NATIONAL RESEARCH
UNIVERSITY

Introduction to Network Science

Network Science
Lecture 1

Leonid Zhukov

lzhukov@hse.ru
www.leonidzhukov.net/hse/2022/networks

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

January 10, 2022

Class details

- Instructors: Leonid Zhukov, Ilya Makarov, Dmitrii Kiselev
- Course duration: Modules 3-4
- Module 3: 10 lectures, 10 online labs
- Schedule:
  Lectures - Monday 18.10-19.30, ZOOM
  Labs - Monday, 19.40-21.00, ZOOM
- Website: www.leonidzhukov.net/hse/2022/networkscience
- Emails: lzhukov@hse.ru, iamakarov@hse.ru, dkiseljov@hse.ru
- Programming: Python, iPython notebooks (Anaconda)
- Python libraries: NetworkX
- Visualization: yEd, Gephi

Prerequisites

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

Network science

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

Conferences

- The Web conference (former WWW)
- WSDM, ICDM, KDD, ECML-PKDD
- International Conference on Social Network Analysis, INSNA
- Complex Networks
- ACM Conference on Online Social Networks
- Conference on Complex Systems

Course topics

- Statistical properties and modelling of the network
- Network structure and dynamics
- Processes on networks
- Predictions on networks (ML)
- Network embeddings (DL)
- Graph neural networks (DL)

Module 3 lectures

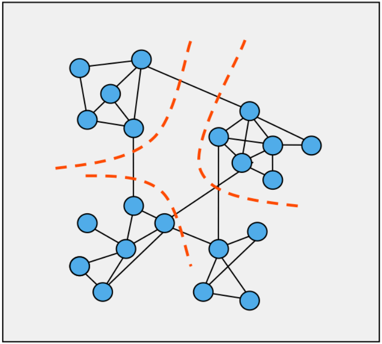
1. Introduction to network science
2. Power law and scale-free networks
3. Random graphs
4. Generative network models
5. Node centrality and ranking on networks
6. Network structure
7. Graph partitioning
8. Network communities
9. Mathematical models of epidemics
10. Epidemics on networks

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.

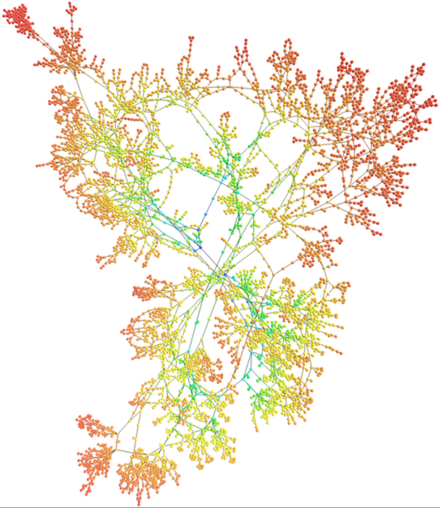
Terminology

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

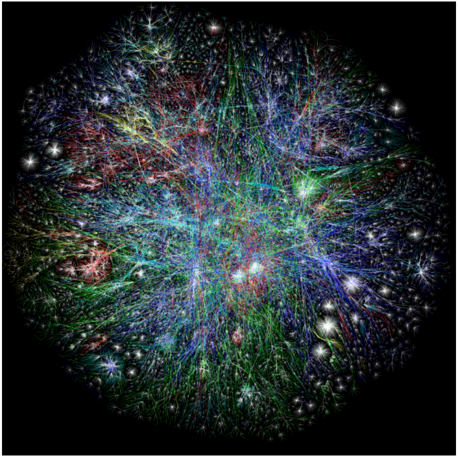


Networks

- they are everywhere
- universal abstract representation
- not regular, but not random
- non-trivial topology
- many universal properties
- complex systems

