning algorithms
10. Community detection

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



Graph theory

- 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 $A_{ij} \in \{0, 1\}$ - availability of an edge between nodes i and j
- 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, $A_{ij} \neq A_{ji}$

Graph theory

- A walk is a sequence of edges $\{i_1, i_2\}, \{i_2, i_3\}, .., \{i_{k-1}, i_{k-2}\}$
- The length of a walk (or a path) is the number of edges on that walk
- A cycle is a closed walk starting and ending at the same vertex.
- 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

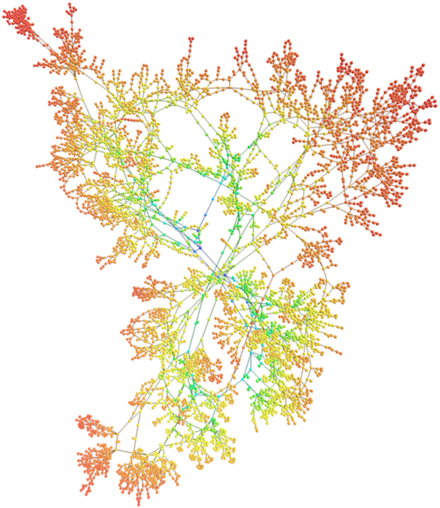
- A graph is connected when there is a path between every pair of vertices.
- A connected component is a maximal connected subgraph of the graph. Each vertex belongs to exactly one connected component, as does each edge.
- A directed graph is called weakly connected if replacing all of its directed edges 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 theory

- The degree of a vertex of a graph is the number of edges incident to the vertex
- A vertex with degree 0 is called an isolated vertex.
- A vertex with degree 1 is called a leaf vertex.
- If each vertex of the graph has the same degree k the graph is called a k-regular graph
- 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

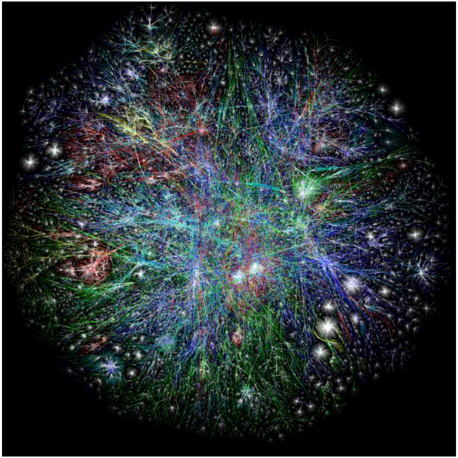
Complex networks

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



Examples: Internet

Internet traffic routing (BGP)



Barret Lyon, 2003

Examples: Social network - Facebook friendship

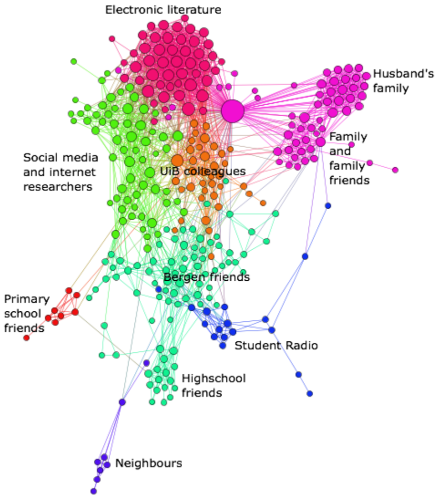
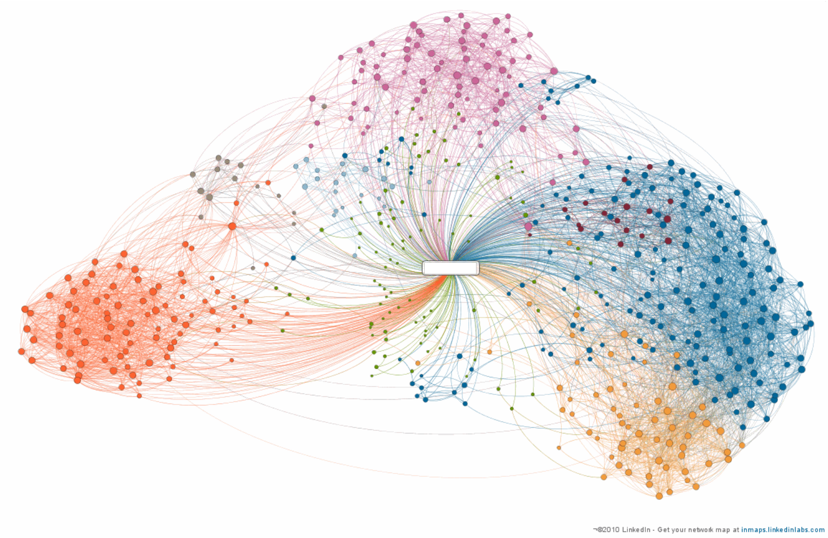


