Complex networks

Introduction to Network Science

Carlos Castillo
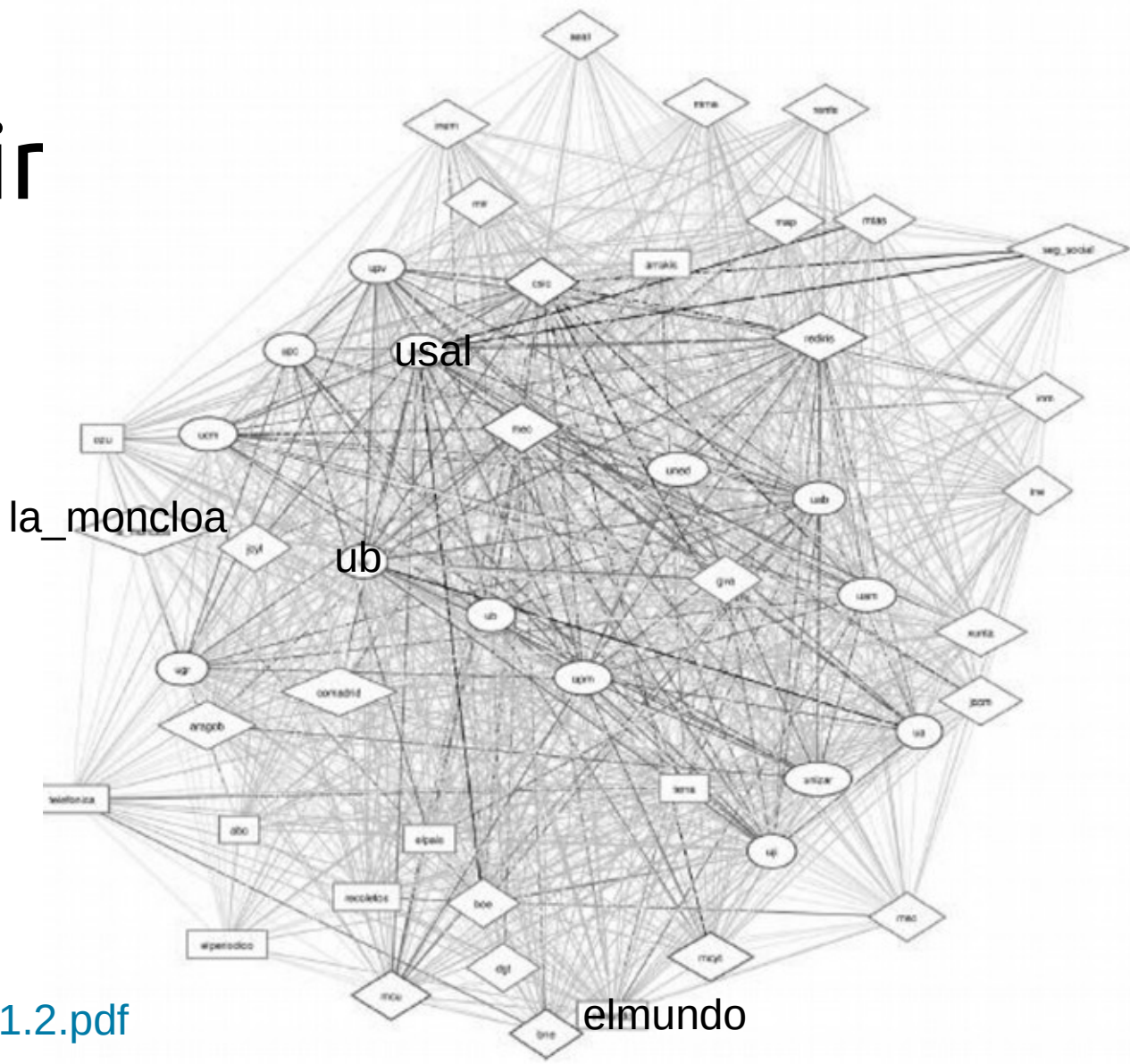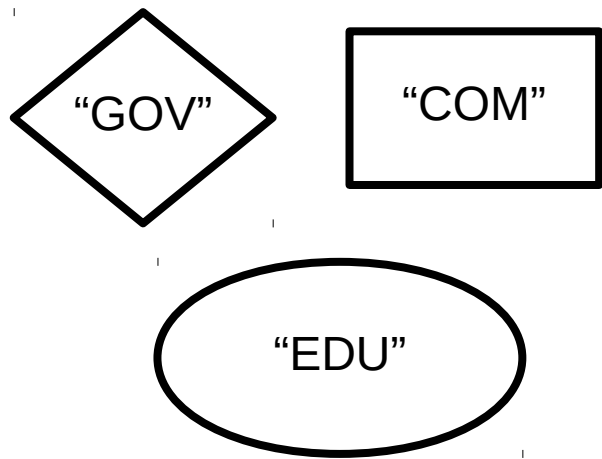
Topic 01

Sources

- Albert László Barabási: Network Science. Cambridge University Press, 2016.
  - Chapters 01-02
- URLs cited in the footer of specific slides

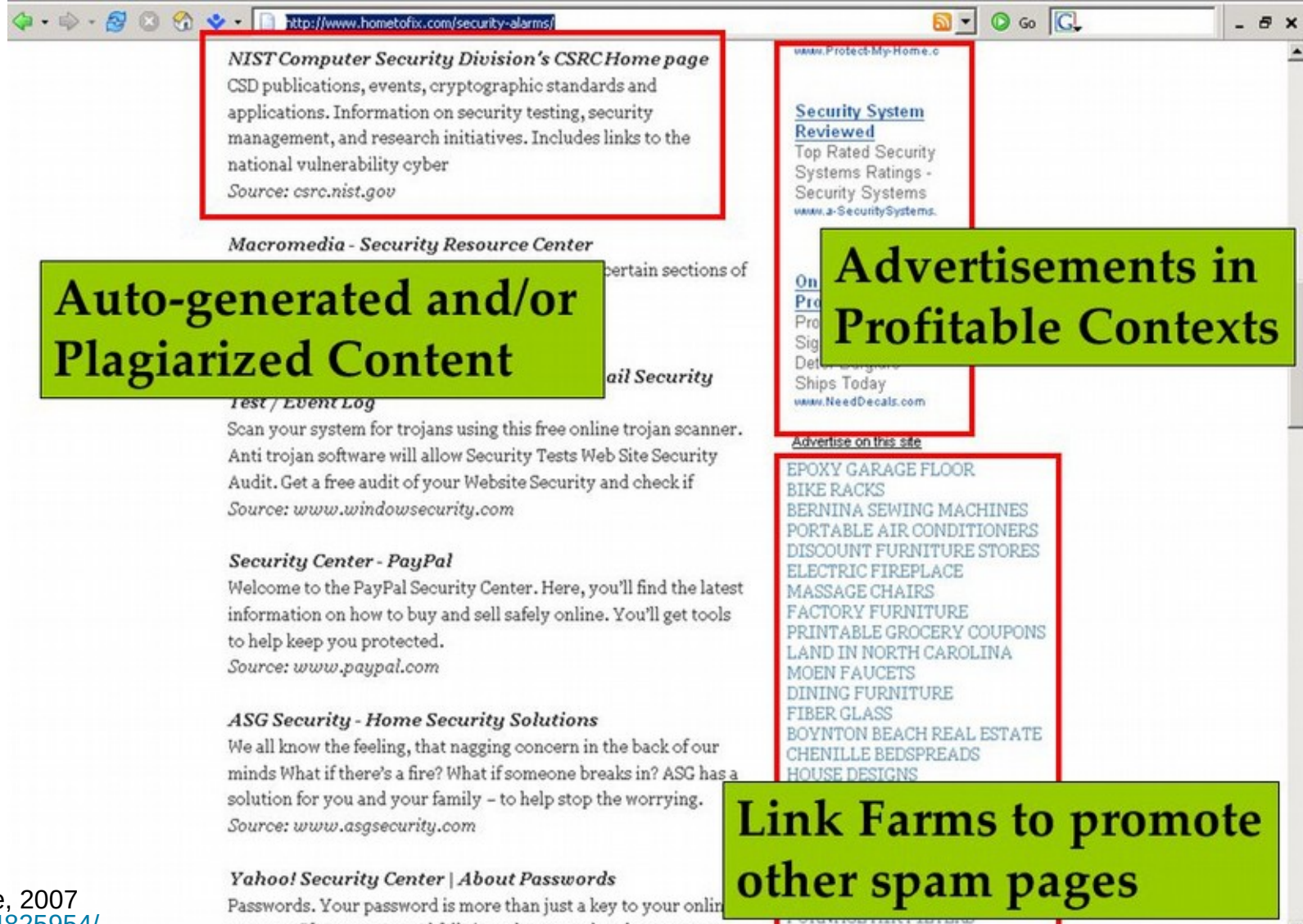# A personal introduction

PhD work (2000-2004)

- Collecting web pages

- Characterizing national web domains
  - Chile, Korea, Greece, Spain ...