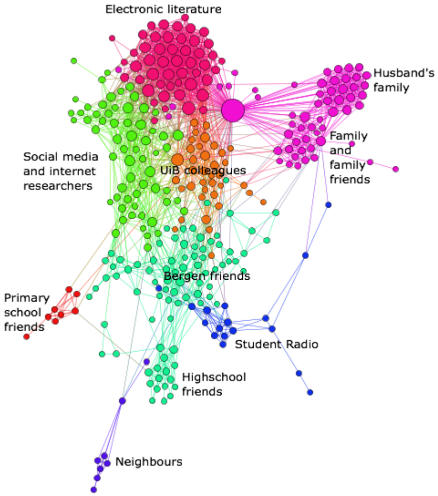


Internet traffic routing (BGP)



Barret Lyon, 2003

Facebook friendship



Political blogs

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

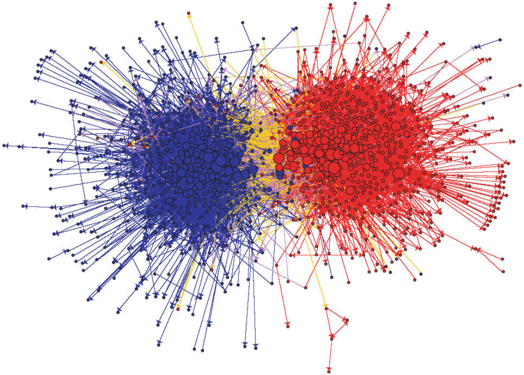


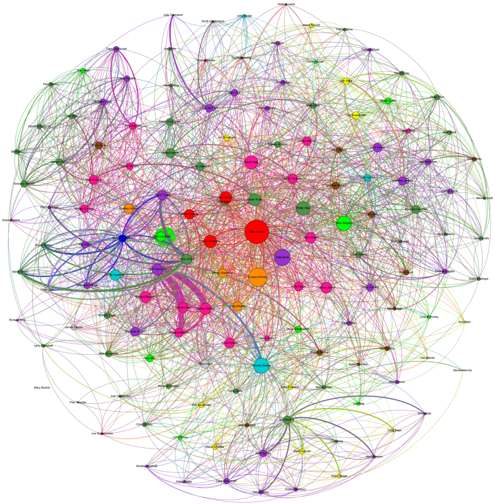
image from L. Adamic, N. Glance, 2005

Examples: Twitter

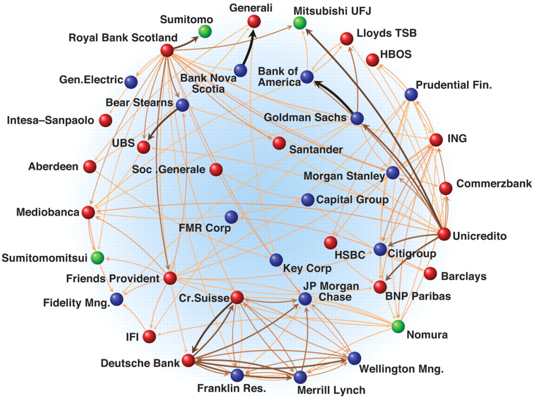
"#usa" hashtag diffusion, retweets - blue, mentions - orange

image from K. McKelvey et.al., 2012

Enron emails



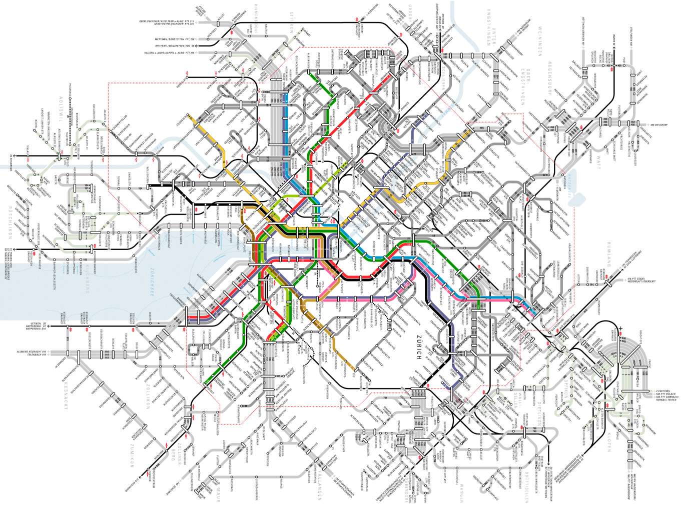
existing relations between financial institutions



