

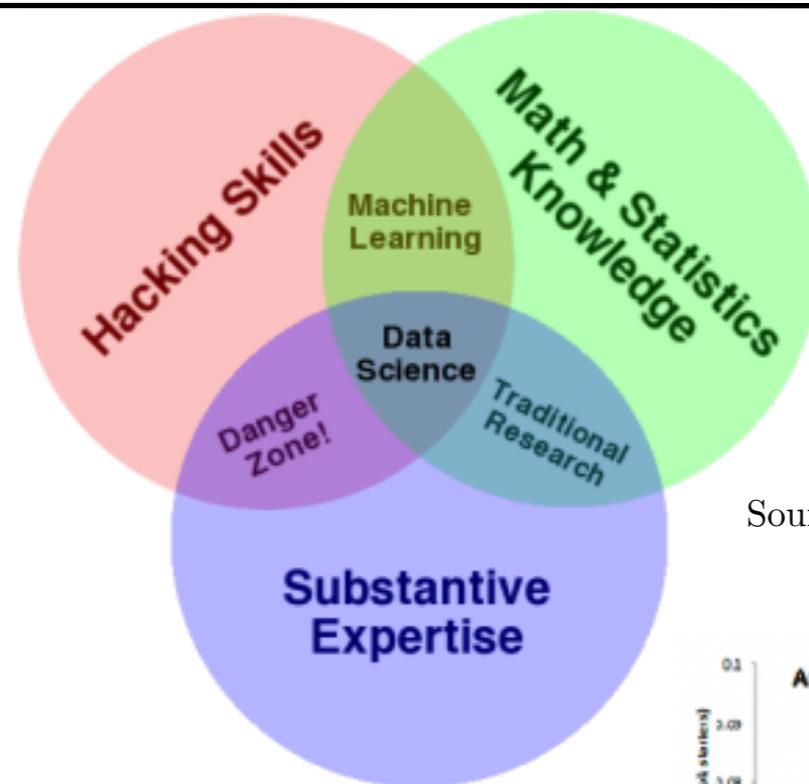
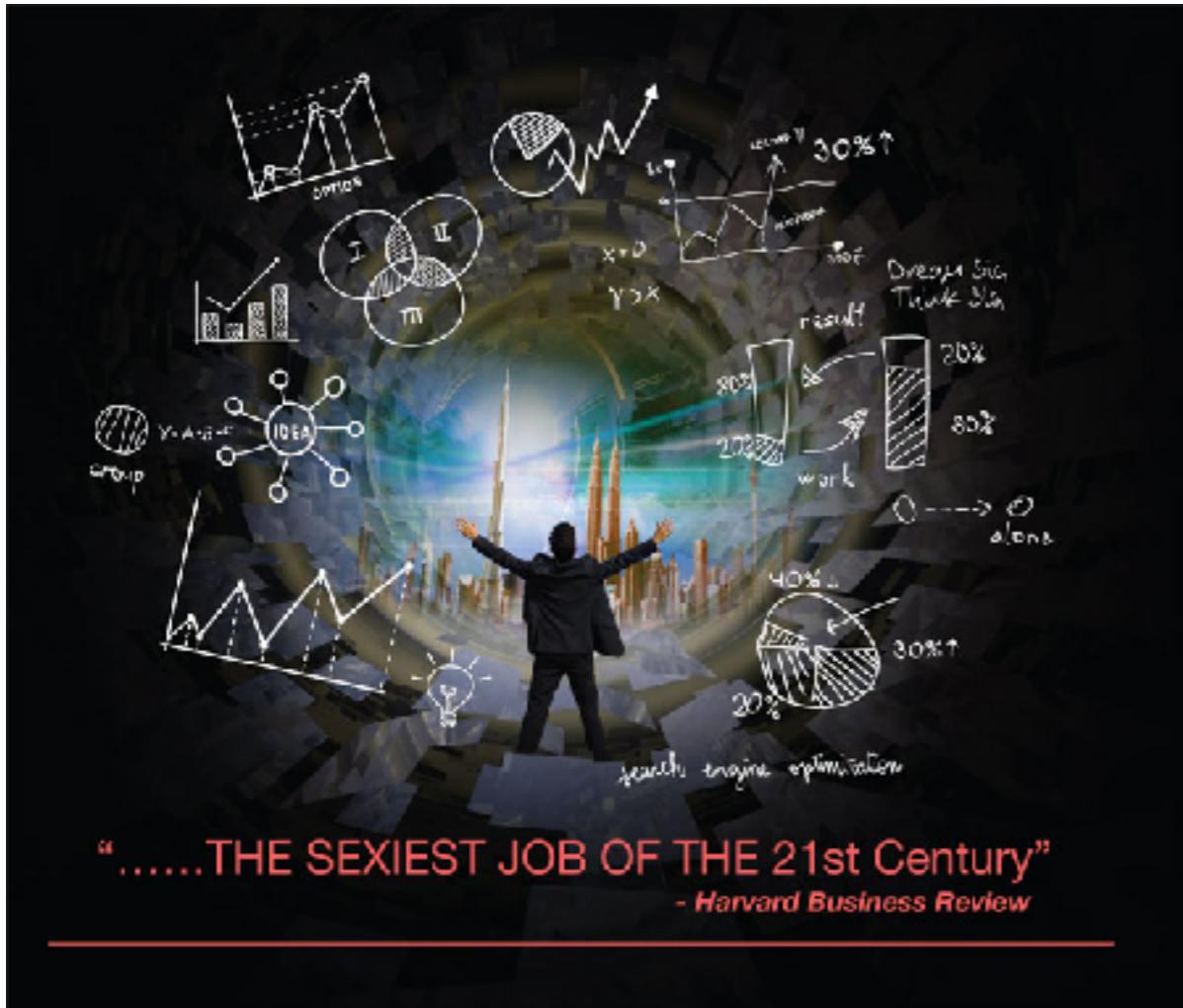
A Network Tour of Data Science - Introduction