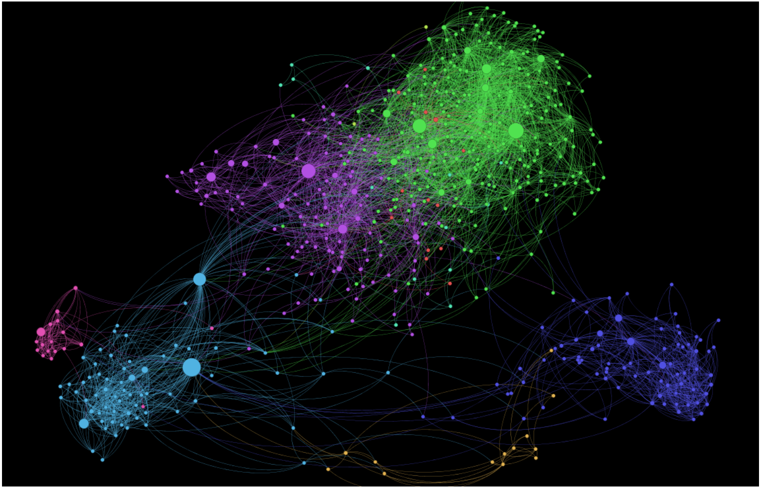
image from Jill Walker Rettberg, jilttxt.net

Examples: Social network - LinkedIn Map

LinkedIn contacts ego-centric network



Examples: Facebook communities structure



Examples: Political blogs

red-conservative blogs, blue -liberal, orange links from liberal to conservative, purple from conservative to liberal