F. Schweitzer, 2009

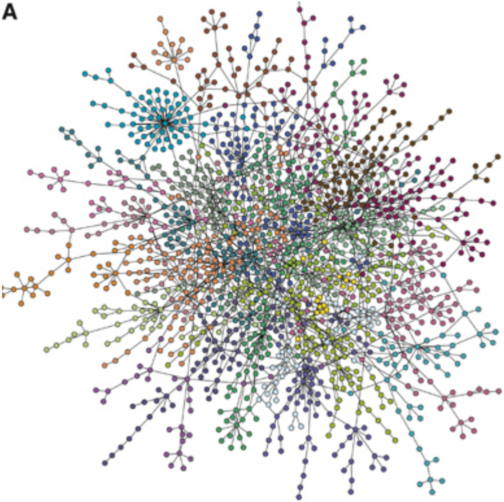
Examples: Transportation

Zurich public transportation map

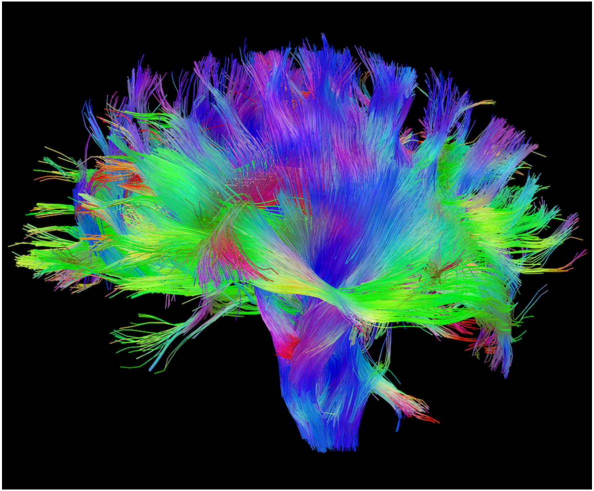


Biology

Yeast protein interaction network



Human Connectome

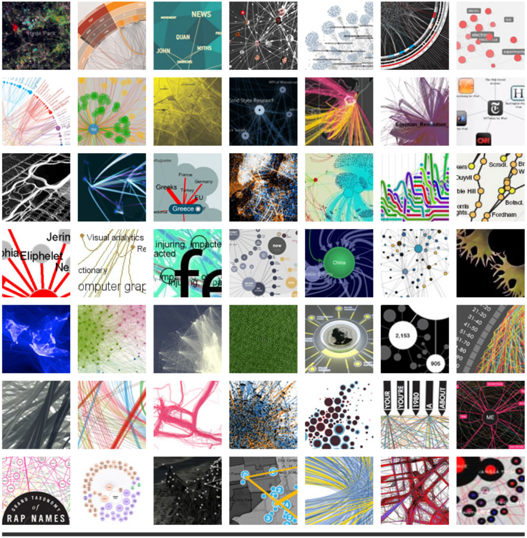


DT-MRI white matter fiber tractography
Human Connectome Project

Facebook

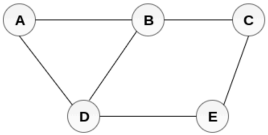
Friendship graph 500 mln people
image by Paul Butler, 2010

