NATIONAL RESEARCH
UNIVERSITY

Introduction to Network Science

Social Network Analysis. MAGoLEGO course.
Lecture 1

Leonid Zhukov

lzhukov@hse.ru
www.leonidzhukov.net/hse/2019/sna

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

Class Technicalities

- Instructors: Leonid Zhukov, Ilya Makarov
- MAGoLEGO course: 1 module
- 10 lectures, 10 labs, 4 homeworks
- Final exam
- Schedule: Fridays, 18.10-21.00 (lecture + lab)
- Website: http://www.leonidzhukov.net/hse/2019/sna
- Emails: lzhukov@hse.ru, iamakarov@hse.ru

Helpful background

Theory:

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

Programming experience:

- R, RStudio
- R libraries: igraph
- Visualization: Gephi

Textbooks

- "Network Science", Albert-Laszlo Barabasi, Cambridge University Press, 2016.
- "Networks, Crowds, and Markets: Reasoning About a Highly Connected World". David Easley and John Kleinberg, Cambridge University Press 2010.
- "Statistical Analysis of Network Data with R", Eric Kolaczyk, Gabor Csardi, Springer, 2014.
- "Social Network Analysis. Methods and Applications". Stanley Wasserman and Katherine Faust, Cambridge University Press, 1994

Lectures

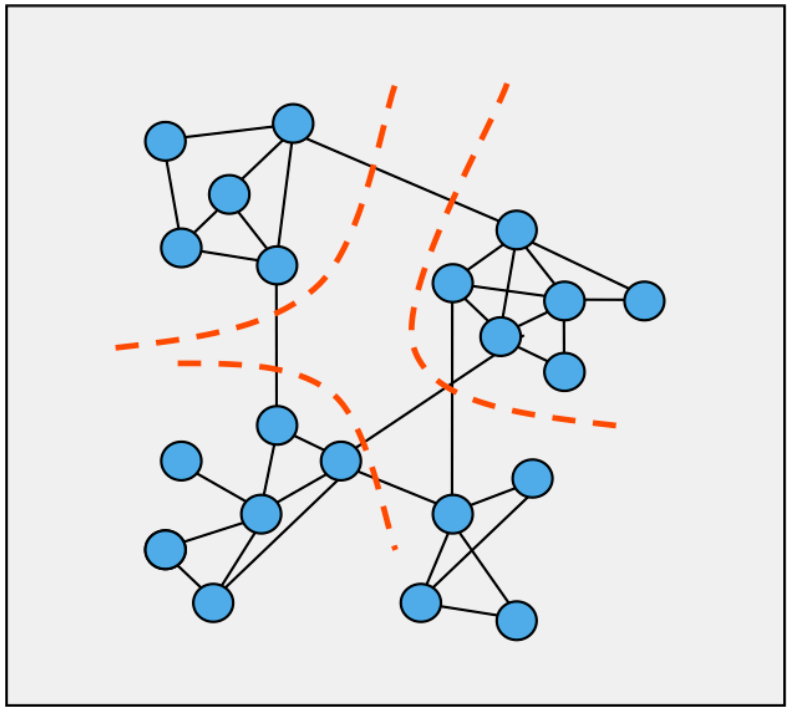
1. Introduction to network science
2. Descriptive network analysis
3. Mathematical models of networks
4. Node centrality and ranking on networks
5. Network communities
6. Network structure and visualization
7. Epidemics and information spreading in networks
8. Diffusion of innovation
9. Strategic network formation
10. Spatial models of segregation

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



Examples: Social network

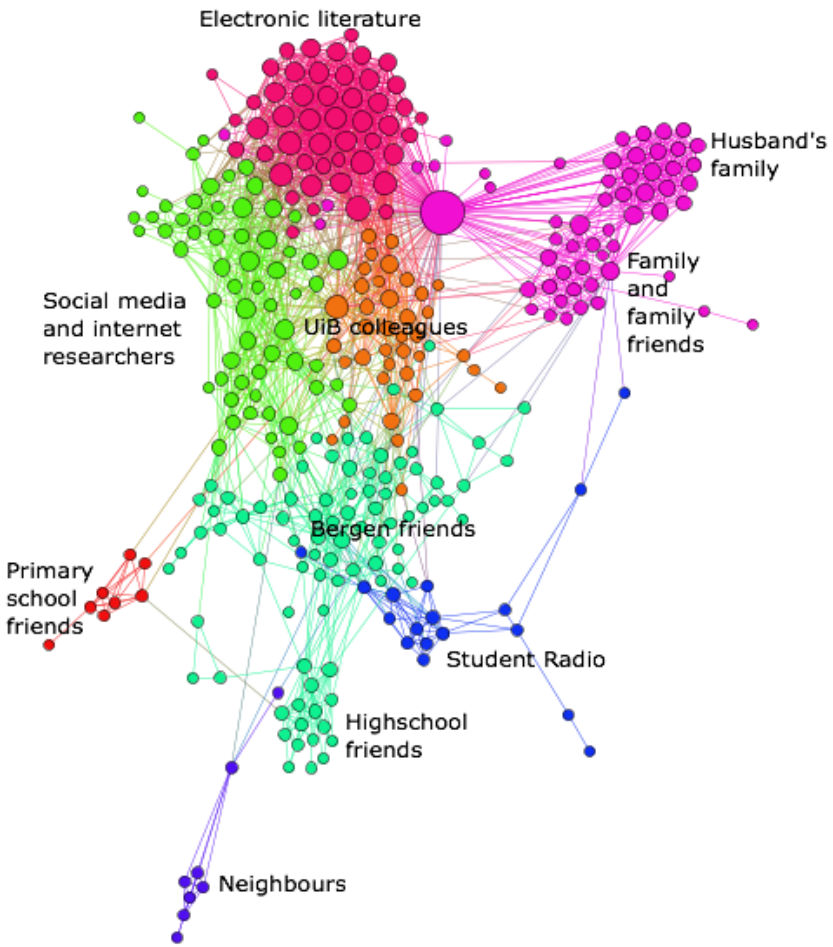
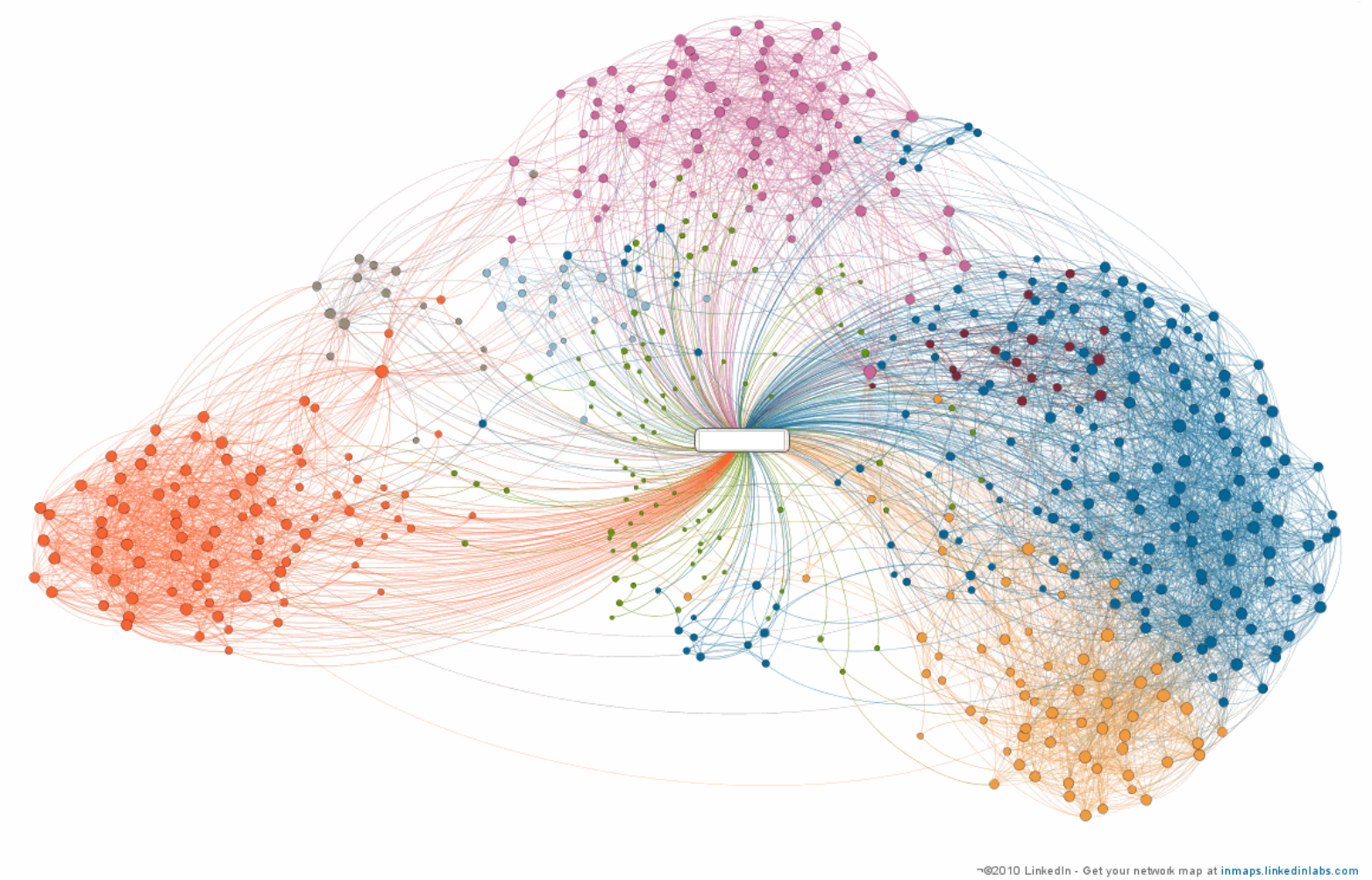