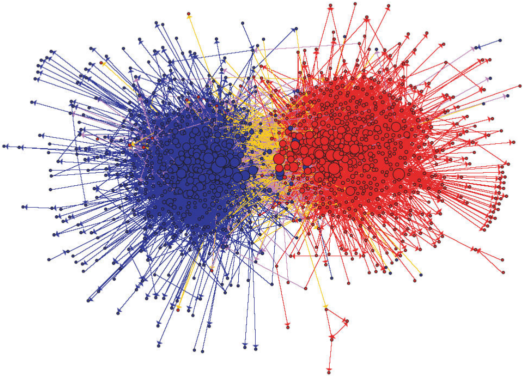


image from L. Adamic, N. Glance, 2005

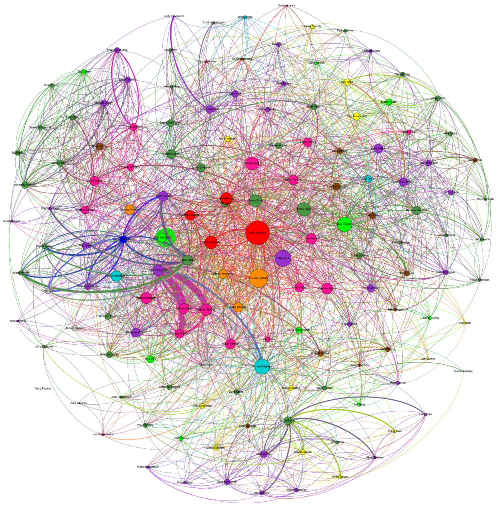
Examples: Twitter

"#usa" hashtag diffusion, retweets - blue, mentions - orange

image from K. McKelvey et.al., 2012

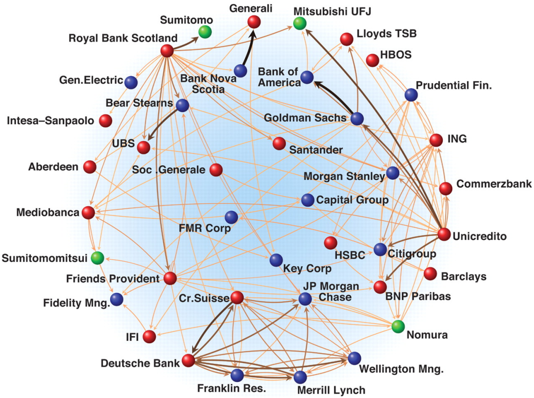
Examples: Communications

Enron emails



Examples: Finance

existing relations between financial institutions



F. Schweitzer, 2009

Examples: Transportation

Zurich public transportation map

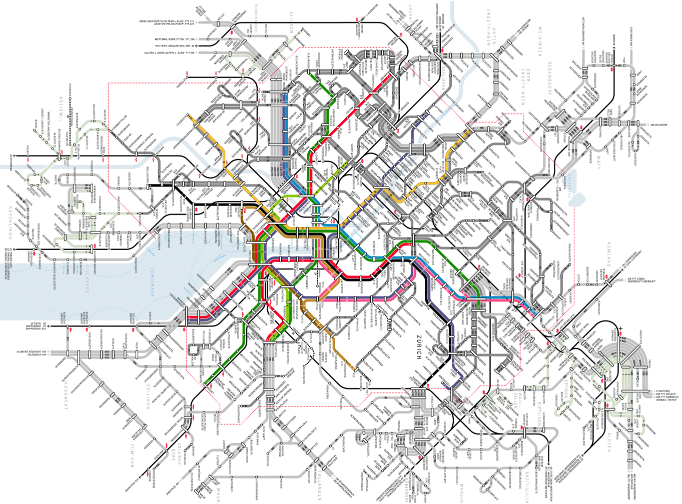


image from http://www.visualcomplexity.com

Examples: Transportation

London bike share

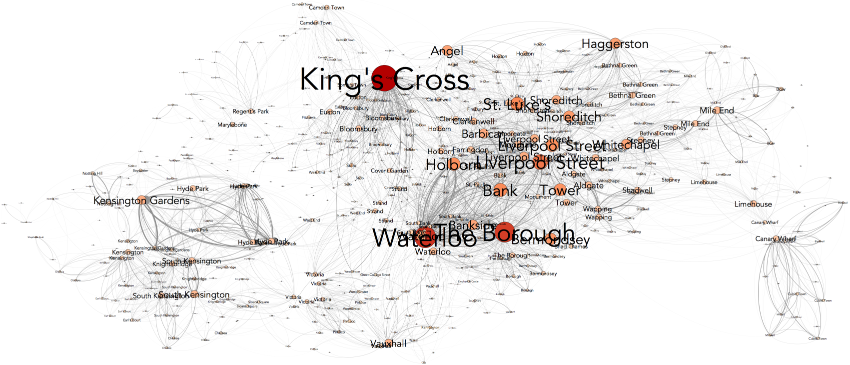
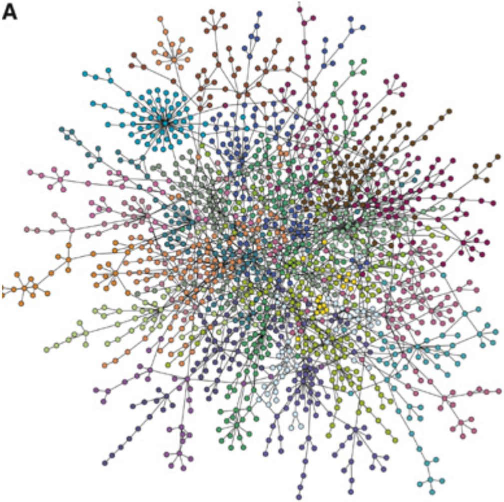