Visual complexity



Graphs

- A *network* is represented by a graph $G(V, E)$,
- A *graph* $G = (V, E)$ is an ordered pair of sets: a set of vertices V and a set edges E, where $n = |V|$, $m = |E|$
- An *edge* $e_{ij} = (v_i, v_j)$ is pair of vertices (ordered pair for directed graph)
- *Adjacency matrix* $A^{n \times n}$ is a matrix with nonzero element a_{ij} when there is an edge e_{ij}

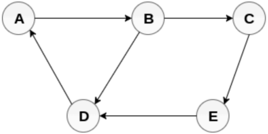


Undirected Graph

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

Adjacency Matrix

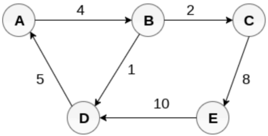
Graphs



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

Nodes degree

- Two nodes/vertices are *adjacent* if they share a common edge
- An edge and a node on that edge are called *incident*.
- The *degree* k_i of a node v_i is the total number of nodes adjacent to it (number of incident edges) , $k_i = |\mathcal{N}(v_i)|$
- Average node degree in the graph:

$$\langle k \rangle = \frac{1}{n} \sum_i k_i = \frac{2m}{n} = \frac{2|E|}{|V|}$$

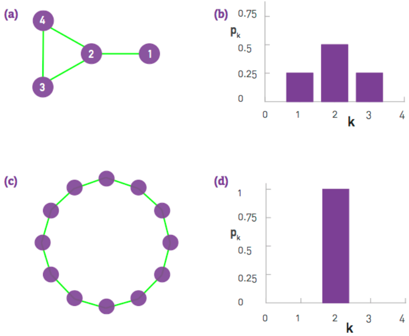
- In directed graphs total node degree $k_i = k_i^{in} + k_i^{out}$
  k_i^{in} - incoming degree, number of edges pointing to node *i*
  k_i^{out} - outgoing degree, number of edges/ pointing from node *i*
- In directed graphs average in and out degrees are equal:

$$\langle k^{in} \rangle = \frac{1}{n} \sum_i k_i^{in} = \langle k^{out} \rangle = \frac{1}{n} \sum_i k_i^{out} = \frac{m}{n} = \frac{|E|}{|V|}$$

Degree distribution

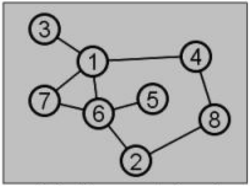
- k_i - node degree, $k_i = 1, 2, ...k_{\text{max}}$
- n_k - number of nodes with degree k, total nodes $n = \sum_k n_k$
- Degree distribution is a fraction of the nodes with degree k

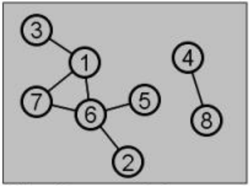
$$P(k_i = k) = P(k) = P_k = \frac{n_k}{\sum_k n_k} = \frac{n_k}{n}$$



Graph connectivity

- A *path* from v_i to v_j is a sequence of edges that joins two vertices. (It also ordered list of vertices such that that there is an edge to the next vertex on the list)
- A graph is *connected* if there a paths between any two vertices.
- *Connected component* is a maximal connected subgraph of G

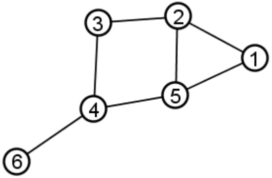




Paths and distances

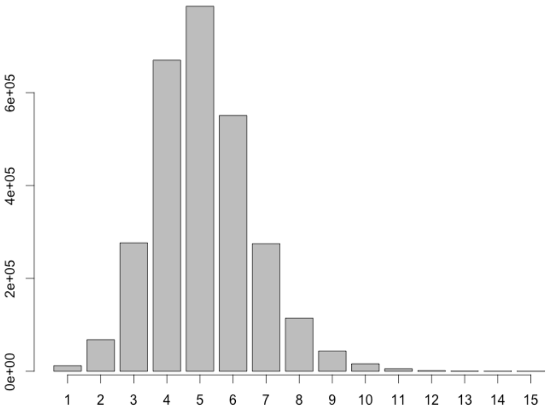
- The *distance* $d_G(v_i, v_j)$ between two vertices is the number of edges in the shortest path from v_i to v_j
- Graph *diameter* is the largest shortest path: $D = \max_{i,j} d_G(v_i, v_j)$
- Average path length (bounded from above by the diameter):