Top domains ir .es (~2006)

"GOV"

"COM"

"EDU"

usal

la_moncloa

ub

elmundo

Post-doctoral work (2005-2009)

- Web spam pages
  - Pages created to deceive search engines
  - Attract traffic by stuffing themselves with keywords
  - Increase link score of other pages
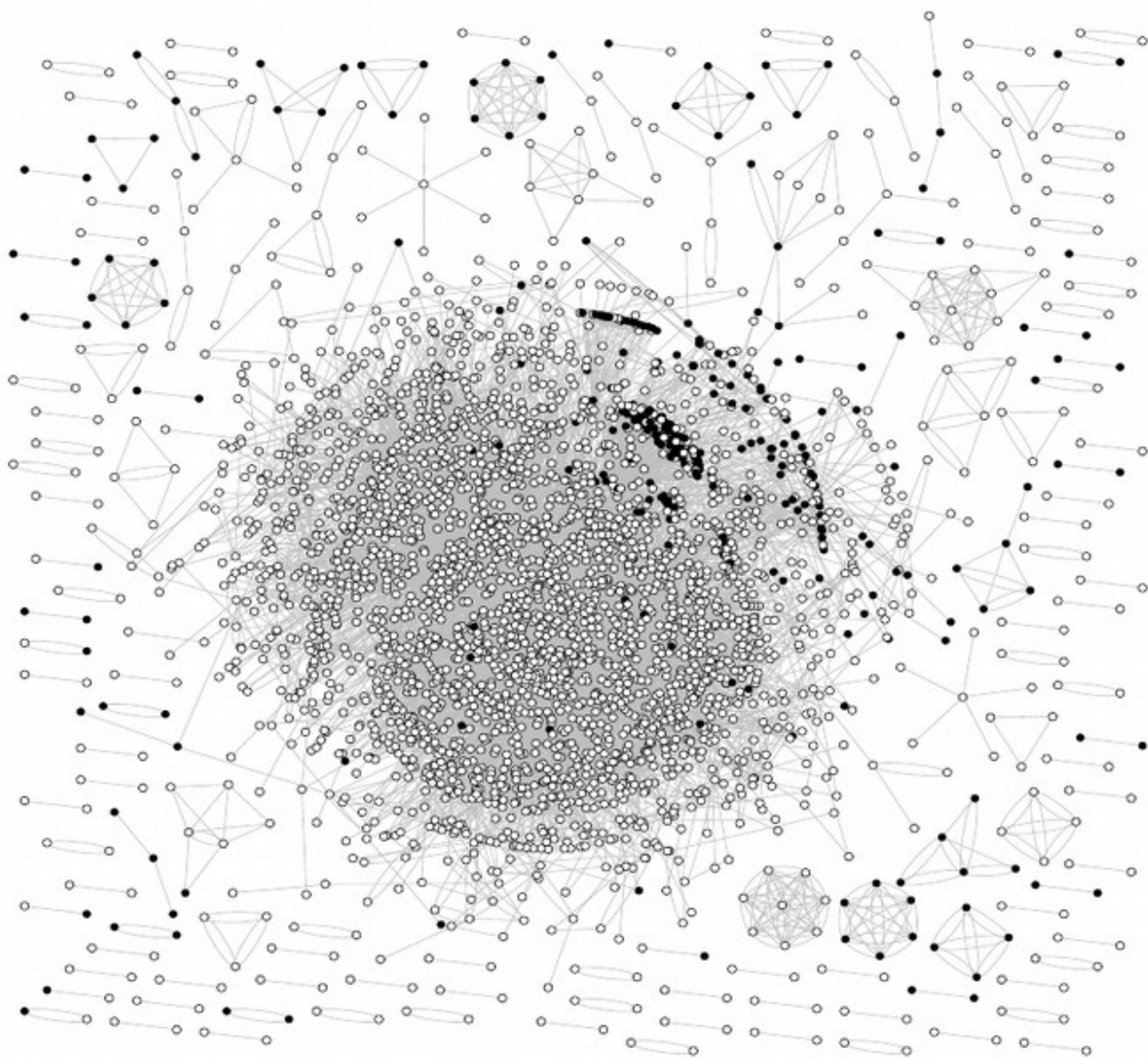 - Methods evolve all the time, how to catch them?

Auto-generated and/or Plagiarized Content

Advertisements in Profitable Contexts

Link Farms to promote other spam pages

Pranam Kolari's PhD defense, 2007
https://slideplayer.com/slide/4825954/

7

Eureka! Moment

2006

Use *gnuplot* to visualize an annotated dataset
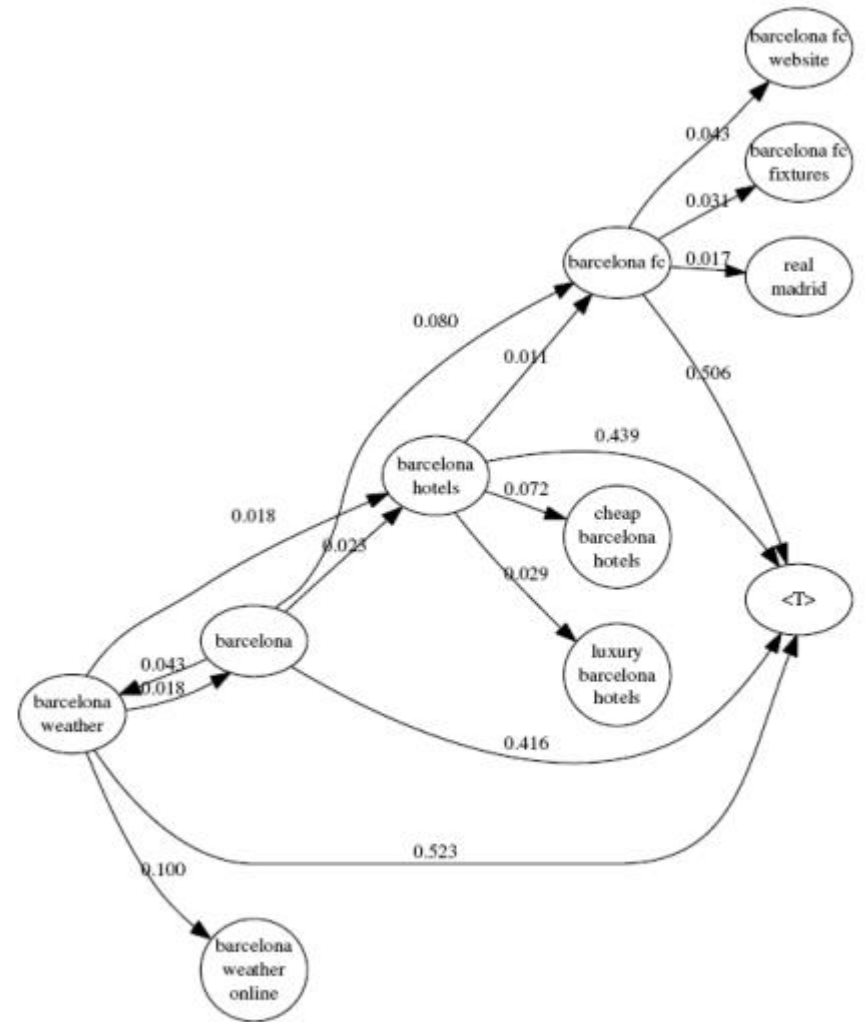
White: non-spam

Black: spam

Query flows

2008

What is the most likely query before or after another query?

How are they connected?

Graphs in my own work

- Characterizing national web domains

- Finding web spam pages

- Suggesting queries to searchers

- …

- Currently:
  - part of a larger toolbox

  - skeptical about structural-only conclusions

Network Science

Wikipedia definition

- **Network science** is an academic field which studies **complex networks** such as
  - telecommunication networks, computer networks, biological networks, cognitive and semantic networks, and social networks,
- considering
  - distinct elements or actors represented by **nodes** and
  - the connections between the elements or actors as **links**.
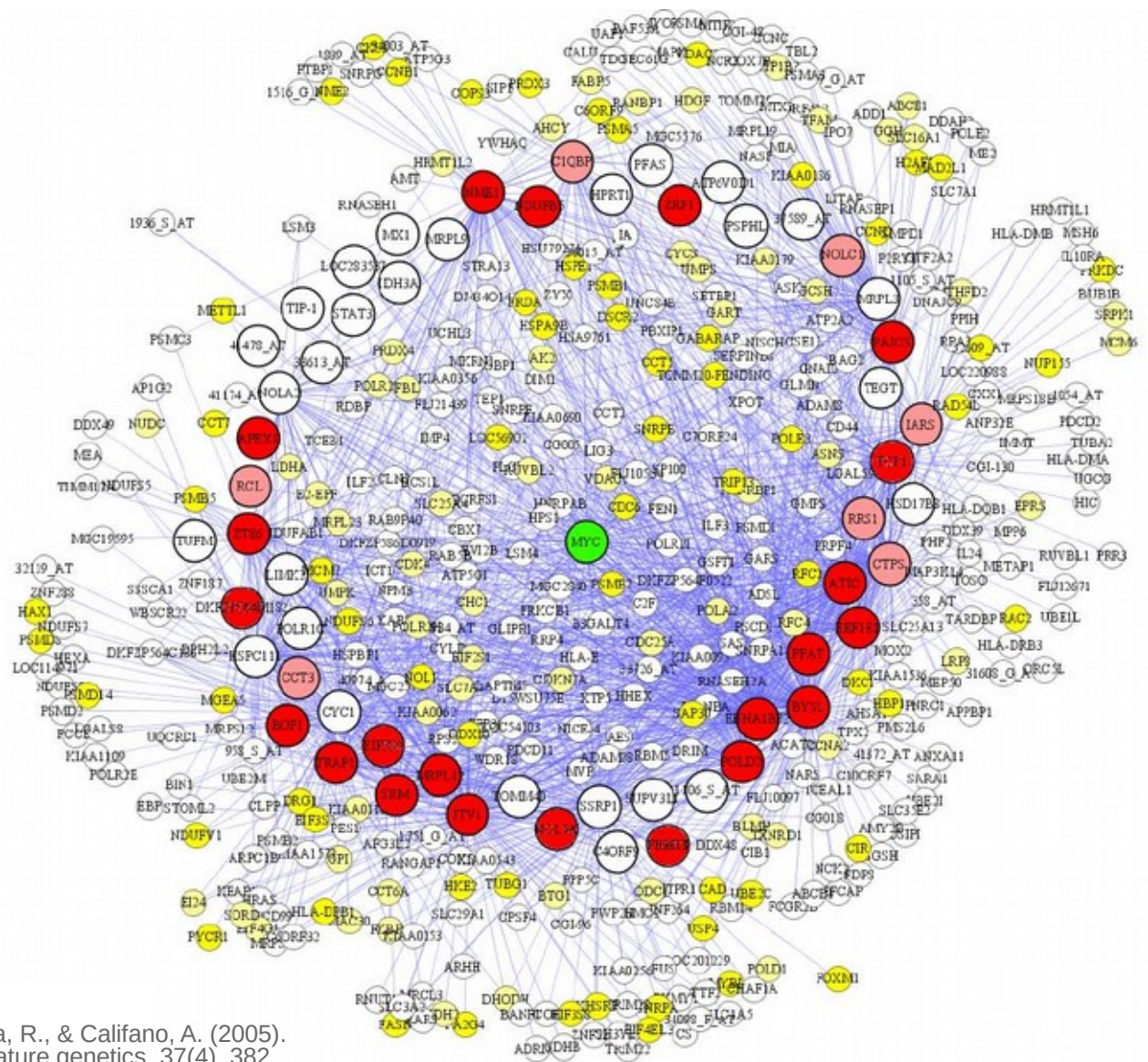
Complex systems