Prof. Pierre Vandergheynst
Prof. Pascal Frossard

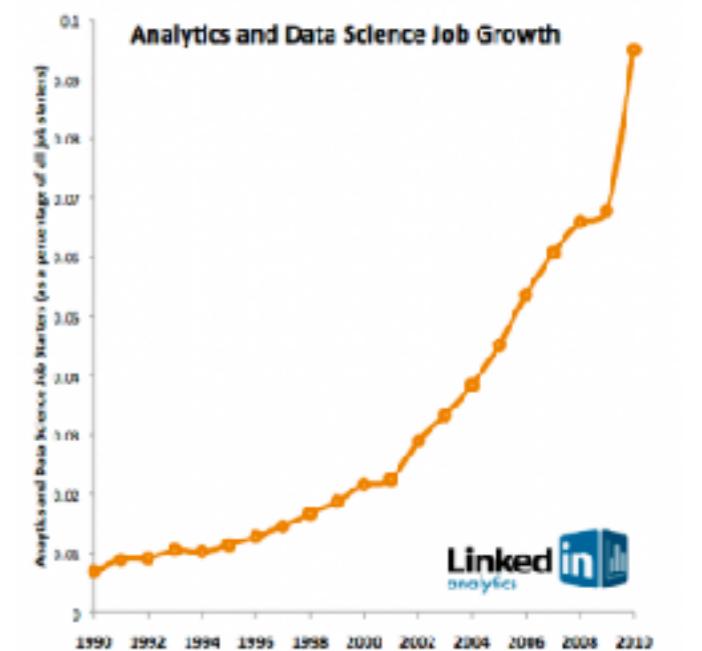
Some slides are taken from Prof. Xavier Bresson's class 'A Network Tour of Data Science (2016)',
and from Prof. Barabási's class on Network Science (www.BarabasiLab.com)



Data scientist



Source: Drew Conway



Best job in the U.S in 2015 [Forbes, LinkedIn].

Salary has jumped from \$125,000 to \$200,000+ [Glassdoor].

McKinsey projects that “by 2018, the U.S. alone may face a 50 percent to 60 percent gap between supply and requisite demand of deep analytic talent.”

News coverage

The New York Times

Violent Protests in Egypt As Leader Extends Power

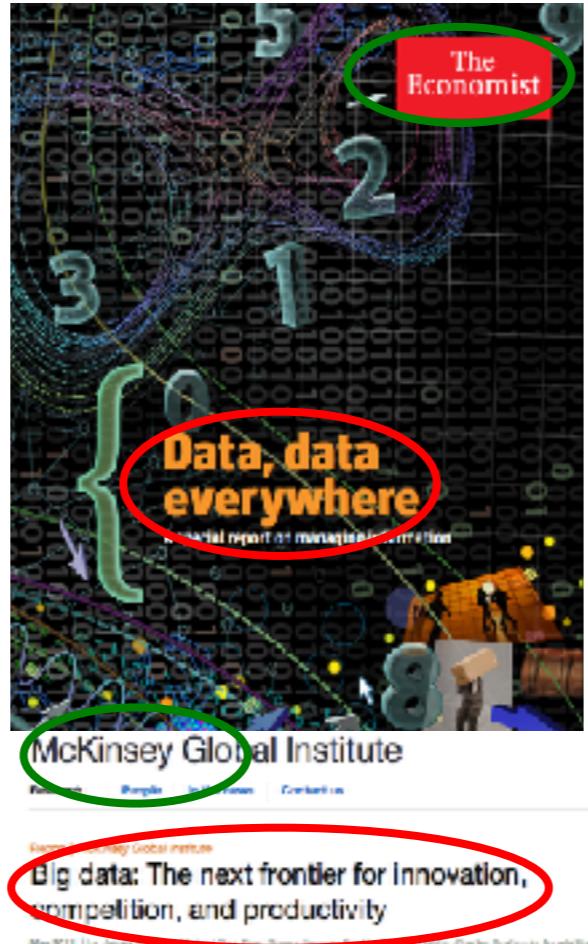
McKinsey: Data, data everywhere

Learning Curve: New Zealand Wants a HollywoodPax on Its Map

In Stock's Debts the Disclosure

A New Jersey for Europe

Le Defense d'Assassins



Harvard Business Review

THE MAGAZINE BLOGS AUDIO & VIDEO BOOKS WEBINARS

Guest | limited access Register today and save 20%* off your first order! Details