$$\langle L \rangle = \frac{1}{n(n-1)} \sum_{i \neq j} d_G(v_i, v_j)$$



Paths and distances

"Yeast" graph, $n = 2617$, $m = 11855$

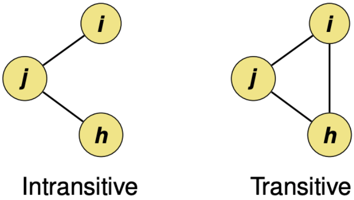


Diameter $D = 15$, average path length $\langle L \rangle = 5.1$

Graph transititvity

- *Transitivity of a graph* (global clustering coefficient):

$$T = \frac{3 \times \text{number of triangles (triads)}}{\text{number of connected triplets of vertices}}$$

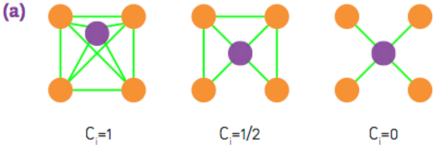


Clustering coefficient

How neighbours of a given node connected to each other

- *Local clustering coefficient* (per vertex):

$$C_i = \frac{\text{number of links in } \mathcal{N}_i}{k_i(k_i - 1)/2}$$



- Average clustering coefficient:

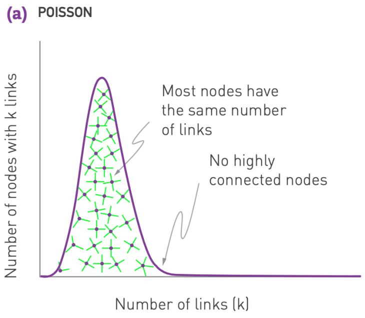
$$\bar{C} = \frac{1}{n}\sum_{i=1}^{n} C_i$$

Complex networks

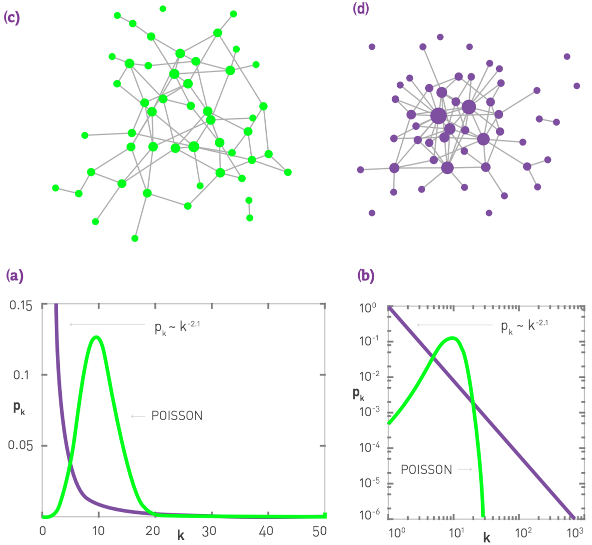
Networks:

1. Scale-free - power law node degree distribution
2. Small-world - small diameter and average path length
3. Transitive - high clustering coefficient