image from Jill Walker Rettberg, jilttxt.net

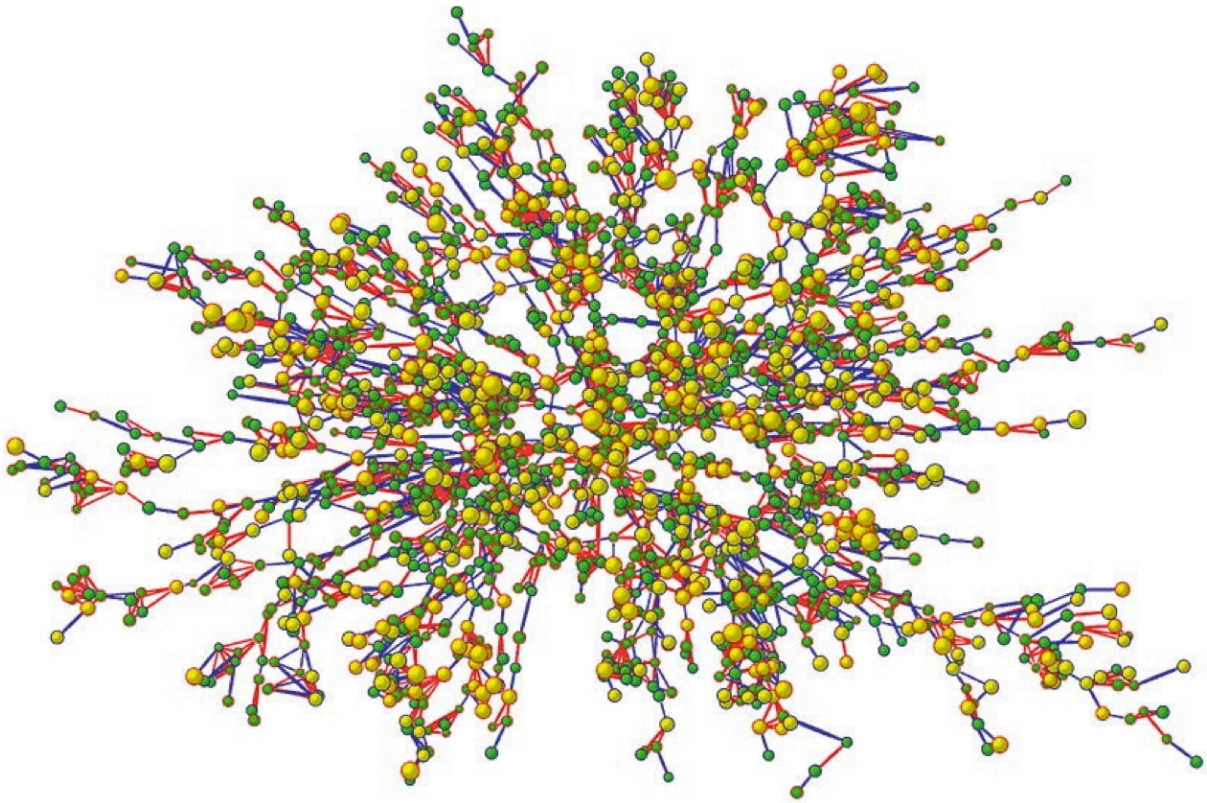
Examples: LinkedIn Map

LinkedIn contacts ego-centric network



Examples: Social network

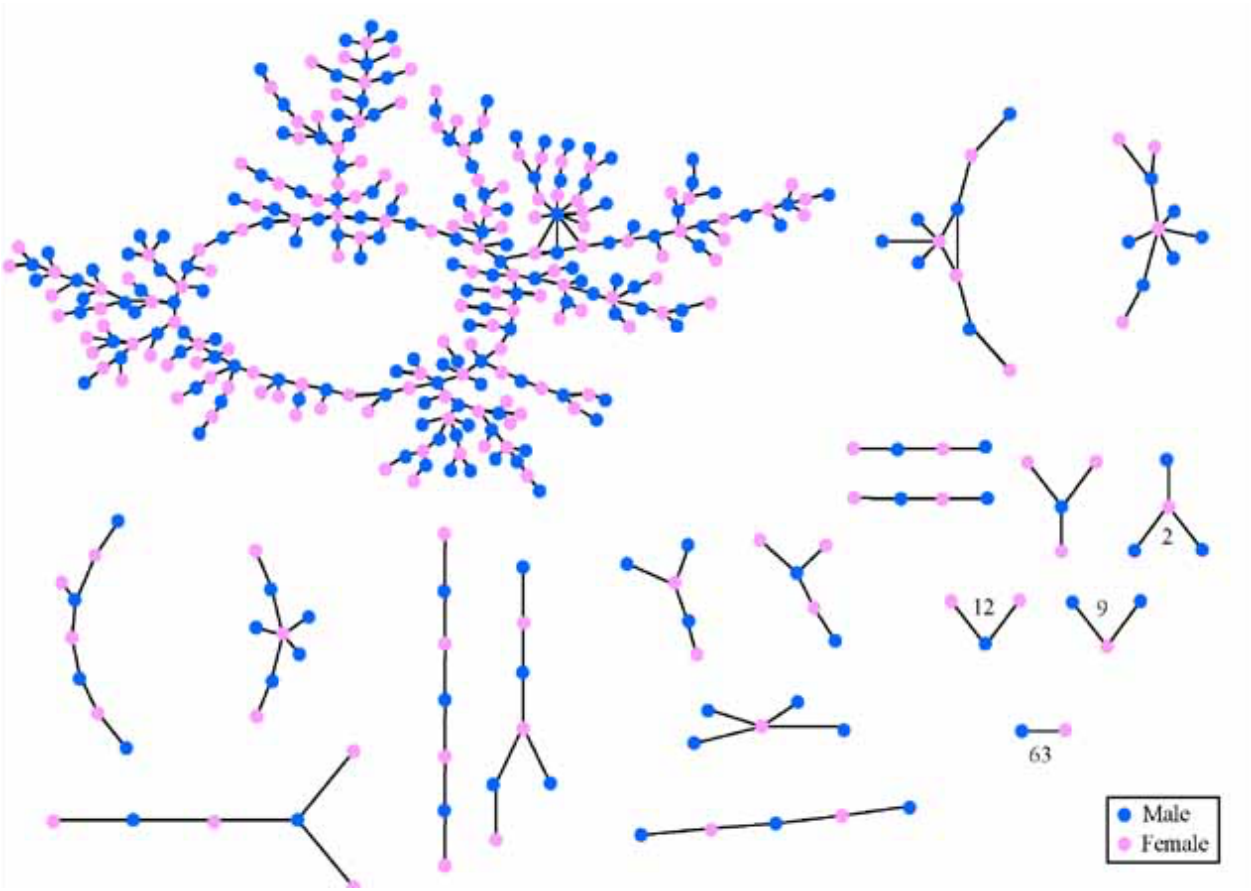
"The Spread of Obesity in a Large Social Network over 32 Years"



N. Christakis , J. Fowler 2007

Examples: high school dating network

"Chains of Affection: The Structure of Adolescent Romantic and Sexual Networks"