- Many interconnected parts
- Complicated arrangement of connections
- Emerging properties

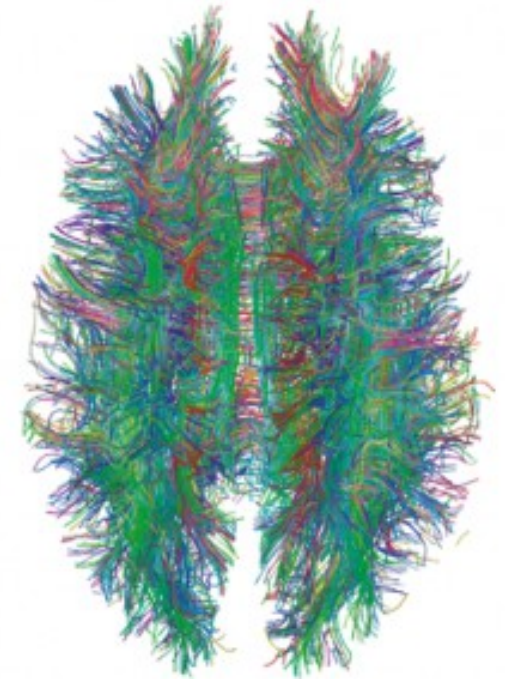# Behind every complex system there is a complex network
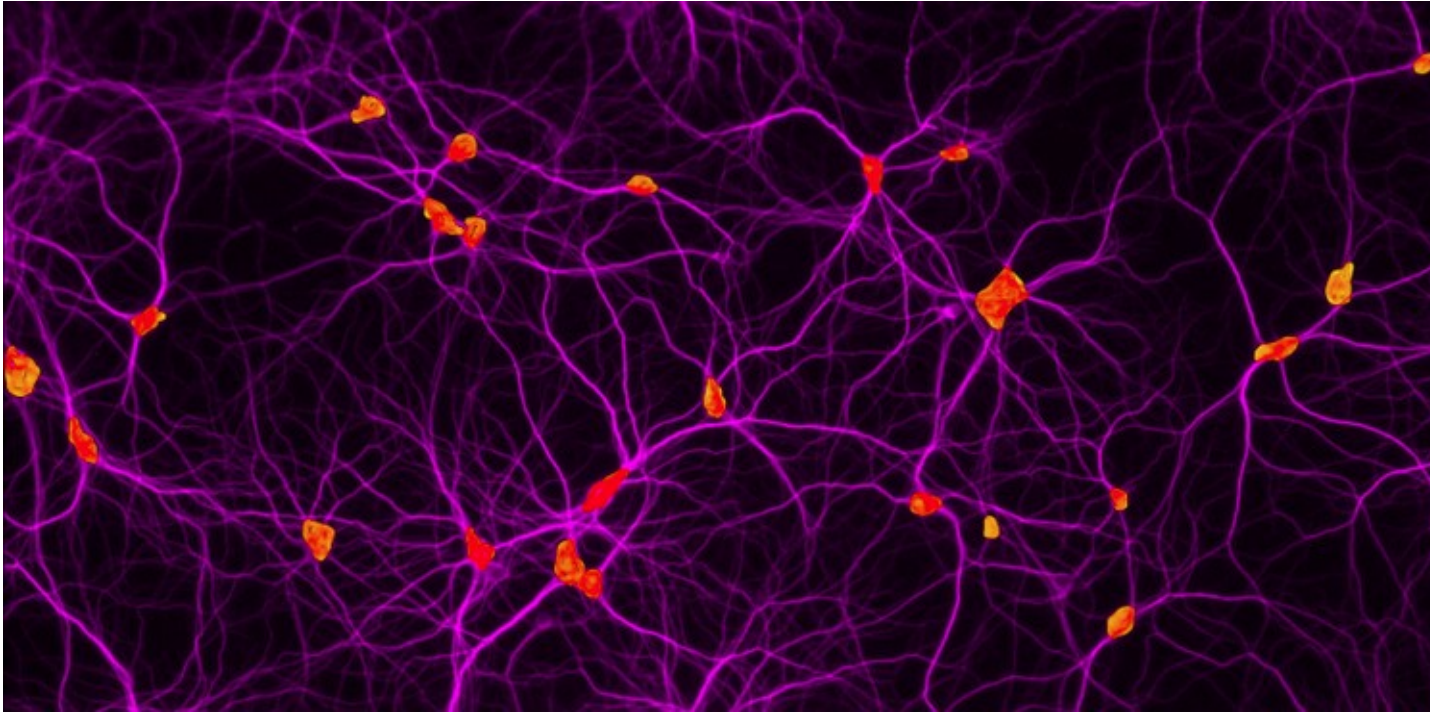
- Connections between neurons in the brain