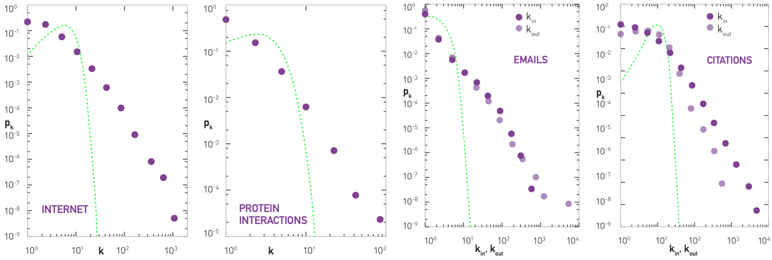
Scale-free networks



Node degree distributions



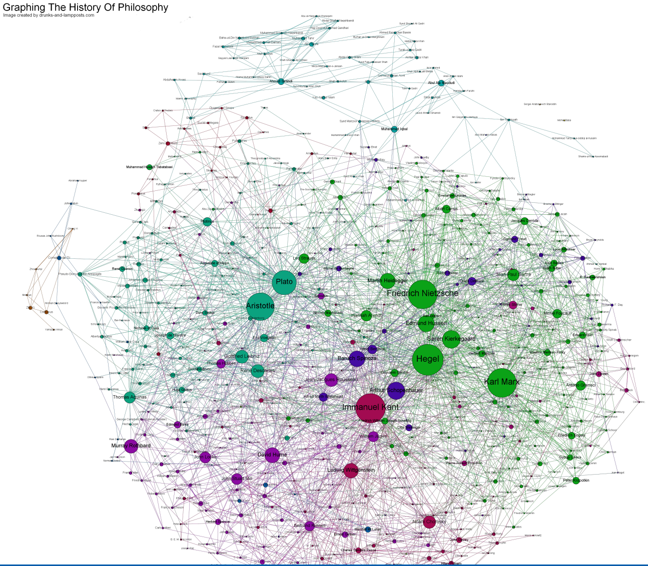
Scale-free networks



$$\log p(k) = c - \gamma \, log \, k$$

$$p(k) = Ck^{-\gamma} = \frac{C}{k^{\gamma}}$$

Power law network



Power law network

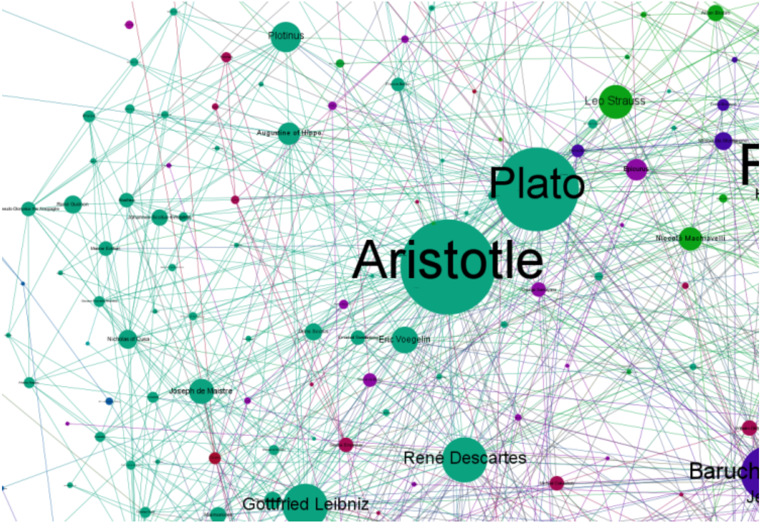


image from http://www.coppelia.io

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

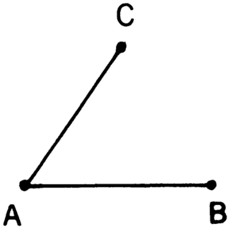


FIG. 1.—Forbidden triad

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

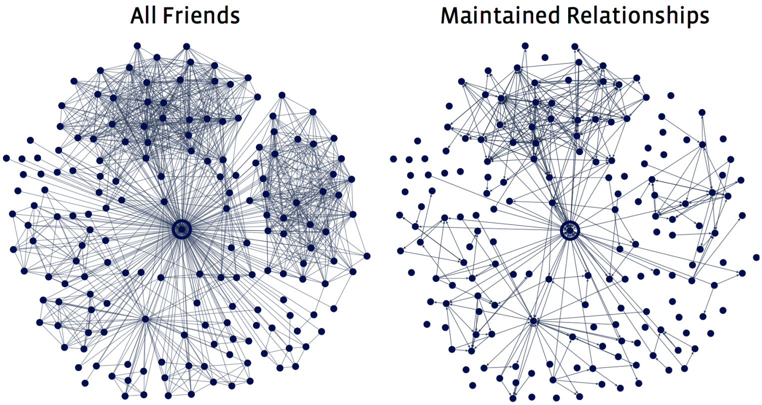
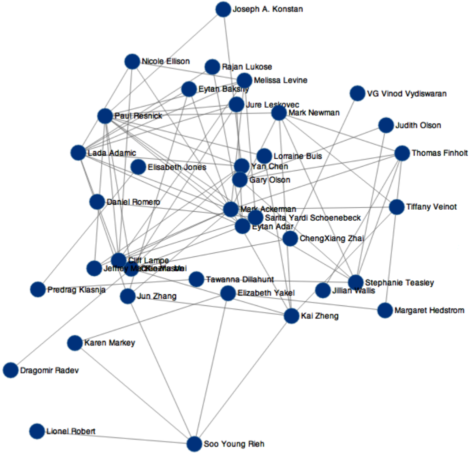


image from Cameron Marlow, Facebook

High clustering

Co-author network



Small world: six degrees of separation

© 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

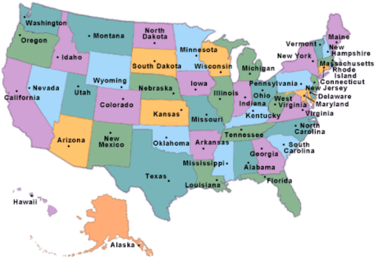
Stanley Milgram's 1967 experiment

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