P. Bearman, J. Moody, K. Stovel, 2004

Examples: Political blogs

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

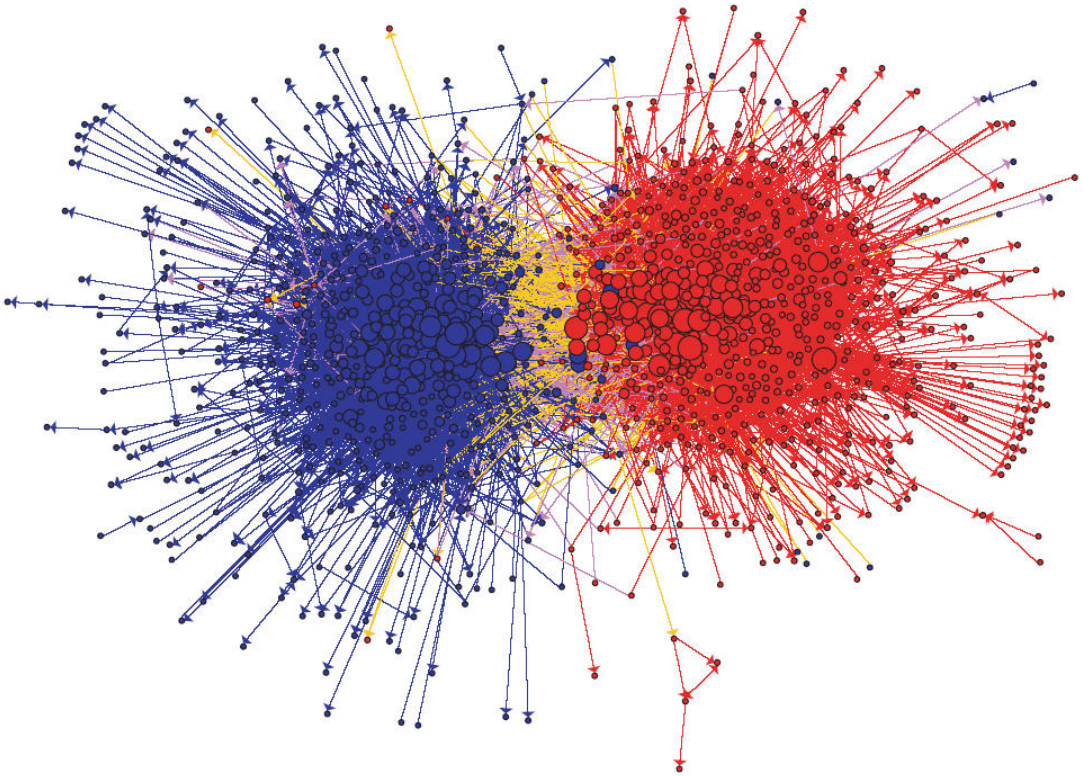


image from L. Adamic, N. Glance, 2005

Examples: Twitter

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

image from K. McKelvey et.al., 2012

Examples: Twitter

News about Bin Laden

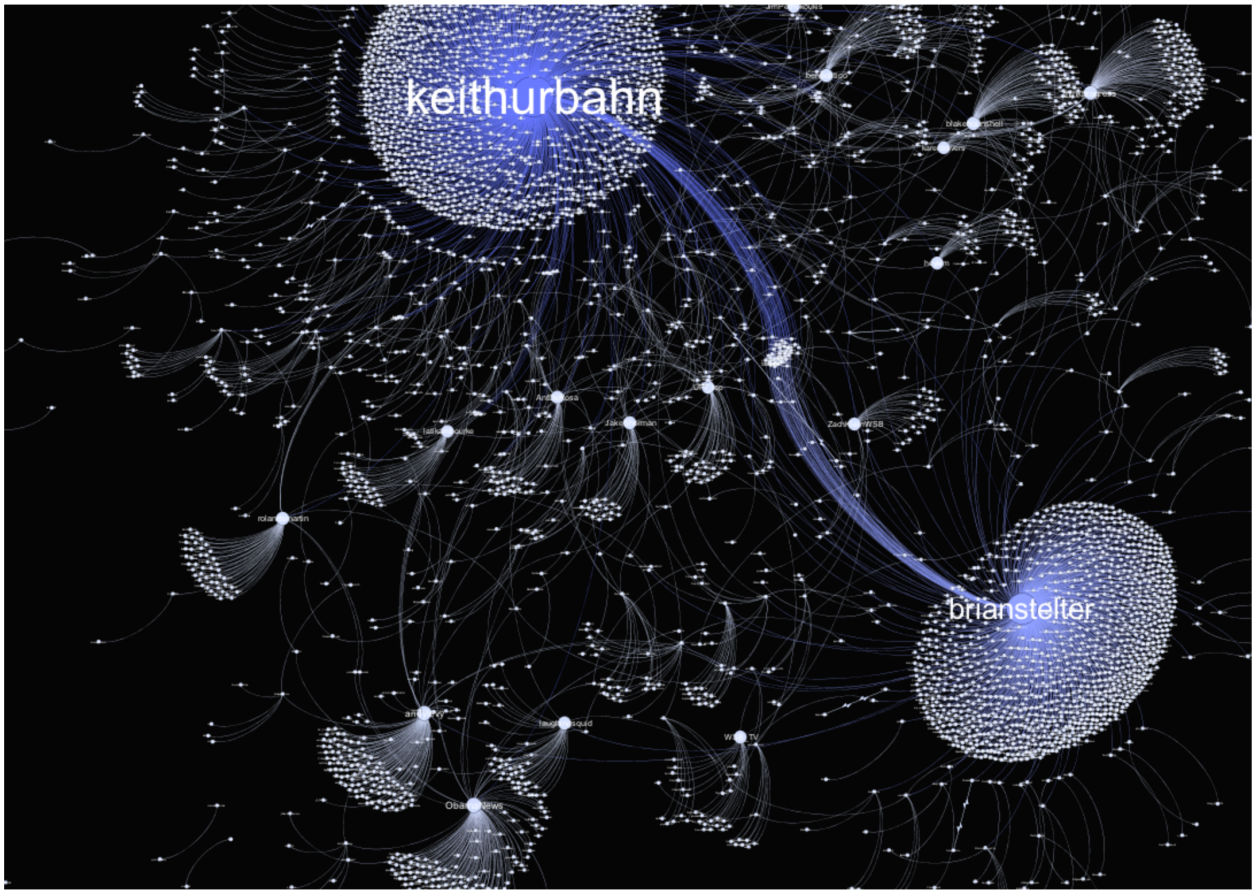
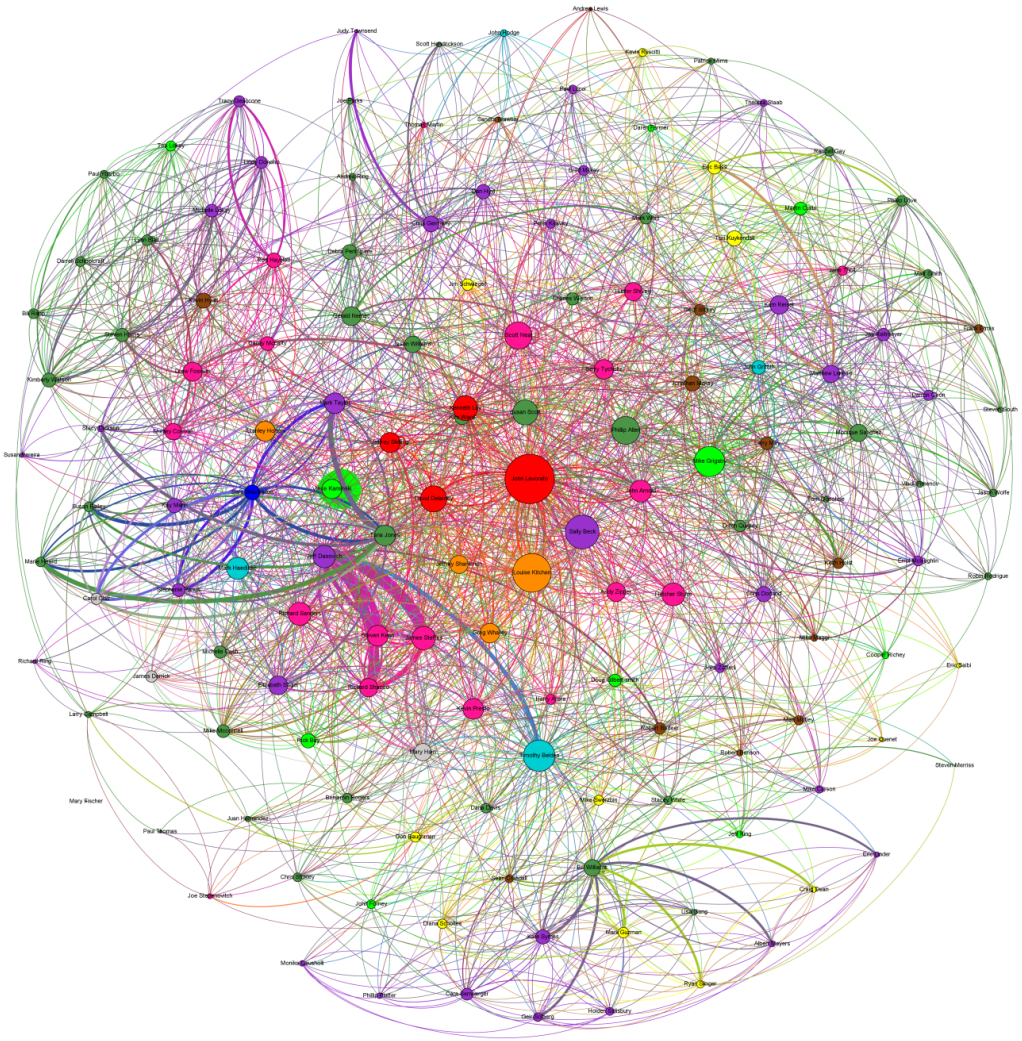