- Interactions between genes and proteins

- Family/friendship links in human and non-human animals

- Infrastructure of telecommunications, electricity

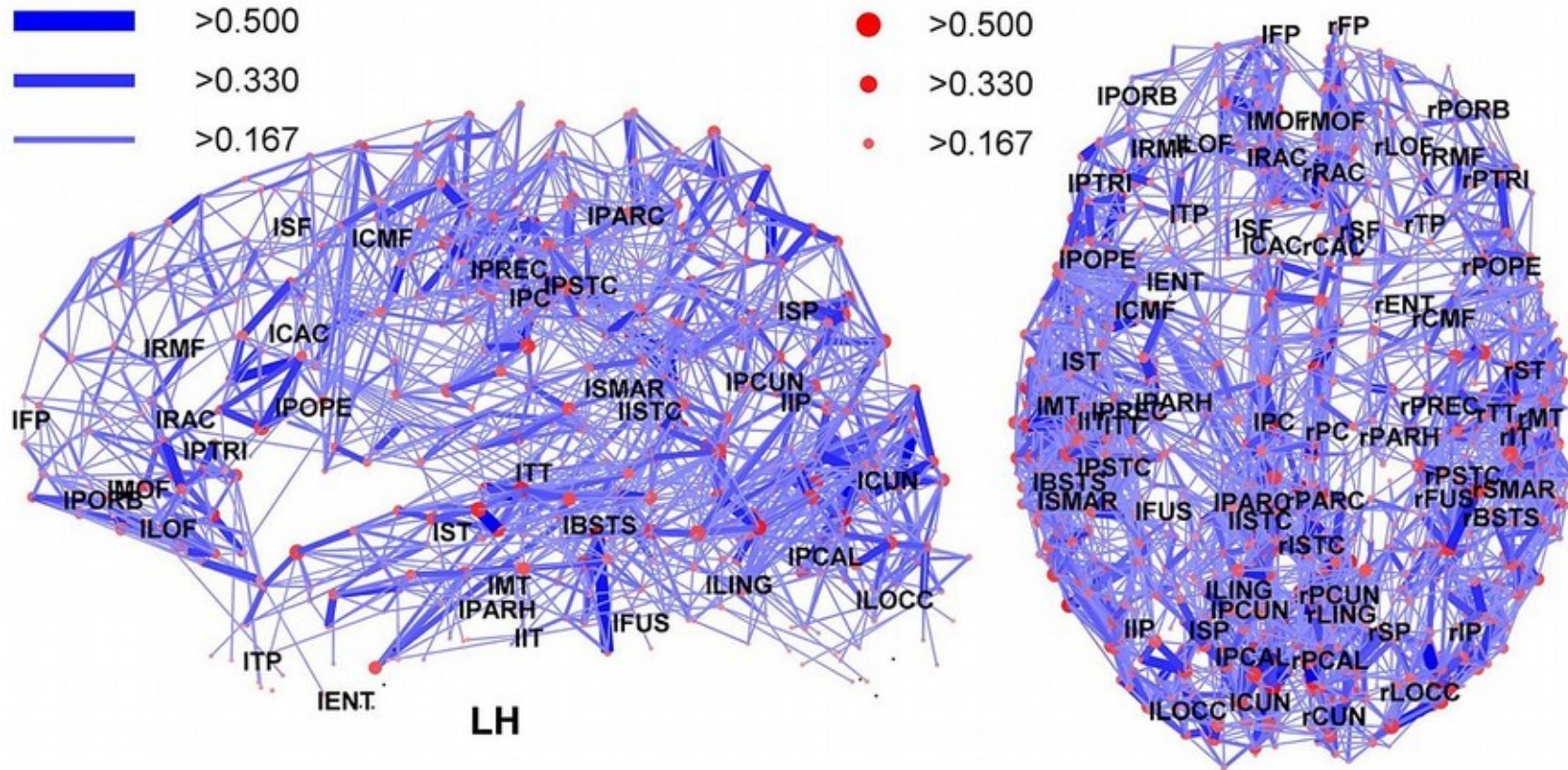- Commerce/trade networks

Co-expression of Genes

|V|=500 in this plot

Basso, K., Margolin, A. A., Stolovitzky, G., Klein, U., Dalla-Favera, R., & Califano, A. (2005). Reverse engineering of regulatory networks in human B cells. Nature genetics, 37(4), 382.
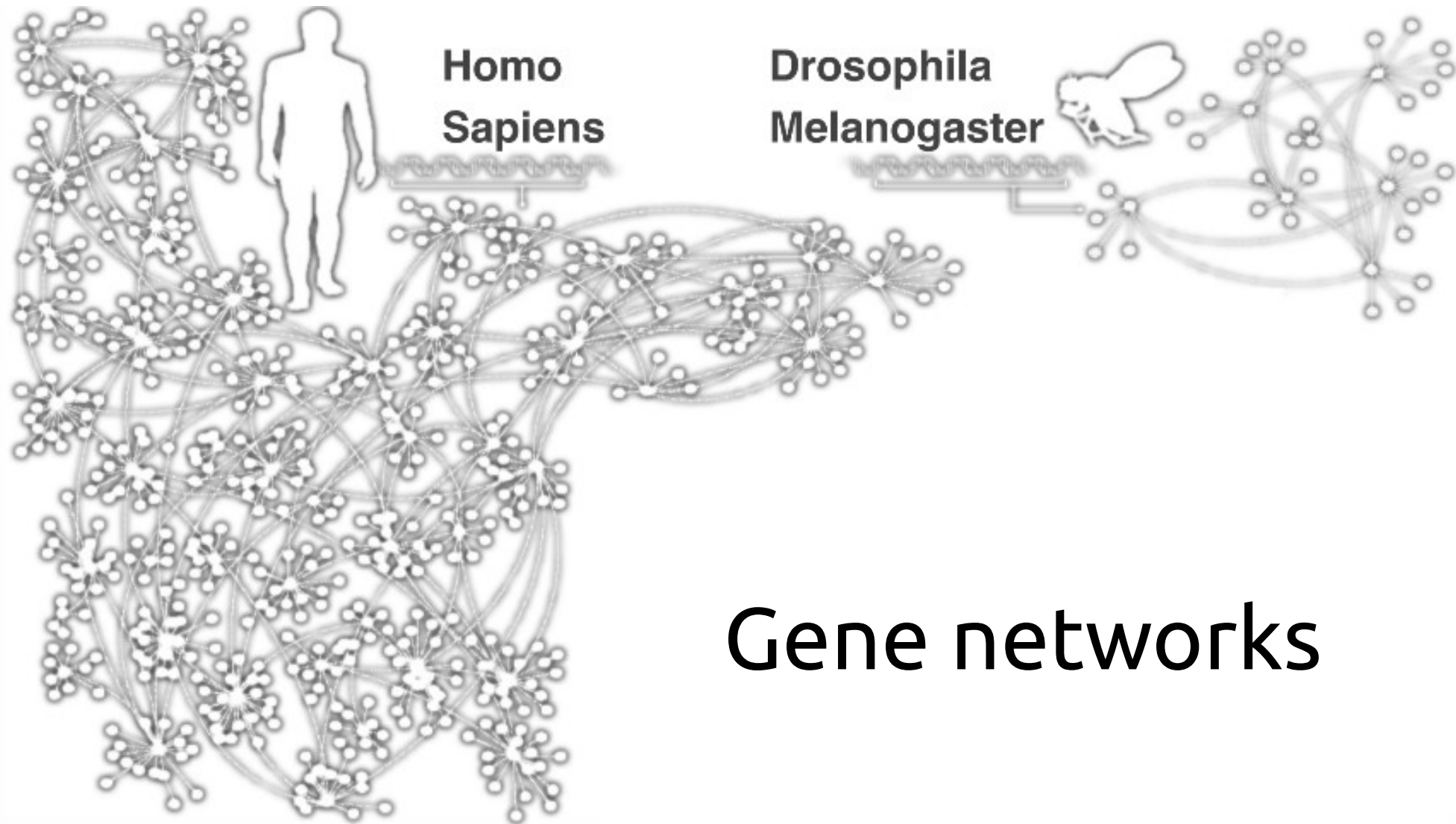
Human brain: $|V| \simeq 90 \times 10^9$

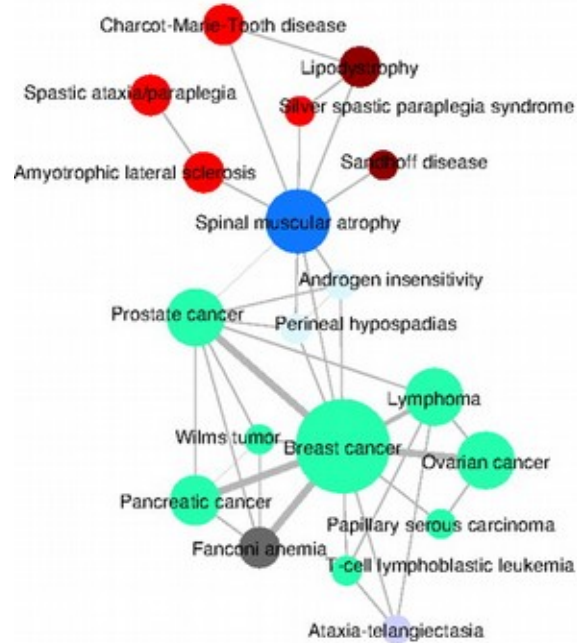# Regions in the human brain

Homo Sapiens

Drosophila Melanogaster

Gene networks
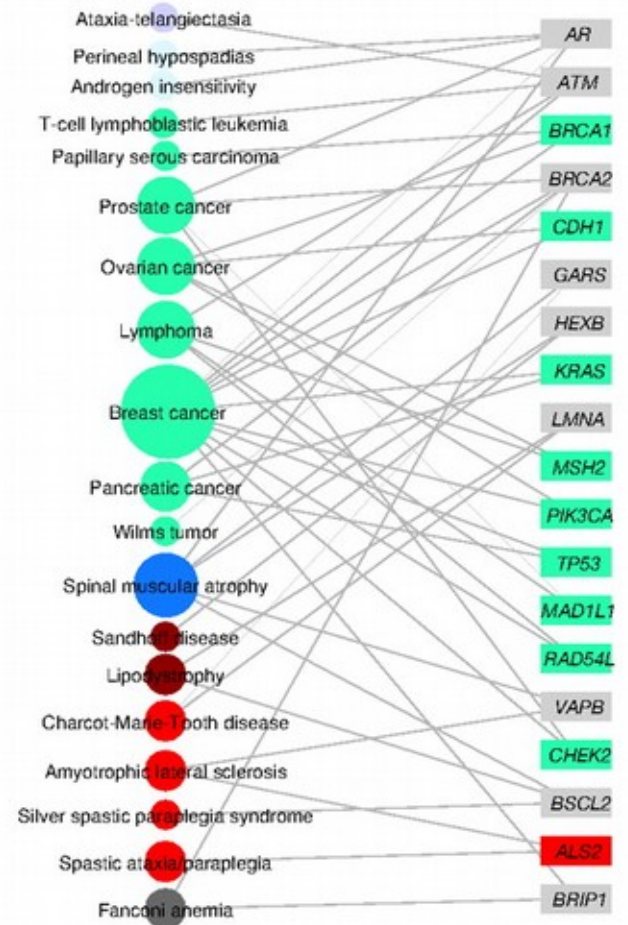
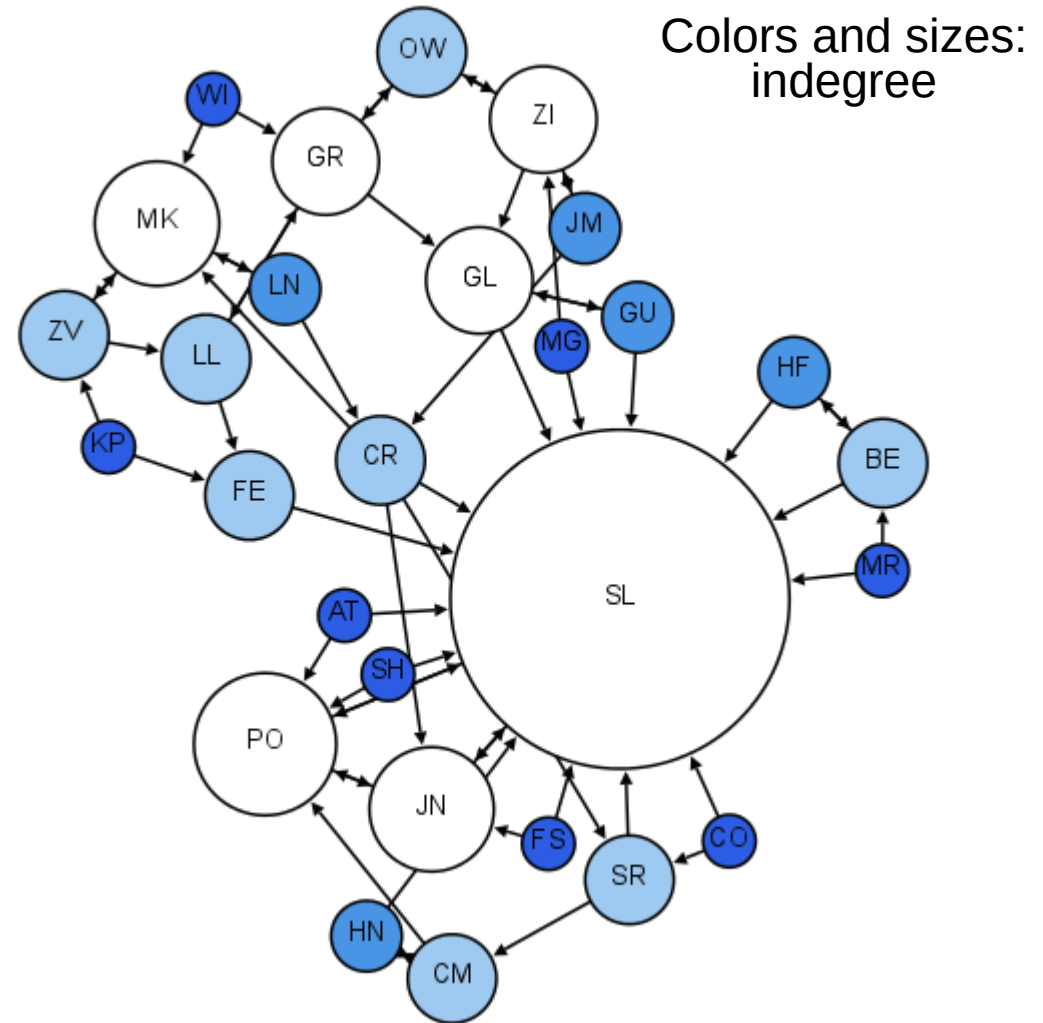# Human disease network