image from vartree.blogspot.com

Examples: Biology

Yeast protein interaction network



H. Jeong et.al., 2001

Examples: Facebook

Friendship graph 500 mln people

image by Paul Butler, 2010

Complex networks

1. Power law node degree distribution: "scale-free" networks
2. Small diameter and average path length: "small world" networks
3. Hight clustering coefficient: transitivity

Power law

Quantity of interest - frequency distribution of node degrees

$$f(k) \sim \frac{1}{k^{\gamma}}$$

- "A study of large sociogram", Anatol Rpoport and William Horrah, 1961
- "Networks of Scientific Papers", Derec J. de Solla Price, 1965
- "Diameter of the World-Wide Web", Reka Albert, Hawoong Jeong, Albert-Laszlo Barabasi, 1999
- "The Web as a graph: Measurements, models and methods", Jon Kleinberg et. al, 1999

Citation of scientific papers for 1961

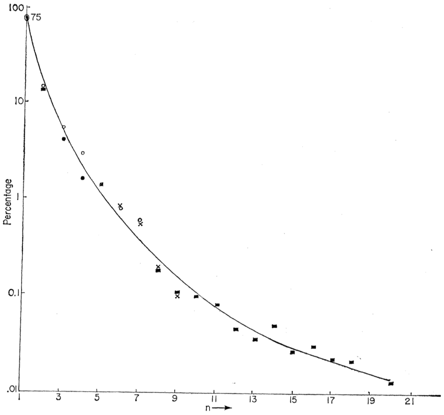


Fig. 2. Percentages (relative to total number of cited papers) of papers cited various numbers (n) of times, for a single year (1961). The data are from Garfield's 1961

from D.Price, 1965

Power law

Node degree distribution in random vs scale-free (power law) network:

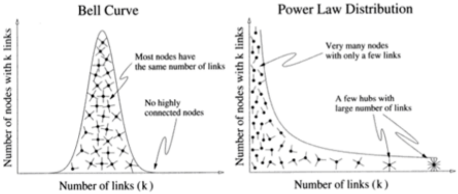


image from A.-L. Barabasi, 2002

Power law

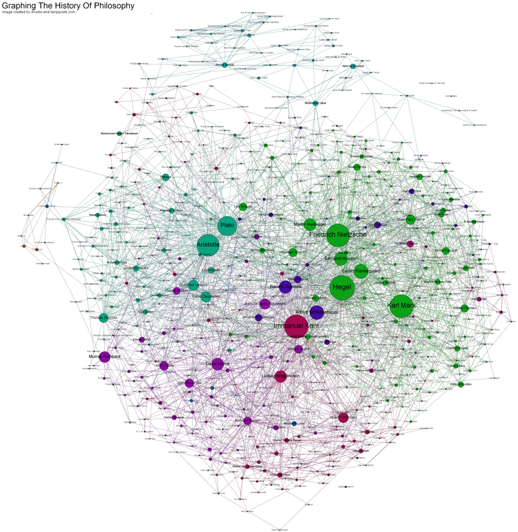


image from http://www.coppelia.io

Power law

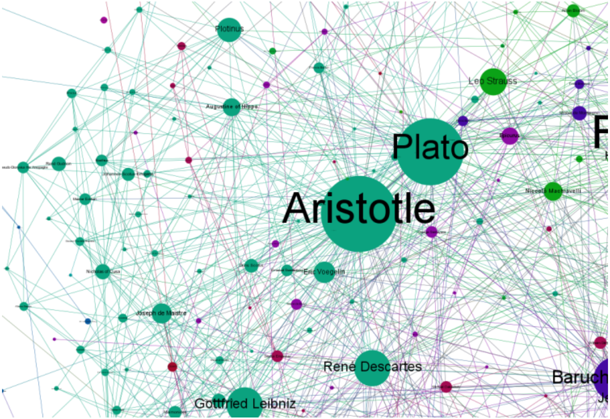


image from http://www.coppelia.io

Triads

The Strength of Weak Ties[1]

Mark S. Granovetter
Johns Hopkins University

Analysis of social networks is suggested as a tool for linking micro and macro levels of sociological theory. The procedure is illustrated by elaboration of the macro implications of one aspect of small-scale interaction: the strength of dyadic ties. It is argued that the degree of overlap of two individuals' friendship networks varies directly with the strength of their tie to one another. The impact of this principle on diffusion of influence and information, mobility opportunity, and community organization is explored. Stress is laid on the cohesive power of weak ties. Most network models deal, implicitly, with strong ties, thus confining their applicability to small, well-defined groups. Emphasis on weak ties lends itself to discussion of relations *between* groups and to analysis of segments of social structure not easily defined in terms of primary groups.

- "The Strength of Weak Ties", Mark Grannoveter, 1973
- "Spread of Information through a Population with Socio-Structural Bias. Assumption of Transitivity", Anatol Rapoport, 1953

Triadic closure

- strength of a tie
- high transitivity
- high clustering coefficient

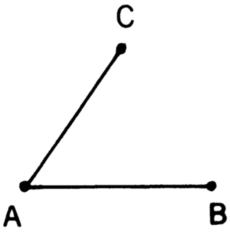


FIG. 1.—Forbidden triad

If A and B and B and C are strongly linked, the the tie between B and C is always present

Grannoveter, 1973

High clustering

Facebook friendship

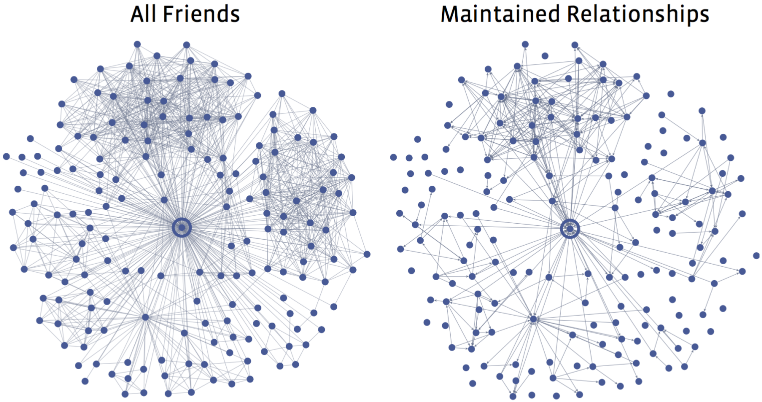
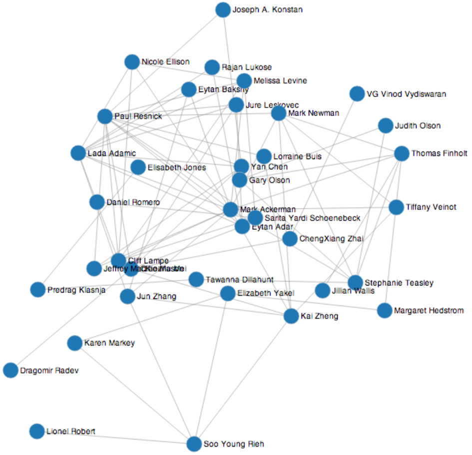