image from SocialFlow

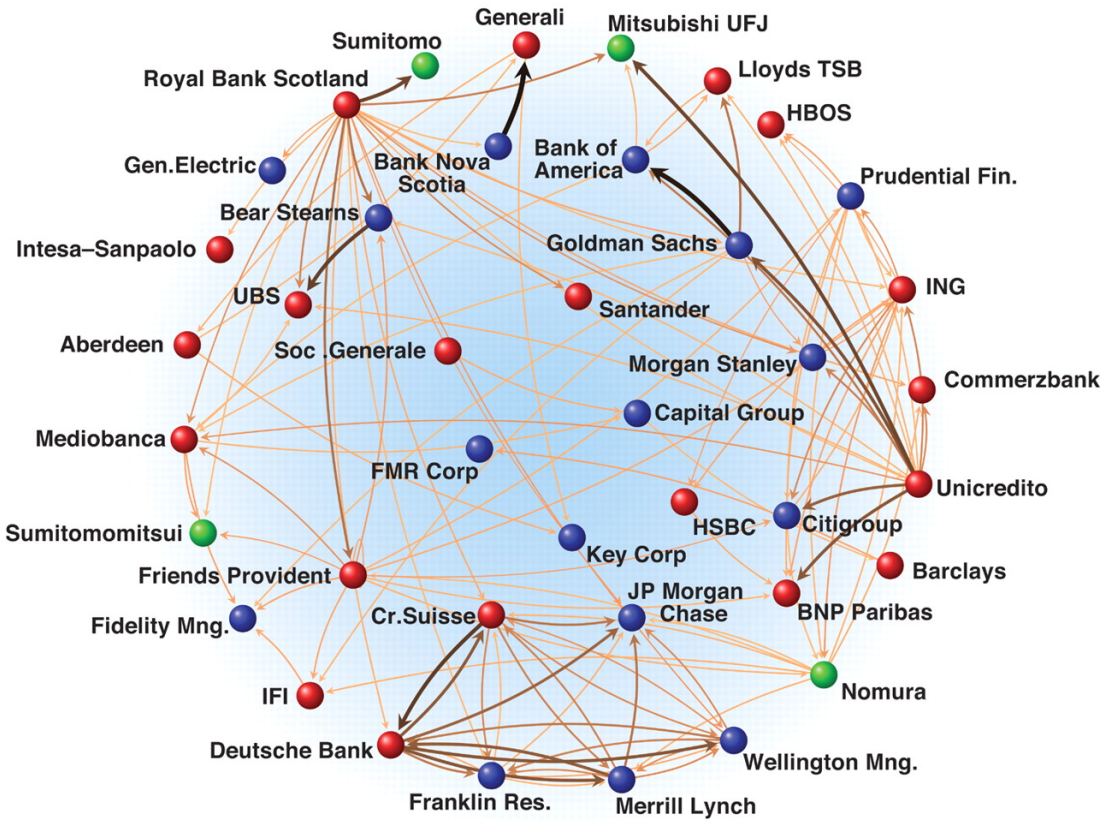
Examples: Emails

Enron emails



Examples: Finance

existing relations between financial institutions



F. Schweitzer, 2009

Examples: Transportation

Zurich public transportation map

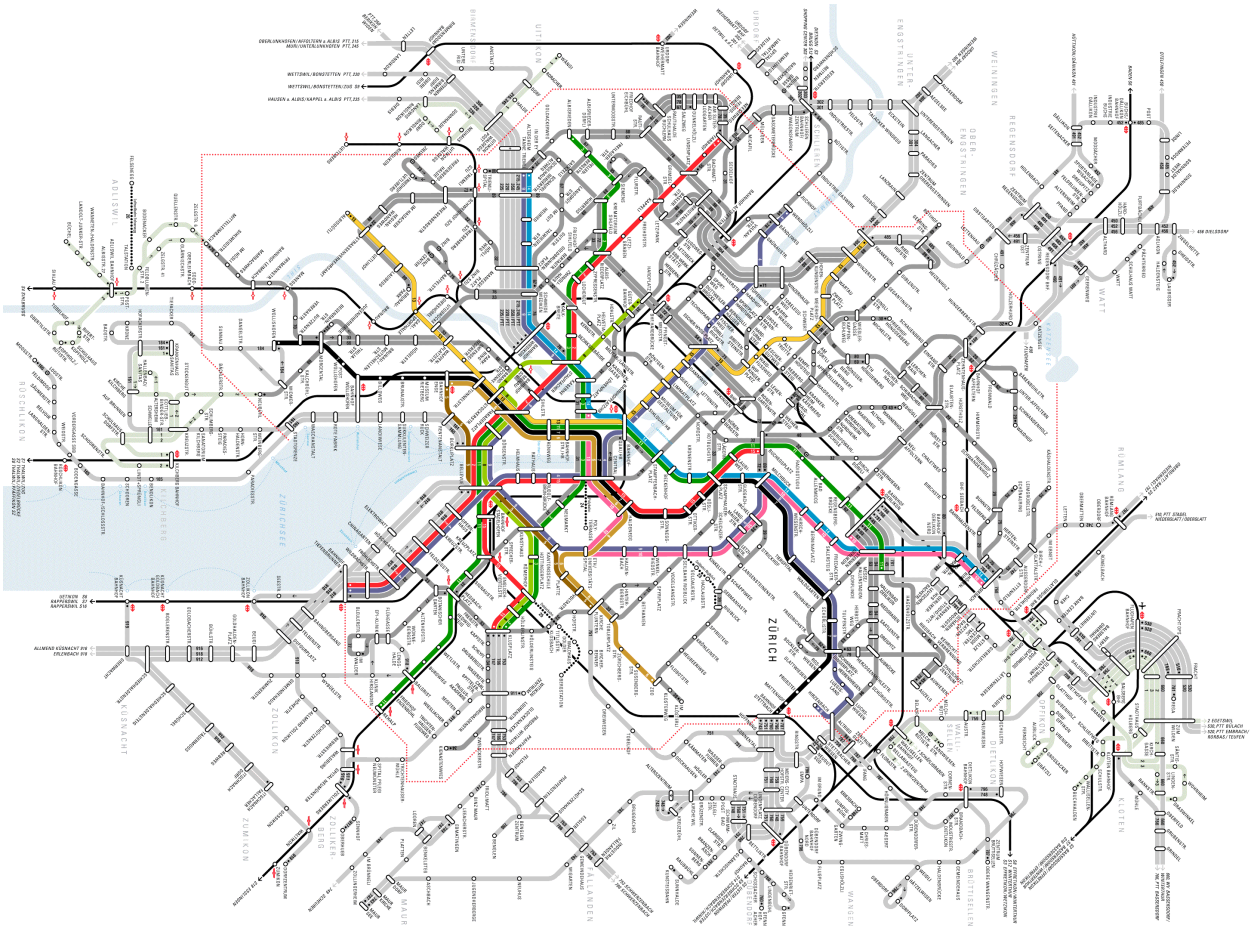


image from http://www.visualcomplexity.com