Human Disease Network (HDN)

DISEASOME
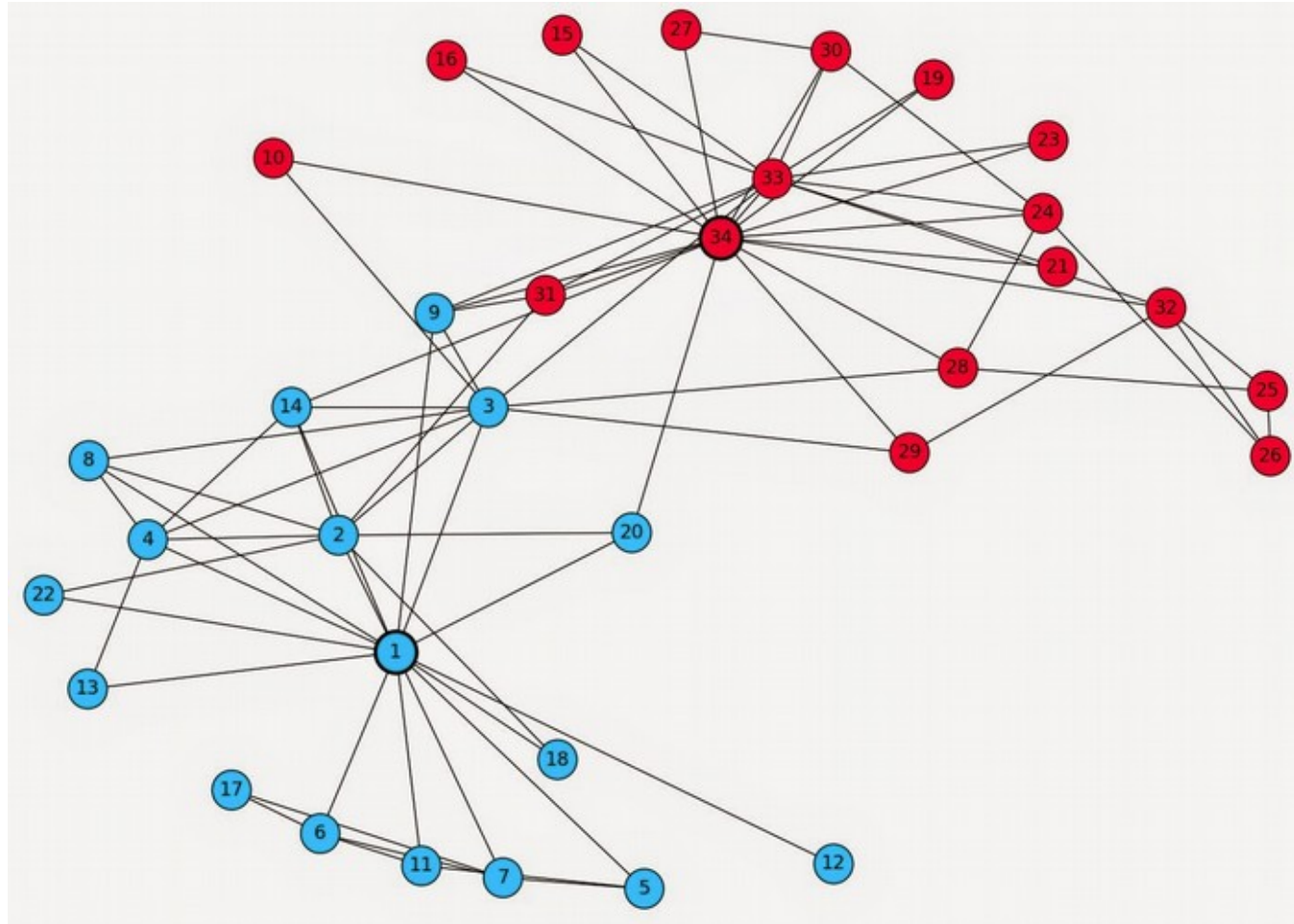
disease phenome — disease genome

Moreno's sociograms

- Early 1930s
- Children in 2nd grade
- Who would you like to sit with?

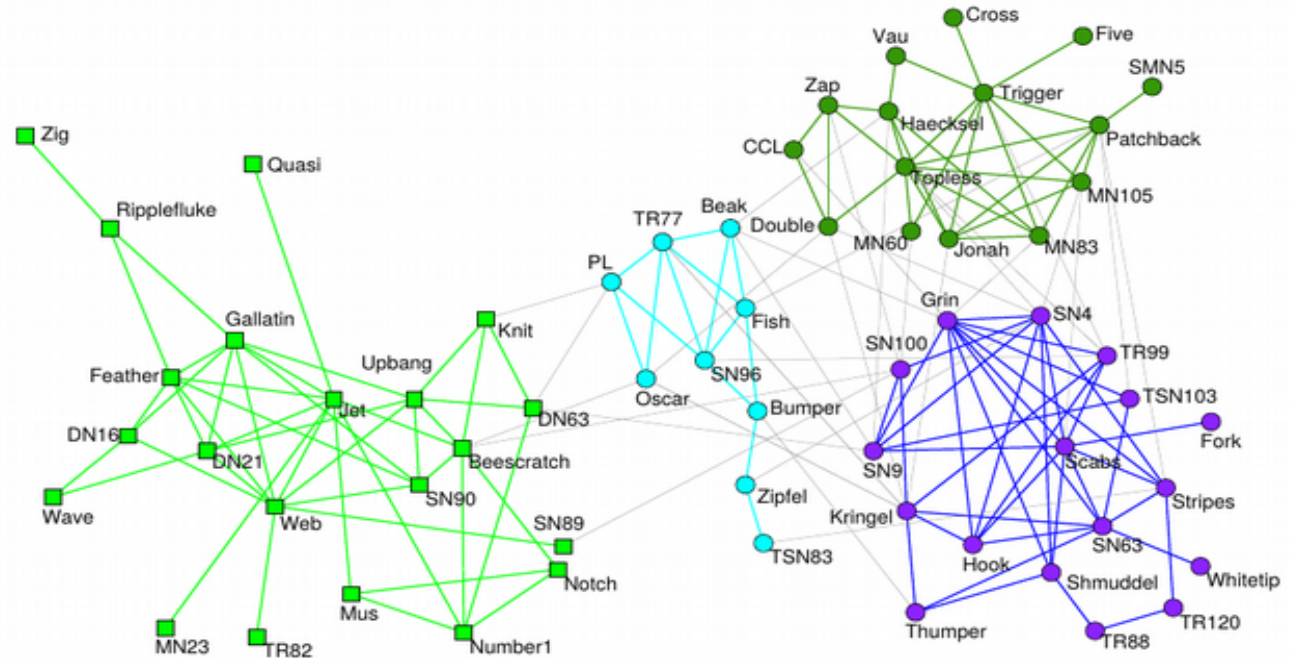Colors and sizes: indegree

Zachary's Karate Club

Karate club that split into two clubs

(led by 1 and 34)

Dolphins in a fjord in New Zealand

- Research following a school of dolphins in the wild (2003)