THE MAGAZINE
October 2012

data Scientist: The Sexiest Job of the 21st Century

by Thomas H. Davenport and D.J. Patil

nature

Home News & Comment Research Careers & Jobs Current Issue Archive Audio & Video For

News & Comment News 2016 March Article

NATURE | NEWS

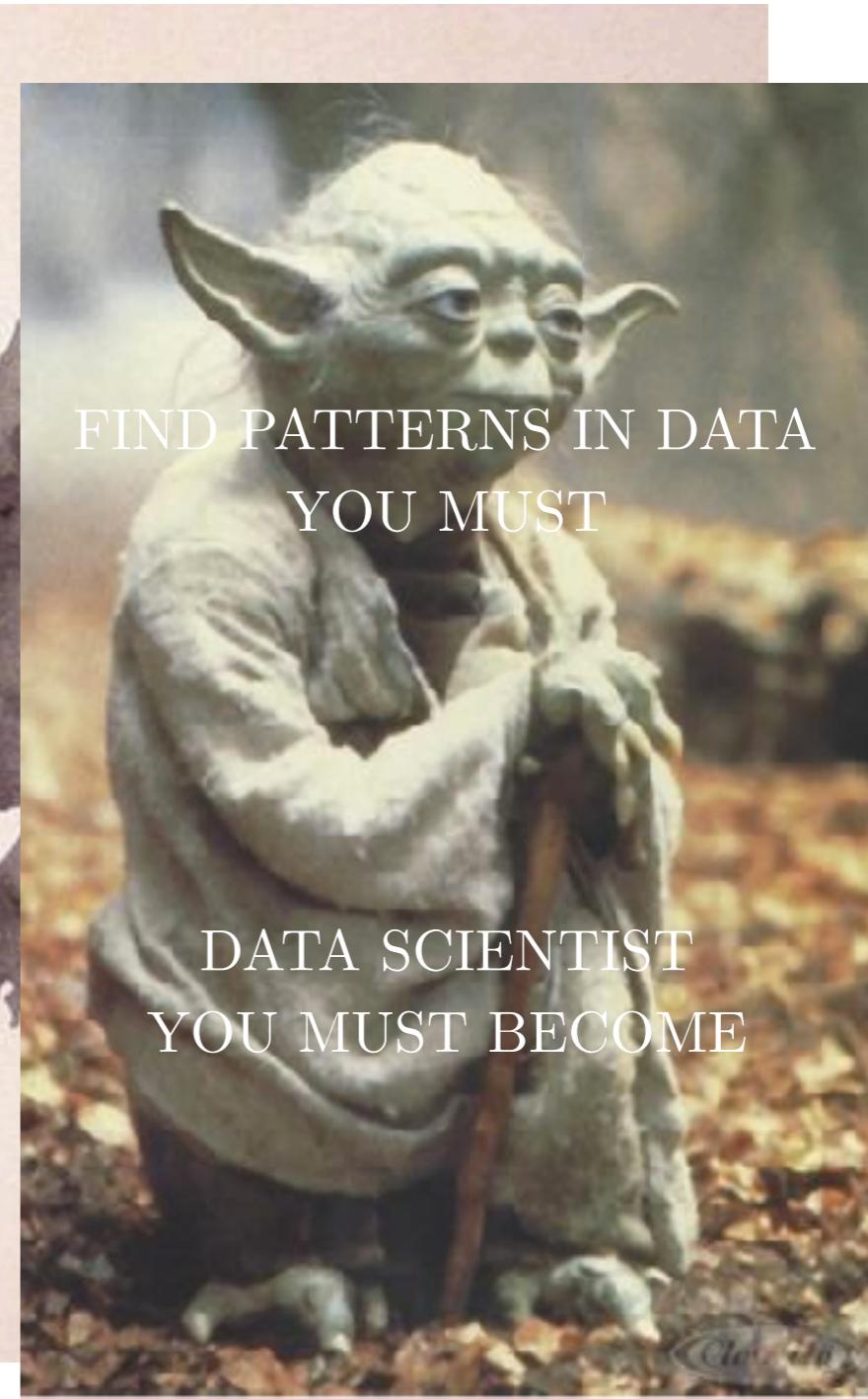
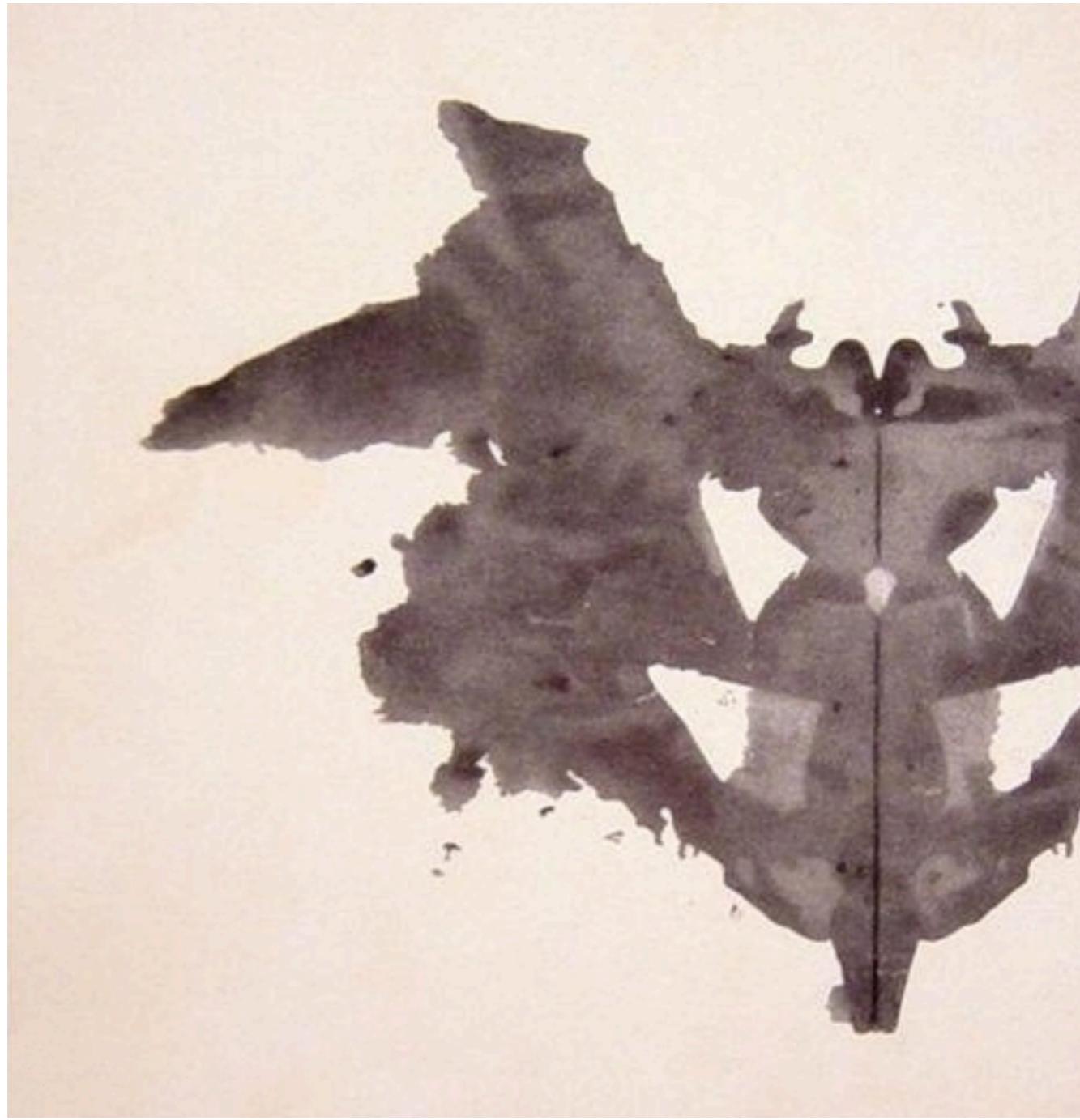
South Korea trumpets \$860-million AI fund after AlphaGo 'shock'