Examples: Transportation

London bike share

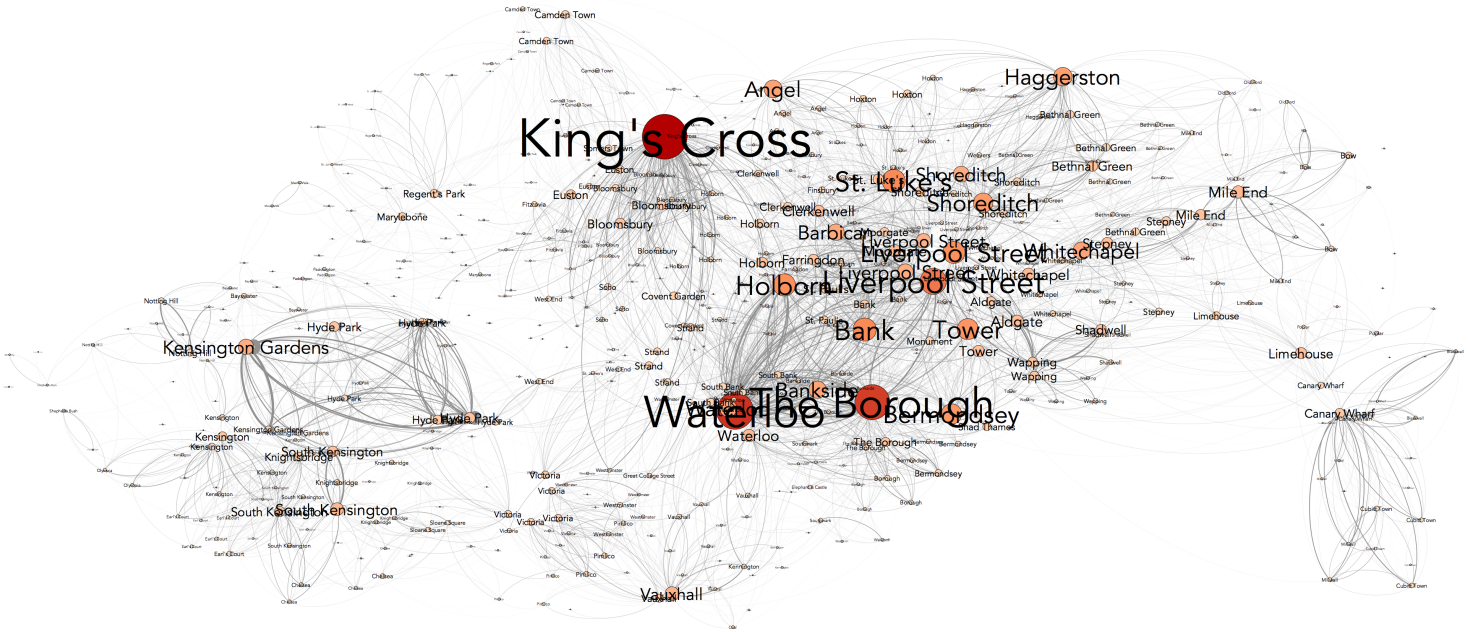
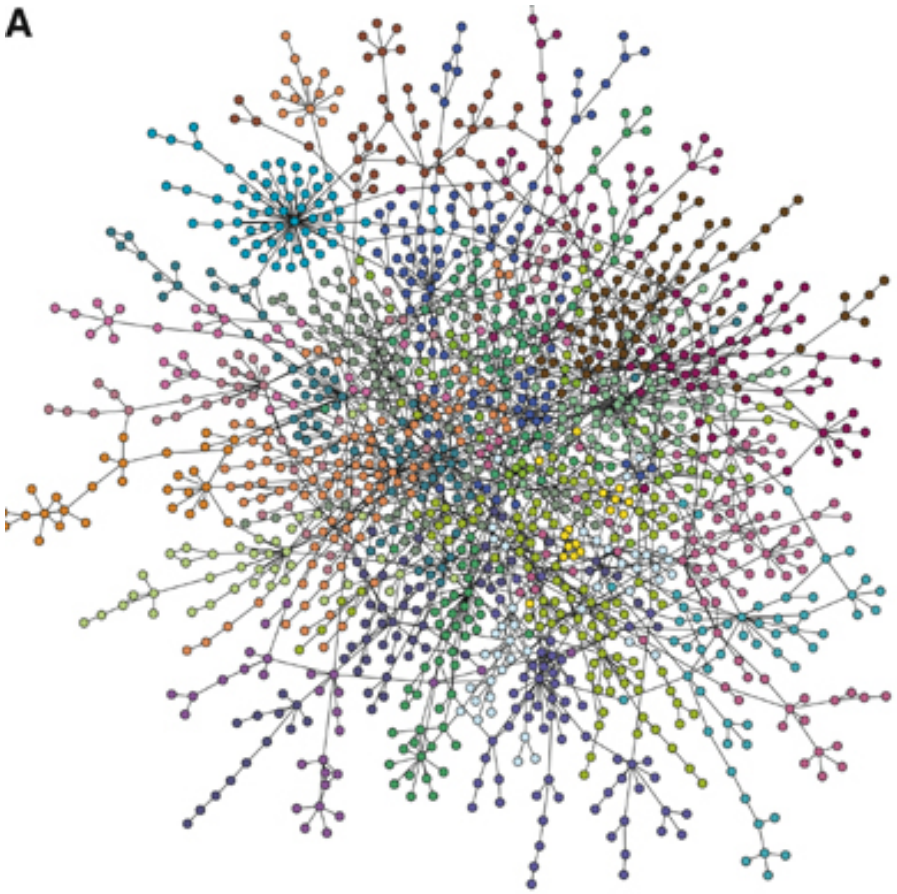


image from vartree.blogspot.com

Examples: Biology

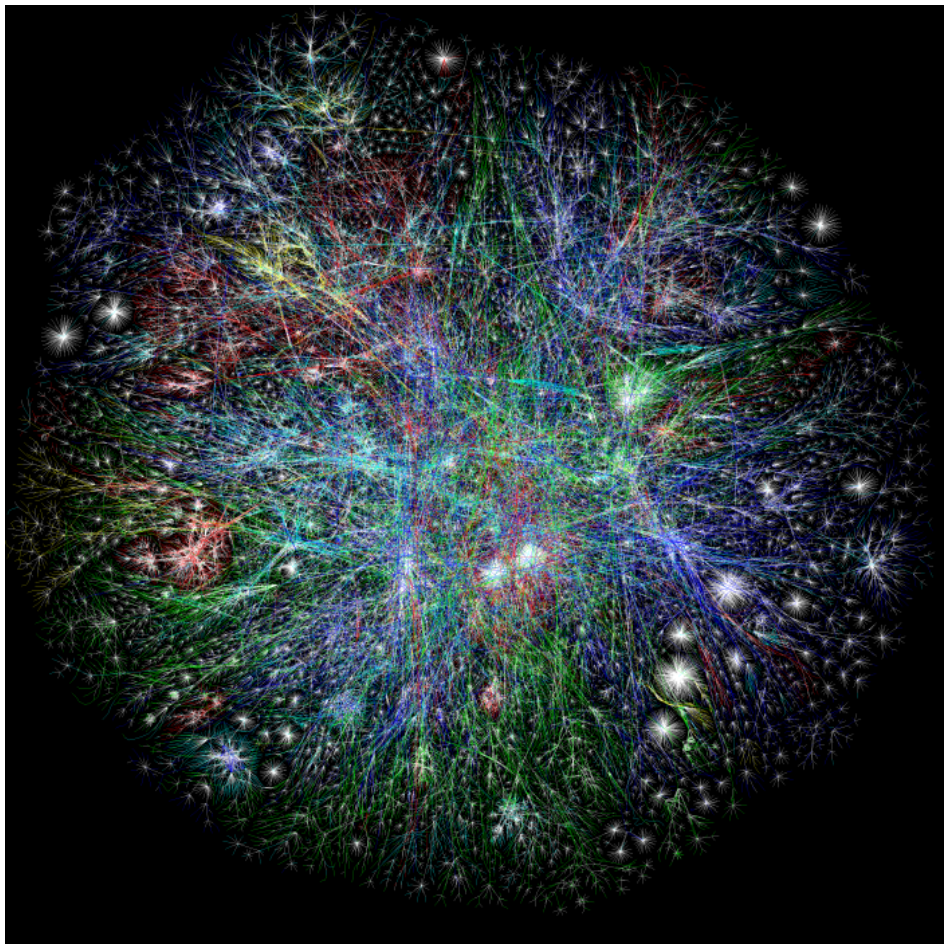
Yeast protein interaction network



H. Jeong et.al., 2001

Examples: Internet

Internet traffic routing (BGP)