- Look for dolphins swimming together

- Found **long-lasting associations;** research has been repeated with other non-human animals (e.g., sheep)
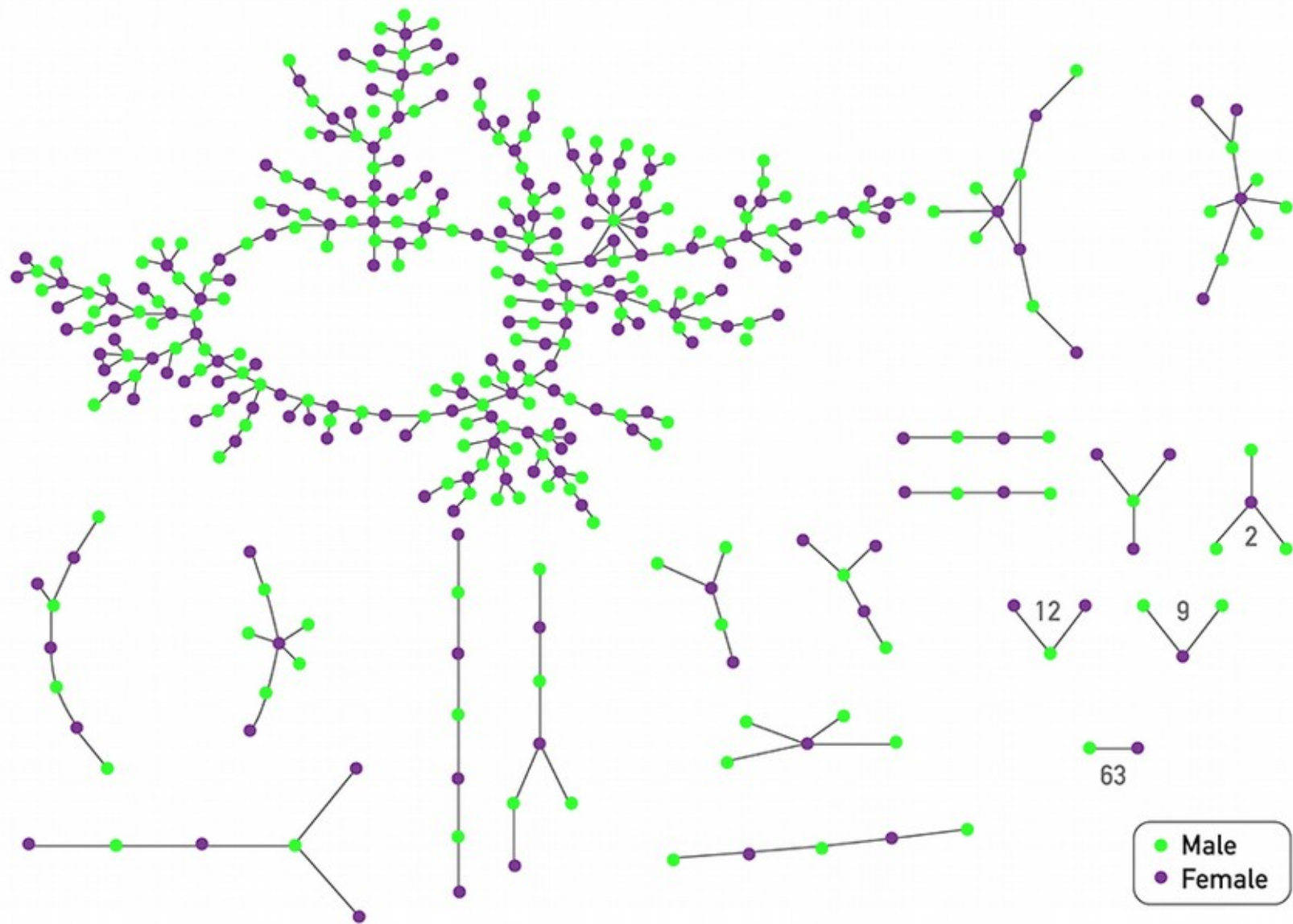
22

Male
Female

Chains of affection

- Early 2000s

- Adolescents in high school

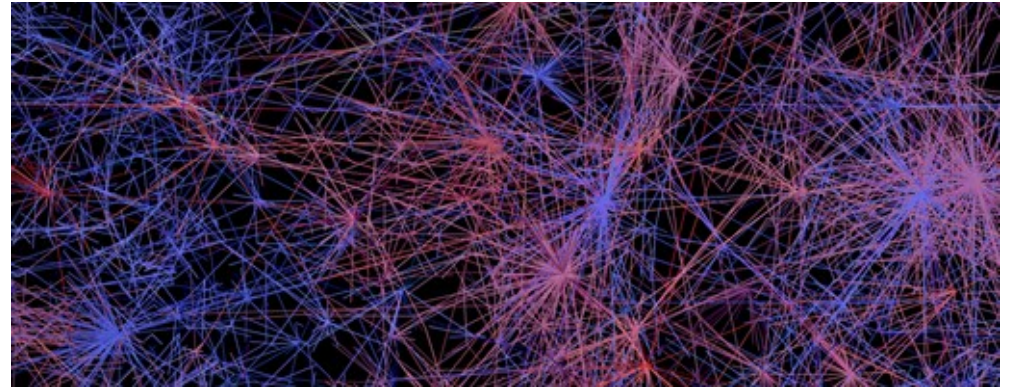- A "*special romantic relationship*" or a "*nonromantic sexual relationship*" in the past 18 months
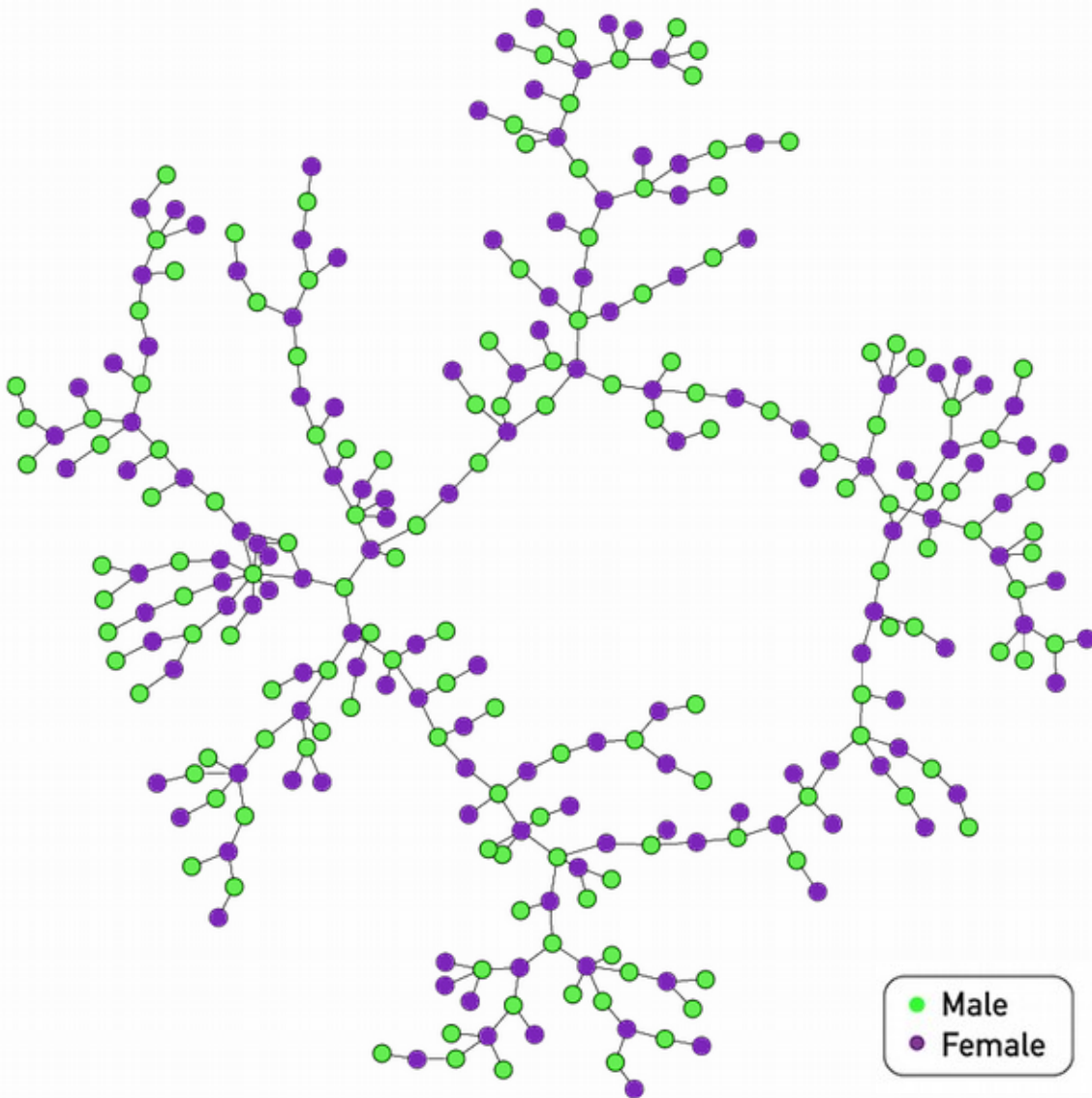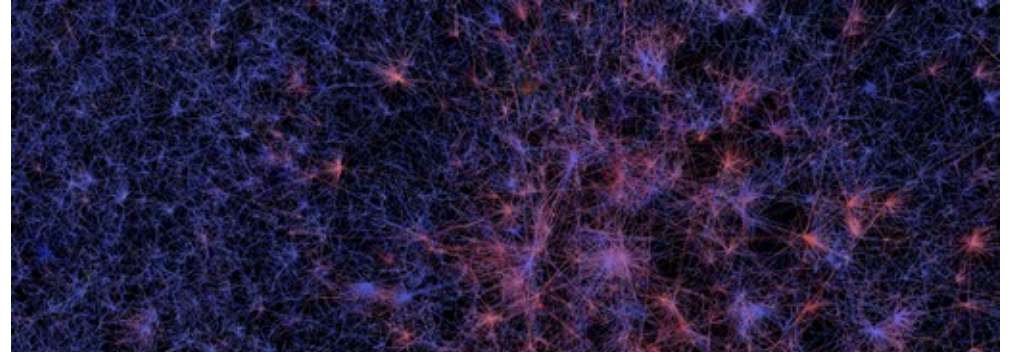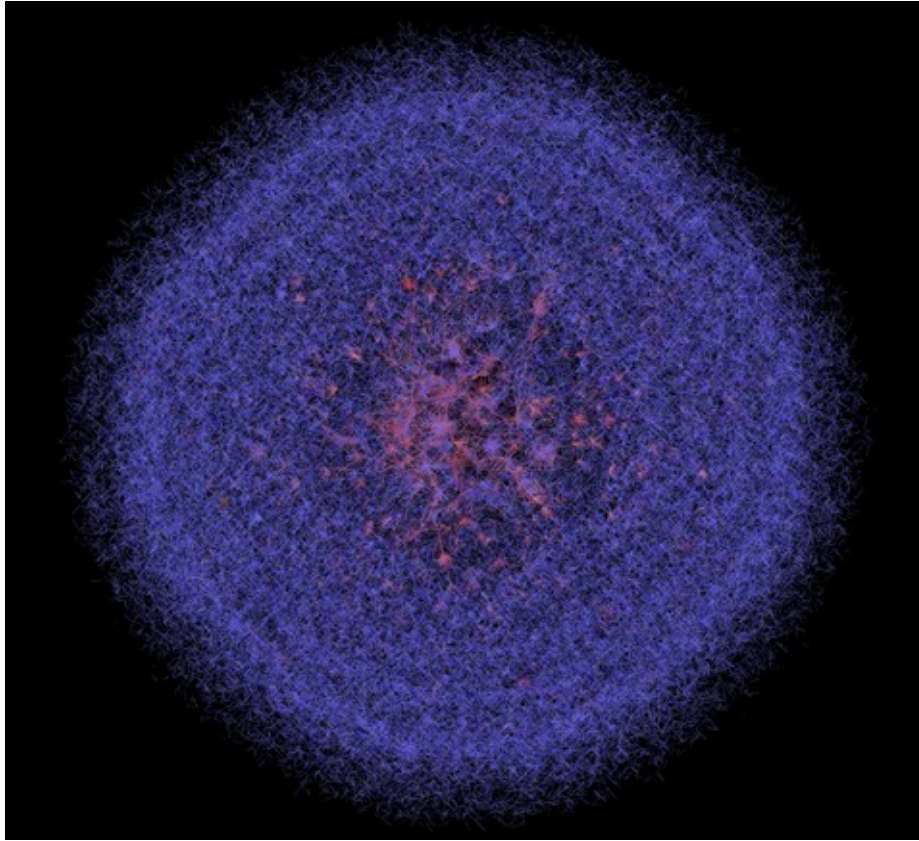
Bearman, P. S., Moody, J., & Stovel, K. (2004). Chains of affection: The structure of adolescent romantic and sexual networks. American journal of sociology, 110(1), 44-91.