Historic win by Google DeepMind's Go-playing program has South Korean government playing catch-up on artificial intelligence.



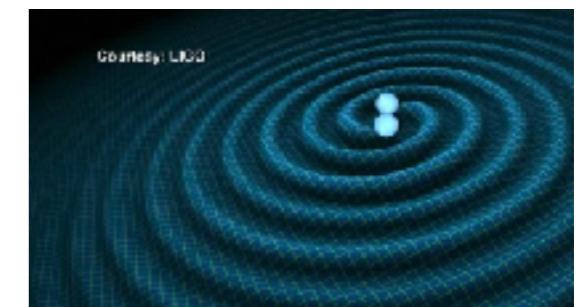
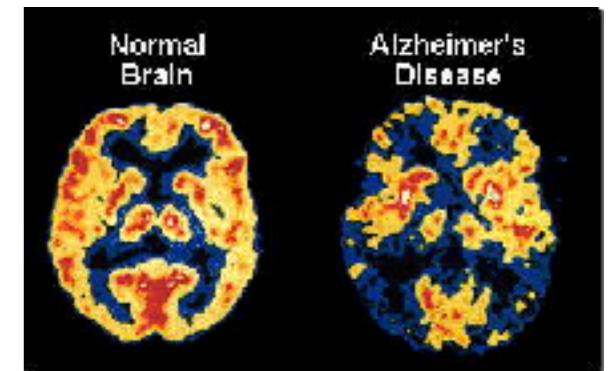
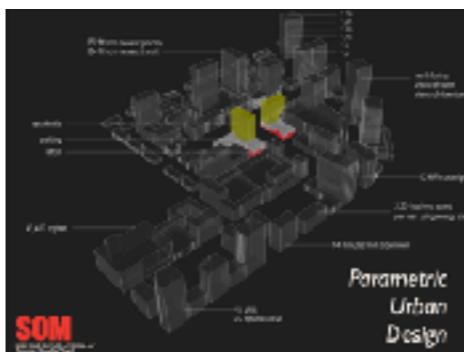
A Network Tour of Data Science, F
Prof. Pierre Vanderghenst
Prof. Pascal Frossard

What is data science?