image from Cameron Marlow, Facebook

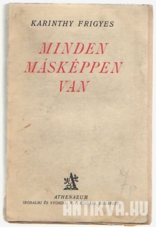
Co-author network

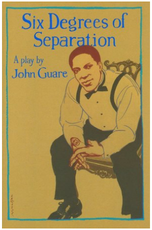


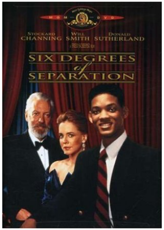
Six degrees of separation

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

- Frigyes Karinthy, short story "Lancszemek" ("Chain-Links"), 1929.
- 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", Jeffrey Travers, Stanley Milgram, 1969

Stanley Milgram's 1967 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.

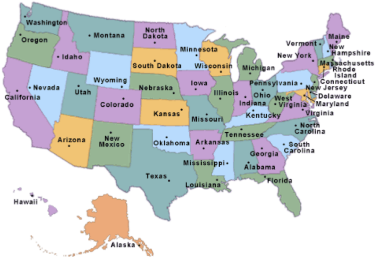
2. DETACH ONE POSTCARD. FILL IT OUT AND RETURN IT TO HARVARD UNIVERSITY. No stamp is needed. The postcard is very important. It allows us to keep track of the progress of the folder as it moves toward the target person.

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 1967 experiment

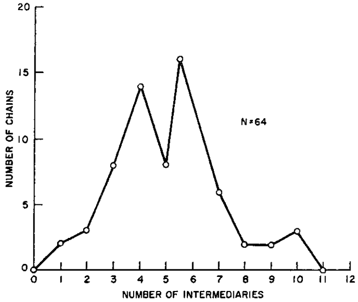
- Starting persons:
  - 296 volunteers, 217 sent
  - 196 in Nebraska
  - 100 in Boston
- Target person - Boston stockbroker
- Information given: target name, address, occupation, place of employment, college, hometown



J. Travers, S. Milgram, 1969

Stanley Milgram's 1967 experiment

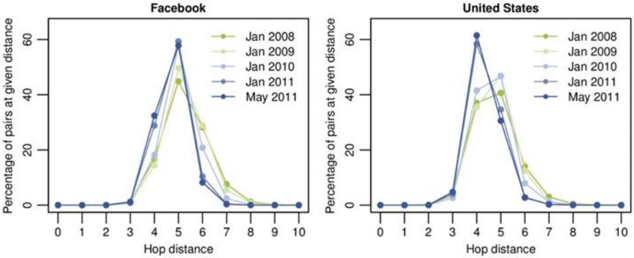
- Reached the target $N = 64(29\%)$
- Average chain length $\langle L \rangle = 5.2$
- Channels:
  - hometown $\langle L \rangle = 6.1$
  - business contacts $\langle L \rangle = 4.6$
  - from Boston $\langle L \rangle = 4.4$
  - from Nebraska $\langle L \rangle = 5.7$



J. Travers, S. Milgram, 1969

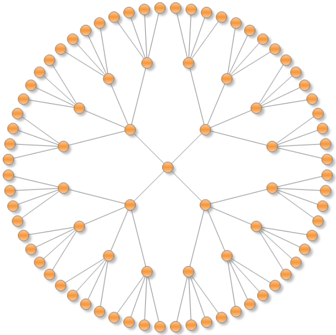
Small world

- Email graph:
  D. Watts (2001), 48,000 senders, $\langle L \rangle \approx 6$
- MSN Messenger graph:
  J. Lescovec et al (2007), 240mln users, $\langle L \rangle \approx 6.6$
- Facebook graph:
  L. Backstrom et al (2012), 721 mln users, $\langle L \rangle \approx 4.74$



figures from L.Backstrom, 2012

Simple model



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

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. 3338.

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