Male
Female

24

1,000 Somali Users of Facebook

25
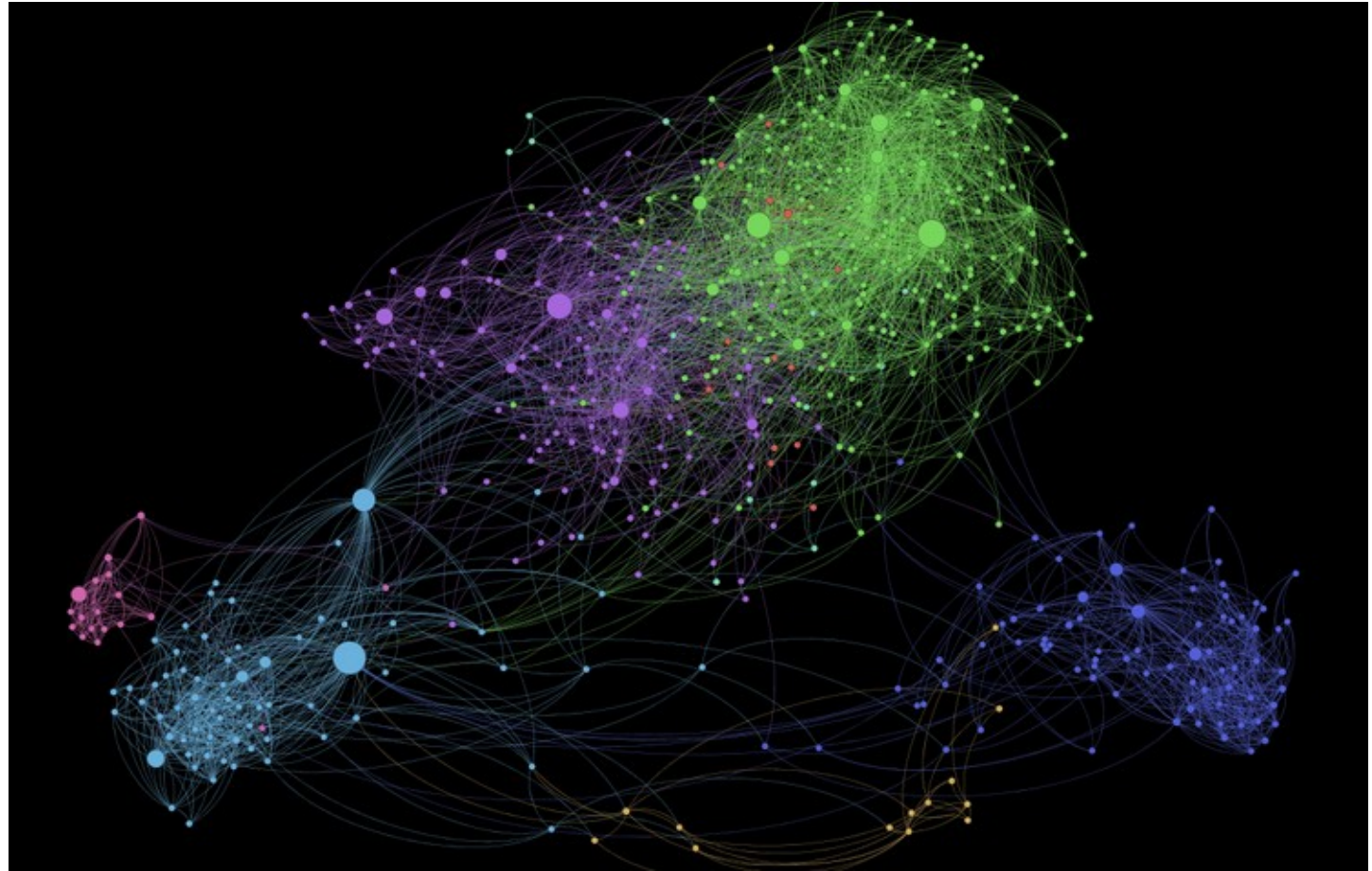
400,000 Twitter Users

What could these networks have in common? Why those regularities could be relevant? How would you find out what they are?

Universality of complex networks

"A key discovery of network science is that the architectures of networks emerging in various domains of science, nature and technology are similar to each other, a consequence of being governed by the same organizing principles."
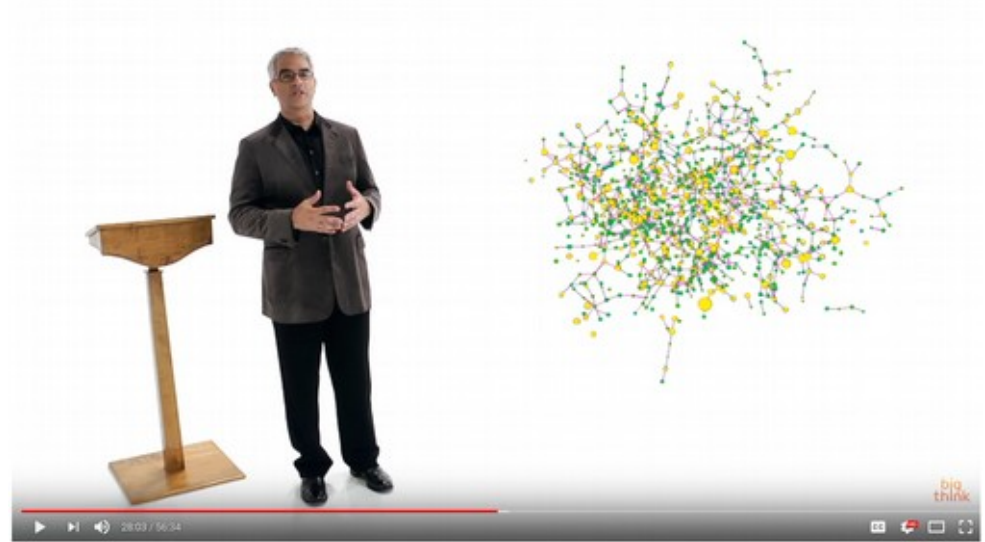(Barabási 2016)