Short answer

Science of transforming raw data into meaningful knowledge to provide smart decisions to real-world problems.



Longer answer



Q: What are the fields of Data Science?

Mathematical Modeling

Design algorithms that transform data into knowledge.
Use Linear algebra, optimization,

Computer Science
Scalable databases for storing, accessing data.
E.g. Cloud computing, Amazon EC2, Hadoop.
Distributed and parallel frameworks for data processing.
E.g. MapReduce, GraphLab.



Data

Collection of massive amounts of data at increasing rate.

E.g. Social networks, sensor networks, mobile devices, biological networks, administrative, economics data

Issues of privacy, security, ownership



Q: What are the main challenges?

Major challenges: Multidisciplinary integration, large-scale databases, scalable computational infrastructures, design math algorithms for massive datasets, trade-off speed and accuracy for real-time decisions, interactive visualization tools.



Q: What are the applications?

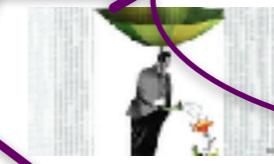
Personalized Services

E.g. Healthcare (enhanced diagnostics)
Commerce (products)



Knowledge Discovery

E.g. Physics, genomics, social sciences.



Data Science



Google

Domain Expertise Sciences

E.g. Economy, Biology, Physics, Neuroscience, sociology.

Government

E.g. Healthcare, Defense, Education, Transportation.

Industry

E.g. E-commerce, Telecommunications, Finance.



Intelligent Systems

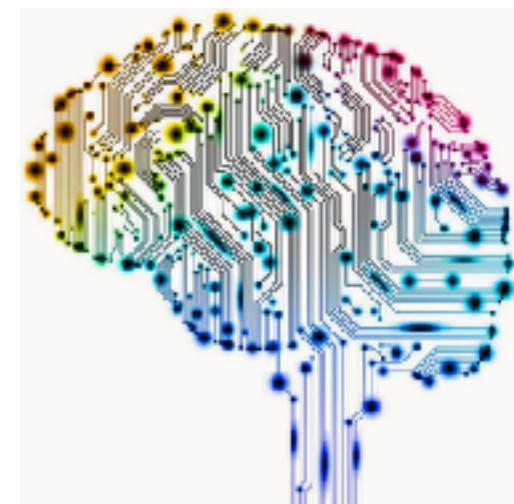
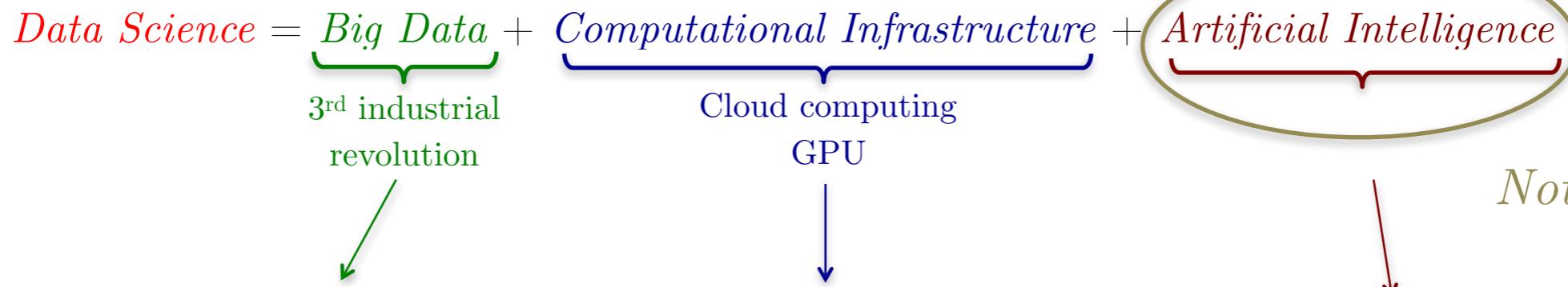
E.g. Autonomous cars, security, interactive tools for data organization and exploration.