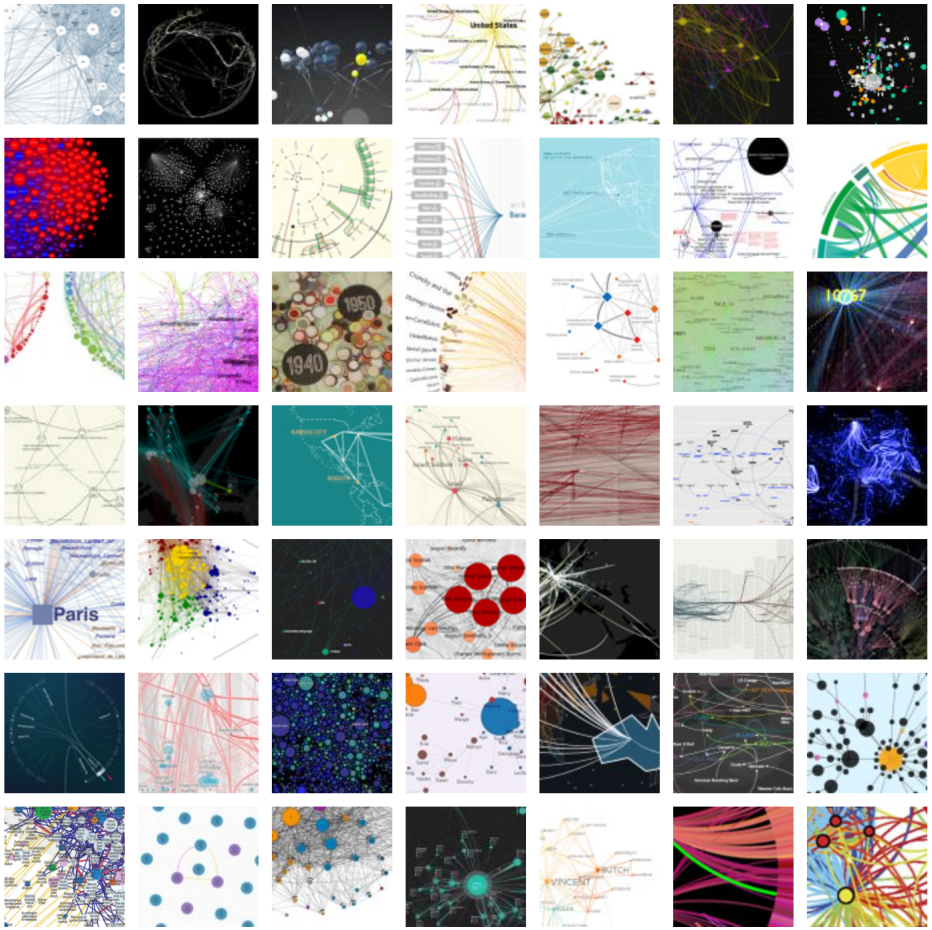
Barret Lyon, 2003

Examples: Facebook

Friendship graph 500 mln people

image by Paul Butler, 2010

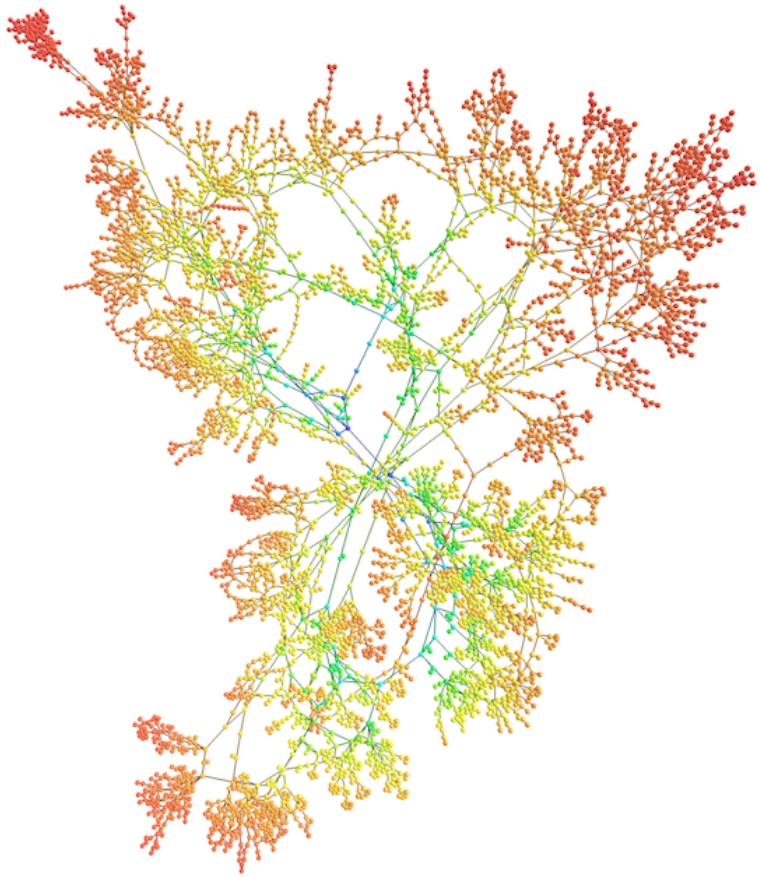
Visual complexity



www.visualcomplexity.com

Complex networks

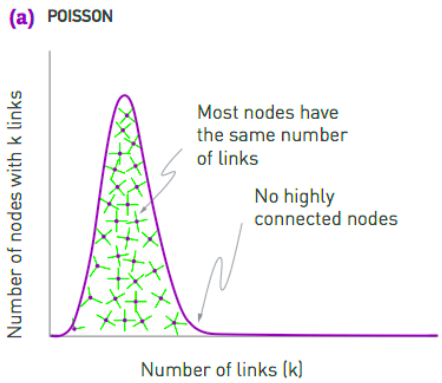
- not regular, but not random
- non-trivial topology
- universal properties
- everywhere
- complex systems

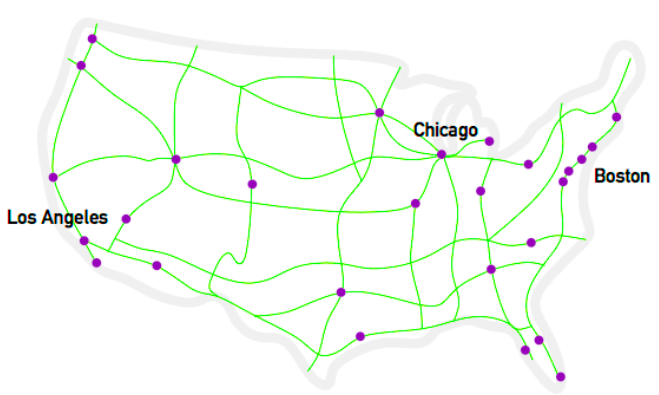


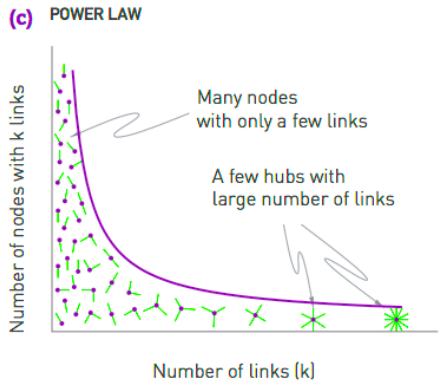
Complex networks

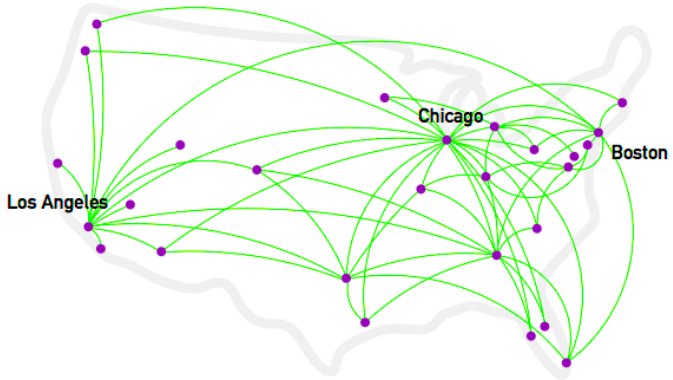
1. Power law node degree distribution: ”scale-free” networks
2. Small diameter and average path length: ”small world” networks
3. Hight clustering coefficient: transitivity