Characteristics of network science

- **Interdisciplinary**; indeed we often address problems from disciplines other than CS

- **Empirical** and data-driven; it is based on the observation of networks

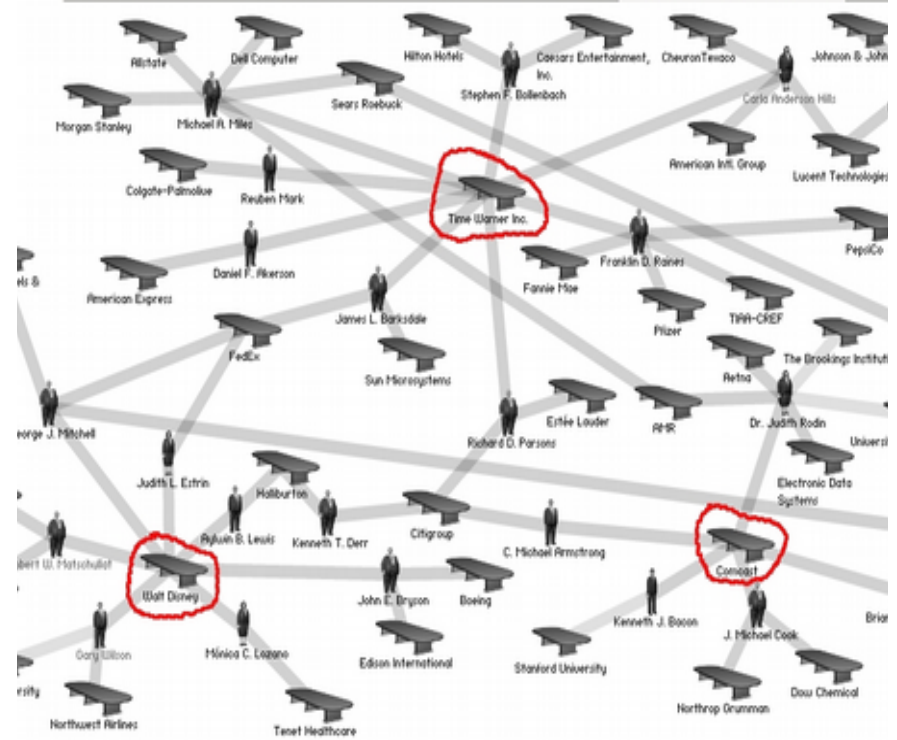- Quantitative, mathematical, **computational**

What can you do with this?

- Help design new treatments and drugs

- ...

Highly recommended:
Nicholas Christakis

https://www.youtube.com/watch?v=wadBvDPeE4E

Help fight organized crime and collusion

https://itnews.iu.edu/articles/2014/complex-networks-researcher-at-iu-fighting-crime-with-mobile-phone-data.php
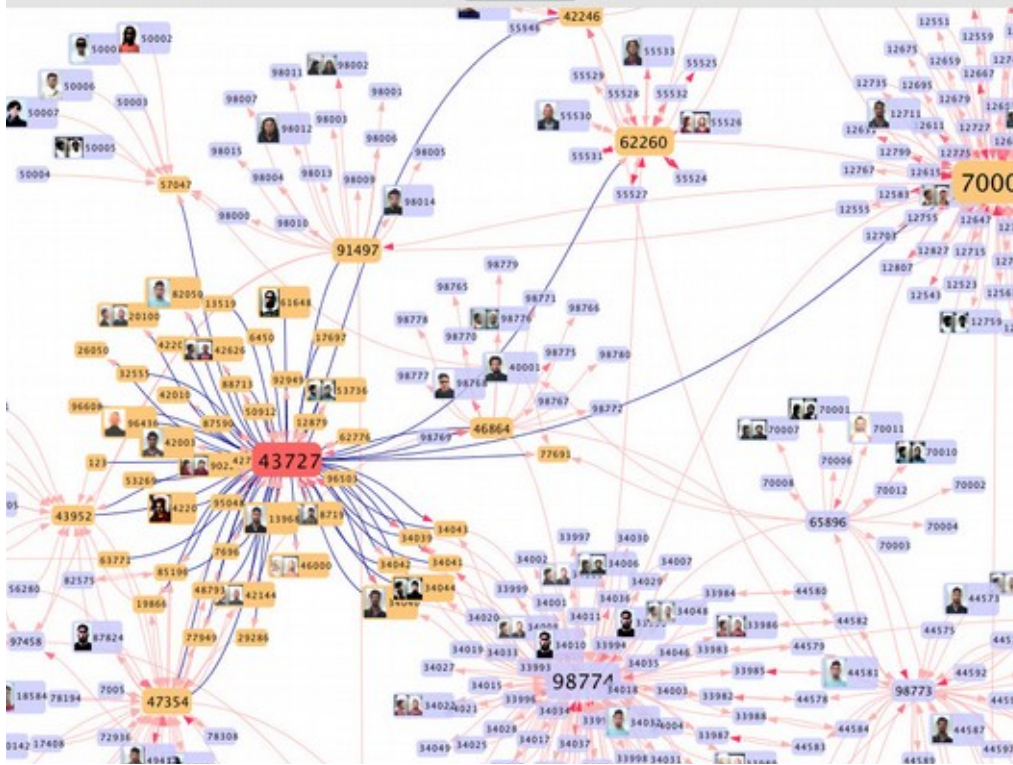
https://en.wikipedia.org/wiki/File:Media_corporation_interlocks_-_2004.jpg

Help fight corruption

https://theintercept.com/2018/08/16/chicago-police-misconduct-social-network/

32

Help to forecast epidemics

https://www.youtube.com/watch?v=mm2u9RKwgsY

Help understand an organization, a society, or a brain

Directors/CEO
Top Managers
Managers
Group Leaders
Associates

What can you do with this?
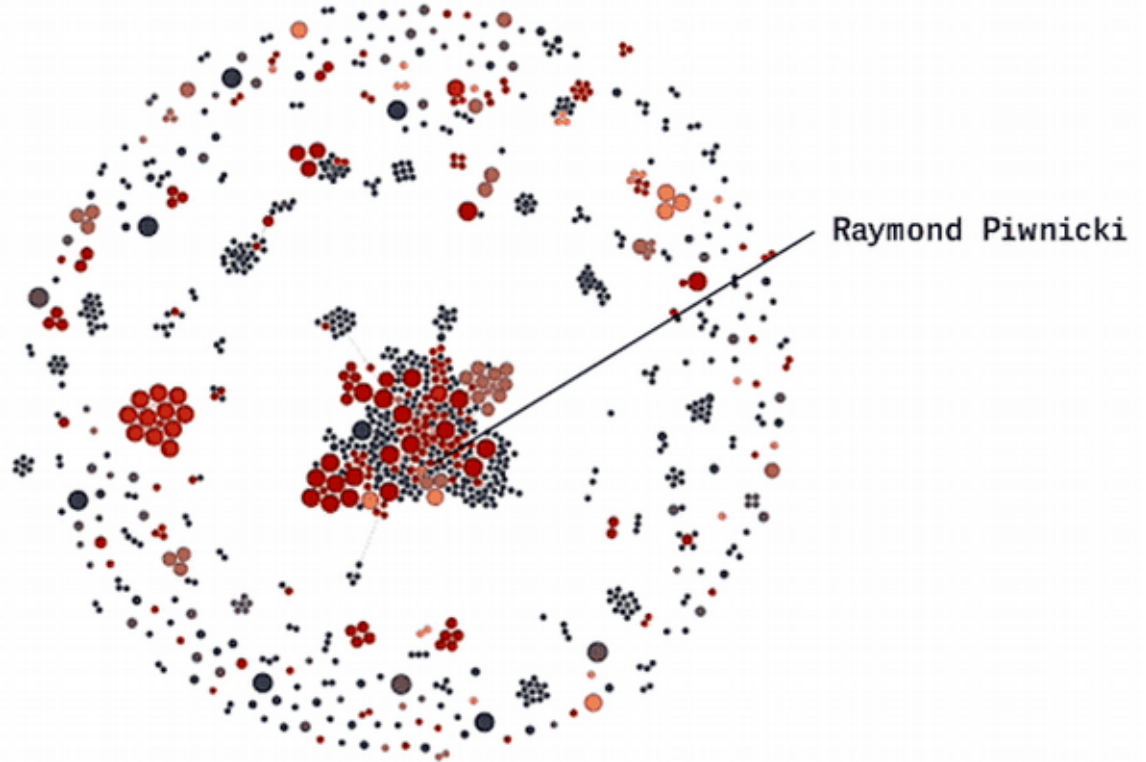
- Help design new treatments and drugs
- Help fight organized crime and corruption
- Help to forecast epidemics
- Help to understand an organization, a society, or a brain

What you can learn in this course

- To describe a network in formal terms
- To identify it as such and characterize it
- To visualize different networks
- To operate with networks programmatically
- To find important nodes and communities
- To make discoveries or help others make them
- **Much more (to a large extent, it's up to you!)**

Organization of the course

- Theory sessions:
  - Help you understand how to model complex networks
  - Help you find important nodes, communities, and track influence
  - Do some simple (and not so simple) exercises to check that you understood correctly each concept, and to help you remember
  - Follow "Network Science" by Barabási

- Practice sessions:
  - Help you work with complex networks
  - Manage and analyze graphs in Python
  - New material

- **My focus is on what I think has value for you as a data scientist**