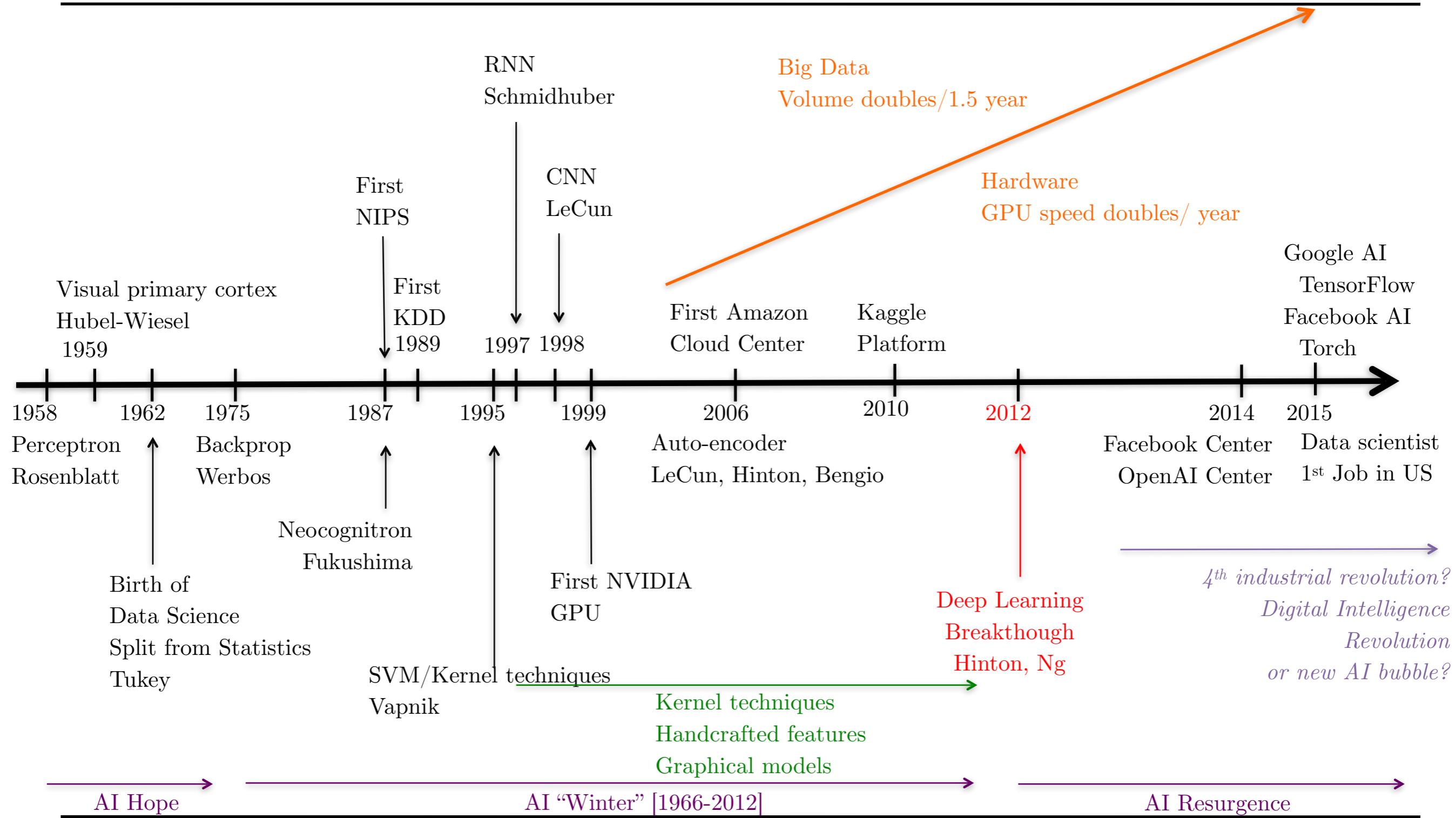
Sort of in-between...