Power law









Frequency distribution of node degrees $f(k) \sim \frac{1}{k^{\gamma}}$

image from A.-L. Barabasi, 2016

Power law

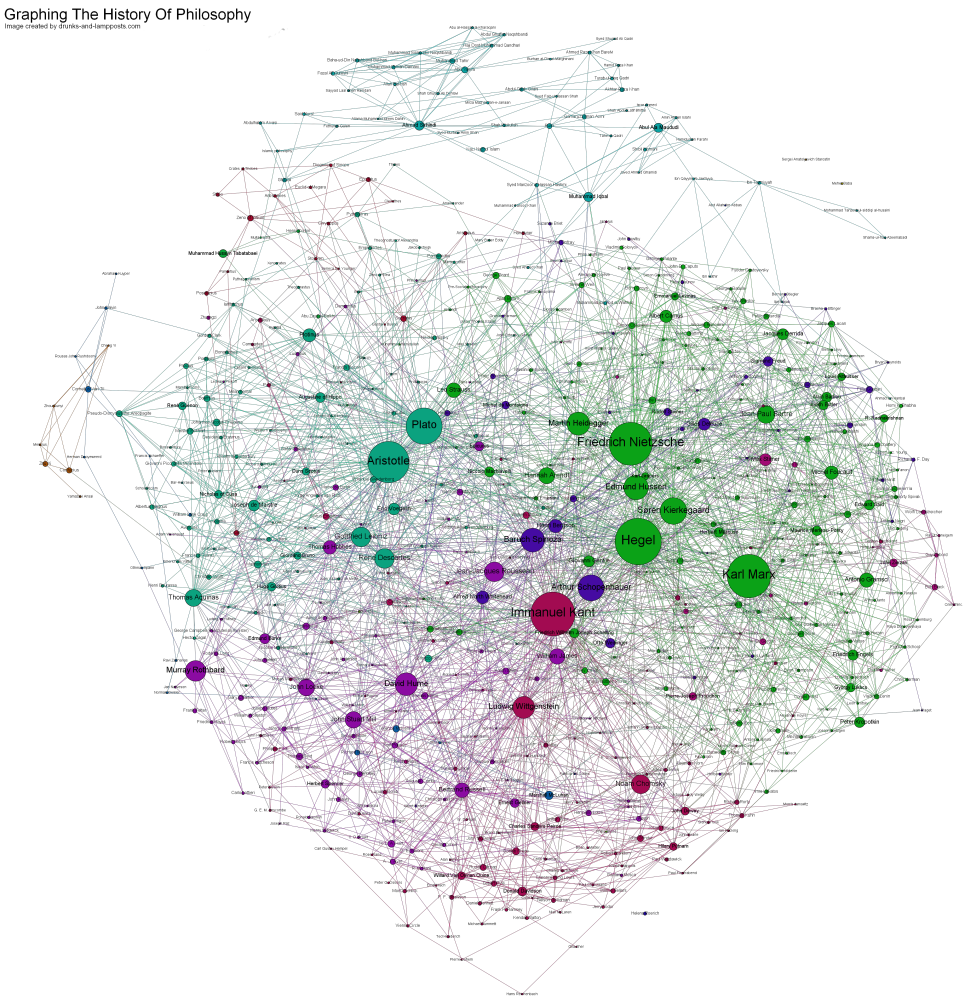


image from http://www.coppelia.io

Power law

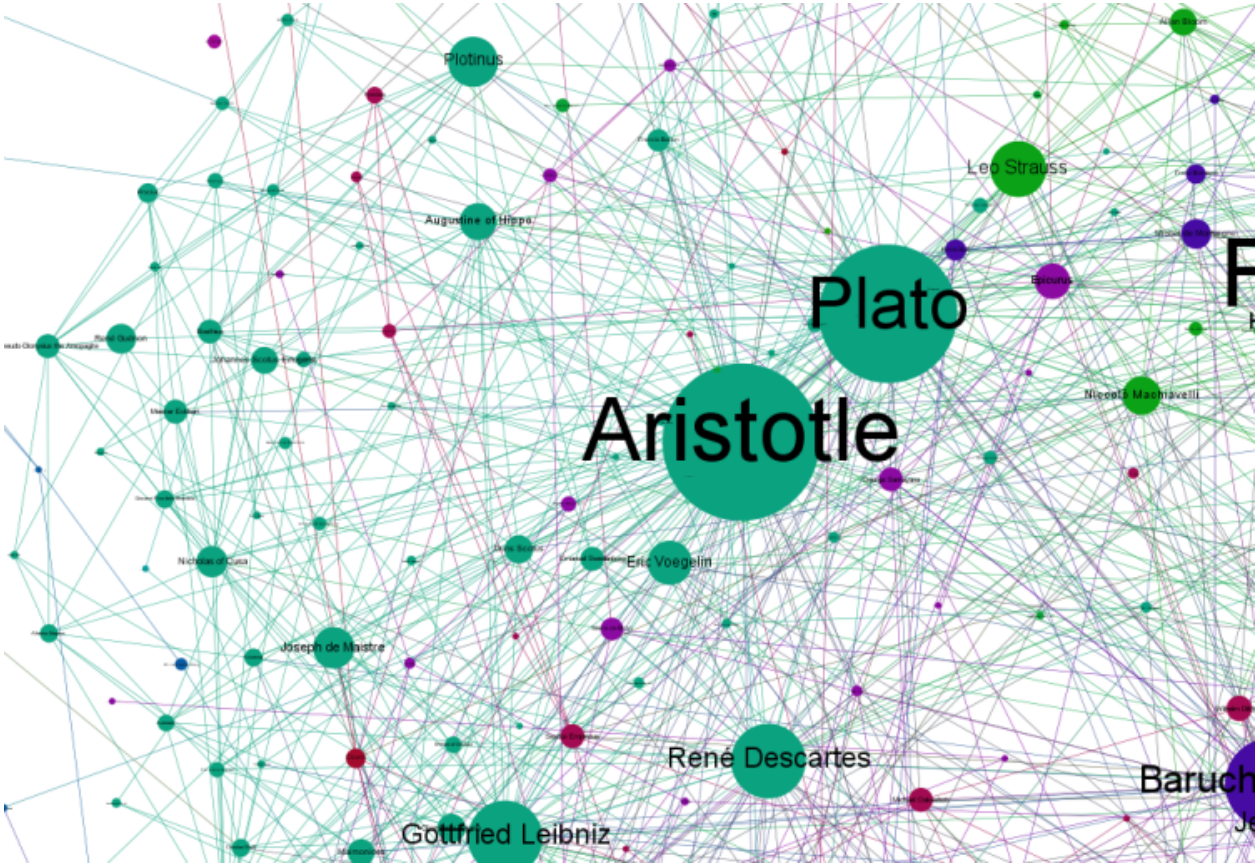


image from http://www.coppelia.io

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

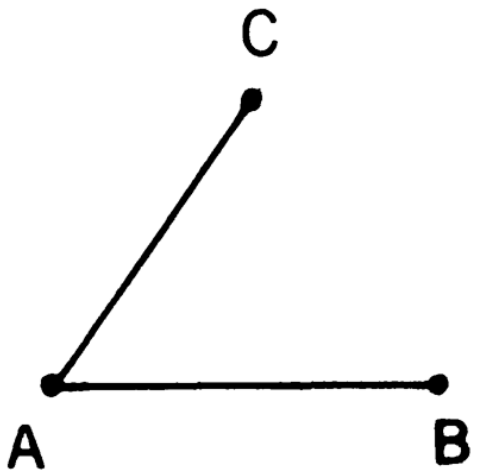


FIG. 1.—Forbidden triad

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

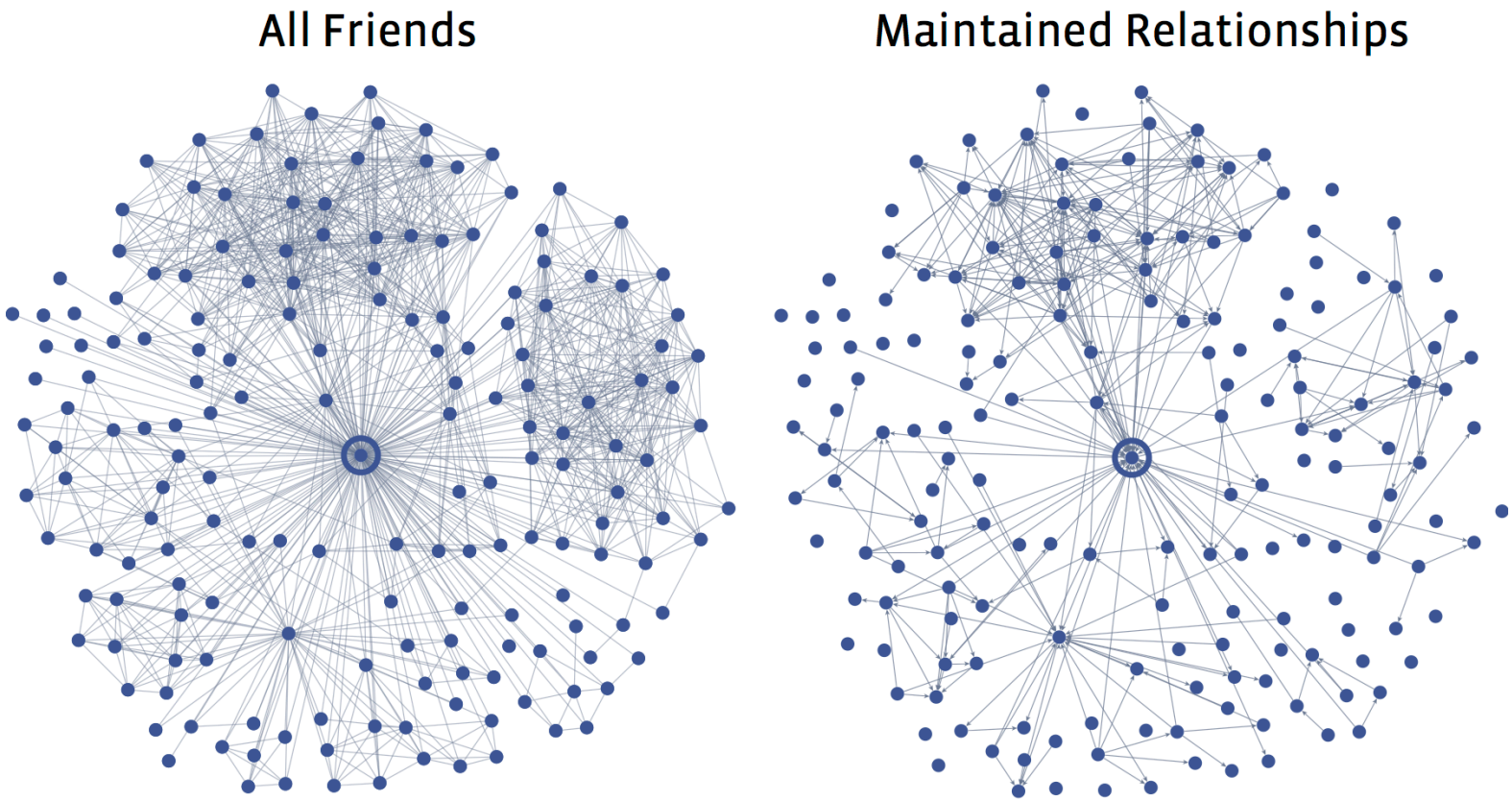
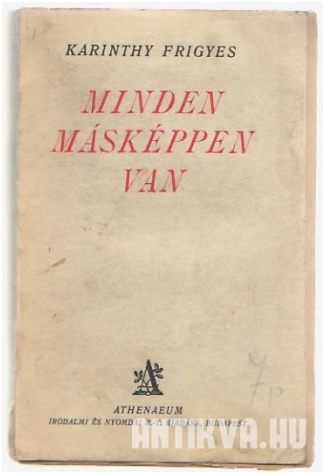


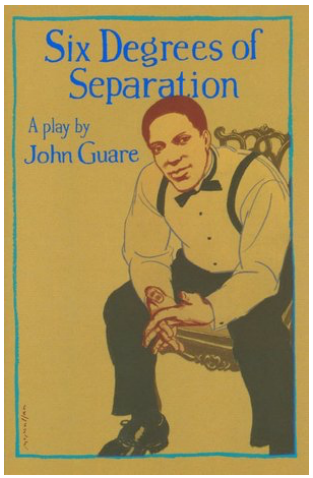
image from Cameron Marlow, Facebook

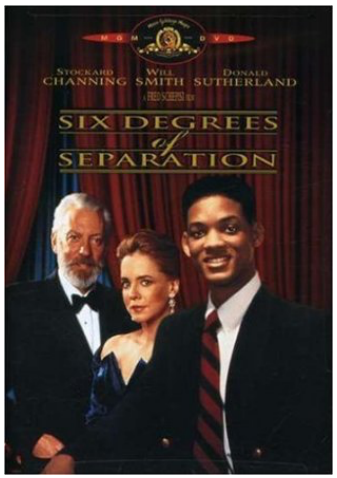
Six degrees of separation

"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.