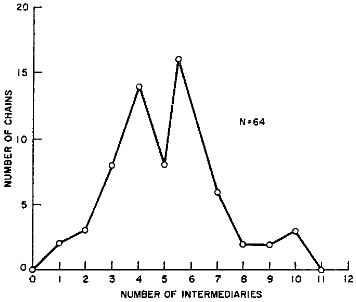


J. Travers, S. Milgram, 1969

Stanley Milgram's 1967 experiment

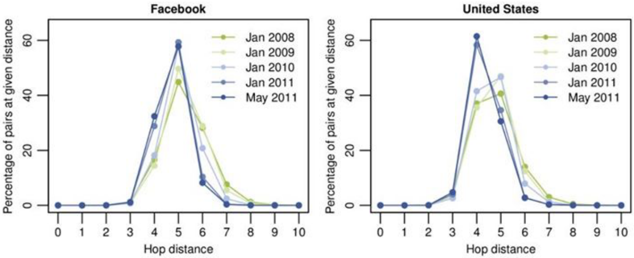
- Reached the target $N = 64 (29\%)$
- Average chain length $\langle L \rangle = 5.2$
- Channels:
  - hometown $\langle L \rangle = 6.1$
  - business contacts $\langle L \rangle = 4.6$
  - from Boston $\langle L \rangle = 4.4$
  - from Nebraska $\langle L \rangle = 5.7$

J. Travers, S. Milgram, 1969



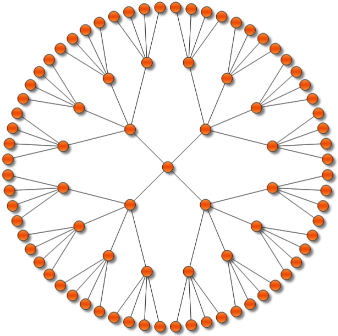
Small world

- 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, $d \approx 5.8$.

Network Data

- The Colorado Index of Complex Networks (ICON)
  http://icon.colorado.edu
- Stanford Large Network Dataset Collection
  http://snap.stanford.edu/data/index.html
- UCI Network Data Repository
  http://networkdata.ics.uci.edu

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