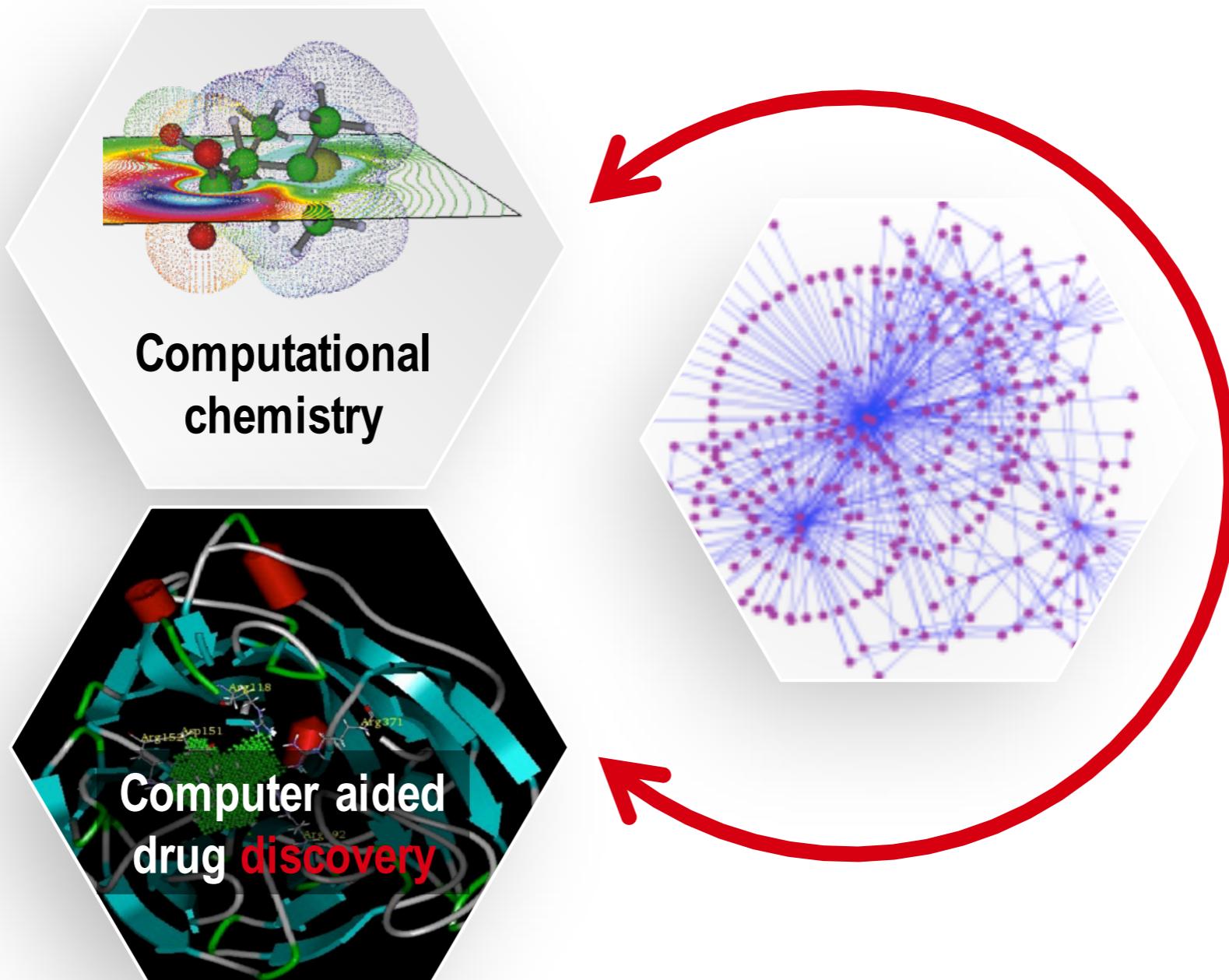
Q: What is big data?



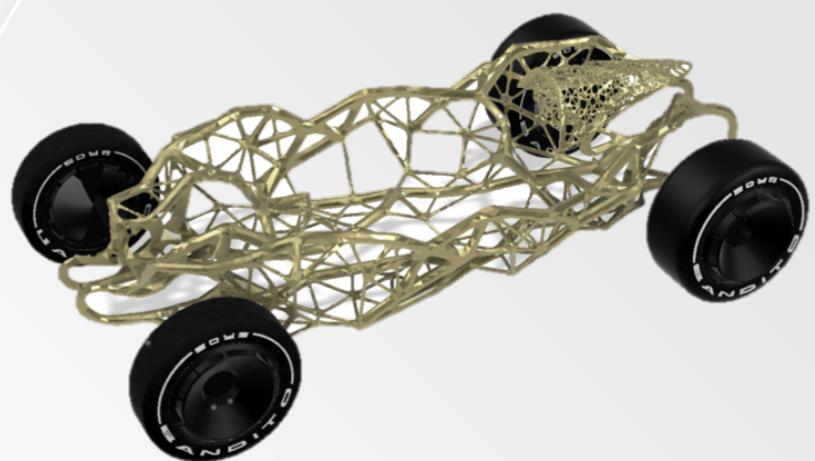
Historical perspective



A fourth industrial revolution ?

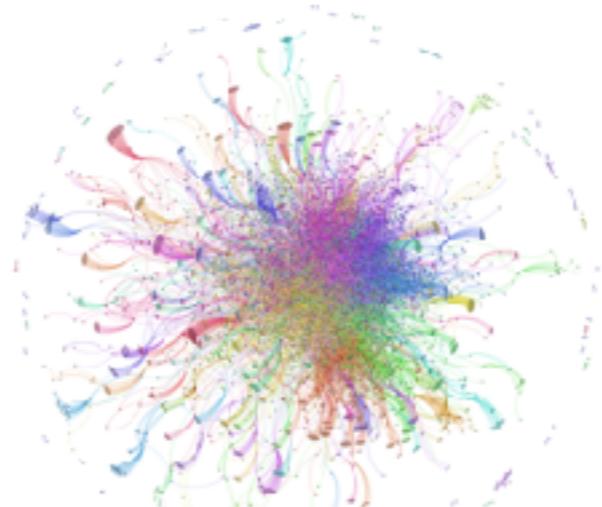


A fourth industrial revolution ?

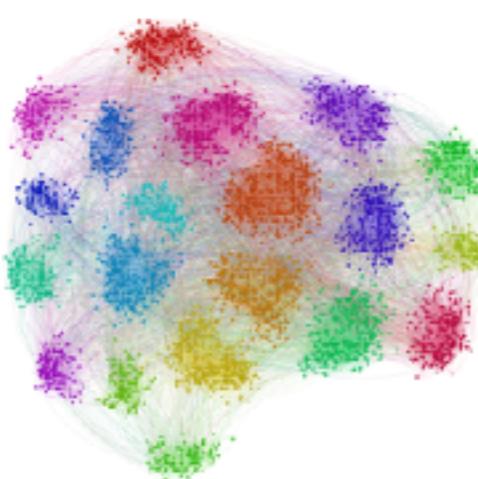


What about networks?