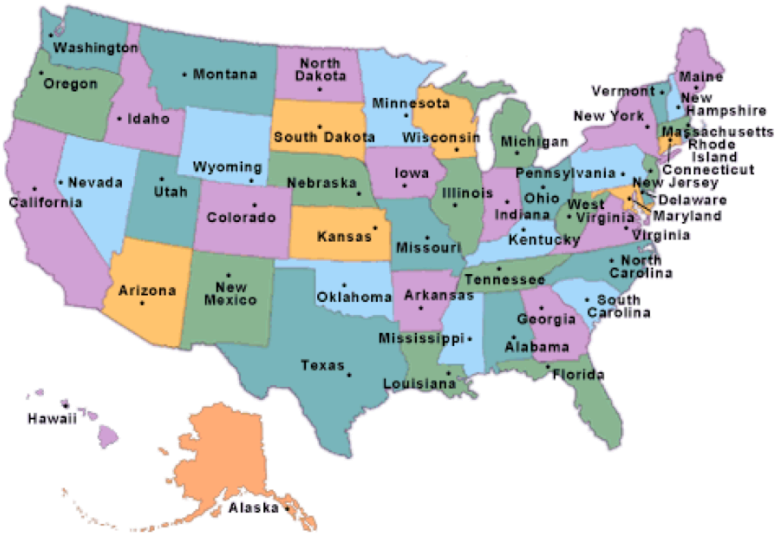
2. DETACH ONE POSTCARD. FILL IT OUT AND RETURN IT TO HARVARD UNIVERSITY. No stamp is needed. The postcard is very important. It allows us to keep track of the progress of the folder as it moves toward the target person.

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

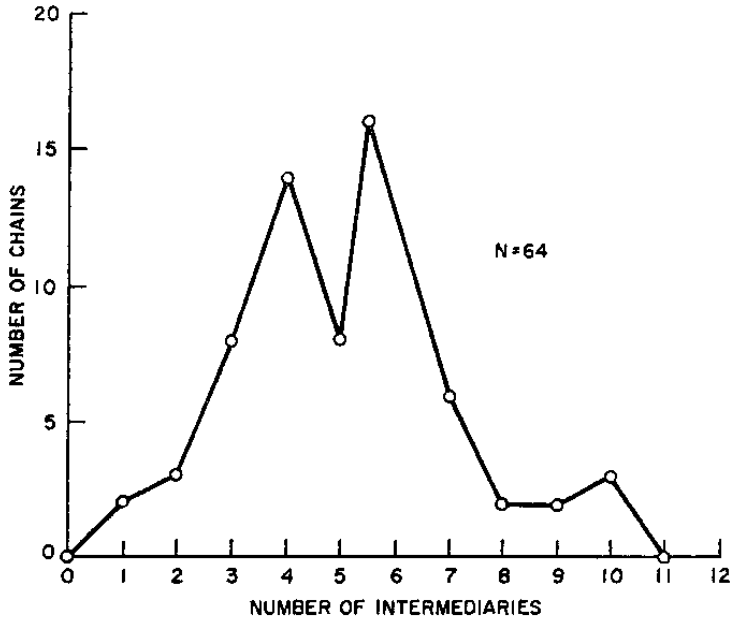
- Starting persons:
  - 296 volunteers, 217 sent
  - 196 in Nebraska
  - 100 in Boston
- Target person - Boston stockbroker
- Information given: target name, address, occupation, place of employment, college, hometown



J. Travers, S. Milgram, 1969

Stanley Milgram's 1967 experiment

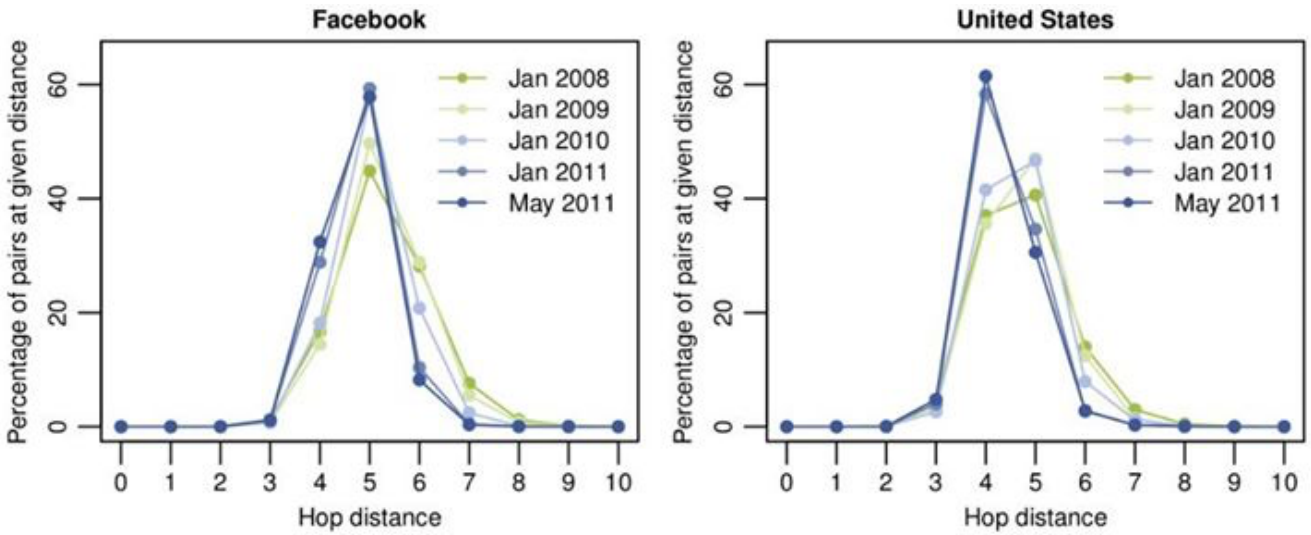
- Reached the target $N = 64 (29\%)$
- Average chain length $\langle L \rangle = 5.2$
- Channels:
  - hometown $\langle L \rangle = 6.1$
  - business contacts $\langle L \rangle = 4.6$
  - from Boston $\langle L \rangle = 4.4$
  - from Nebraska $\langle L \rangle = 5.7$



J. Travers, S. Milgram, 1969

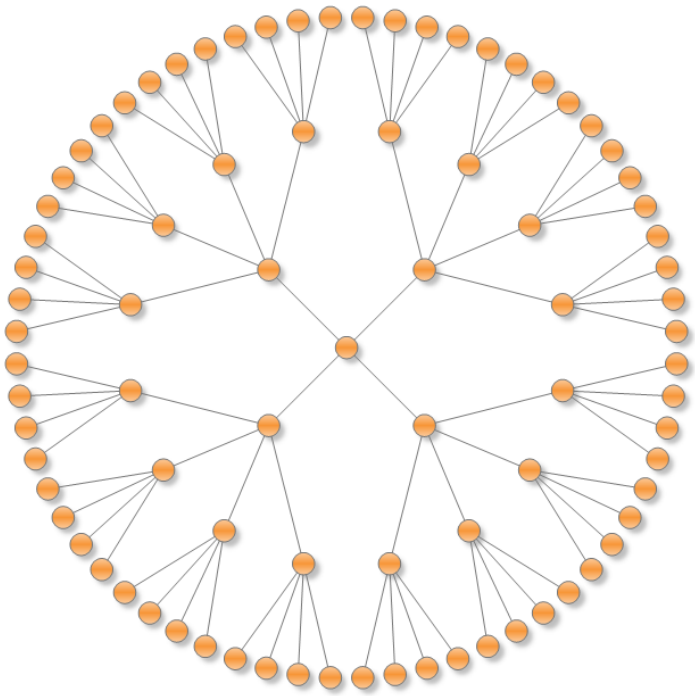
Small world

- 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. 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