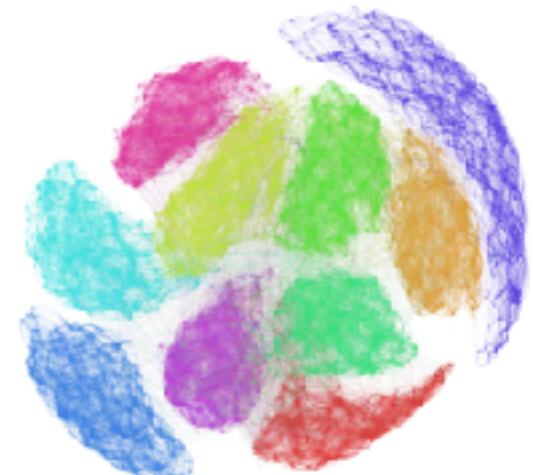
- Graphs encode complex data structures. They are everywhere: WWW, Facebook, Amazon, etc



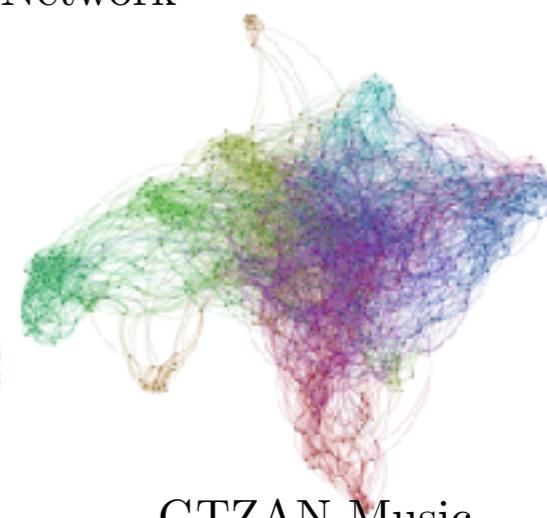
Graph of Google Query
“California”



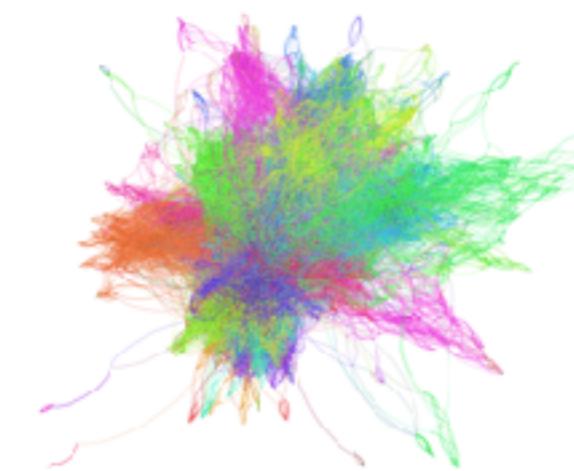
Social Network



MNIST Image
Network



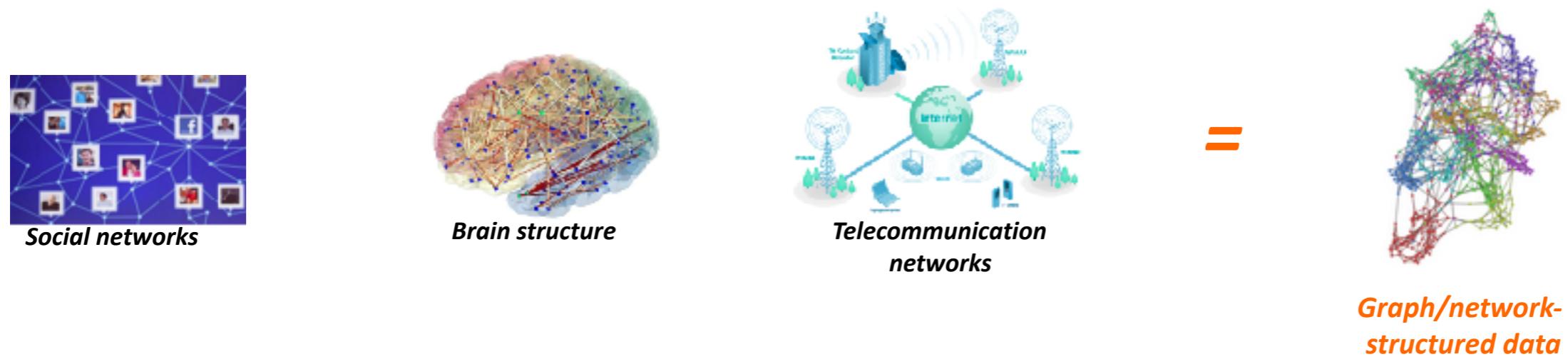
GTZAN Music
Network



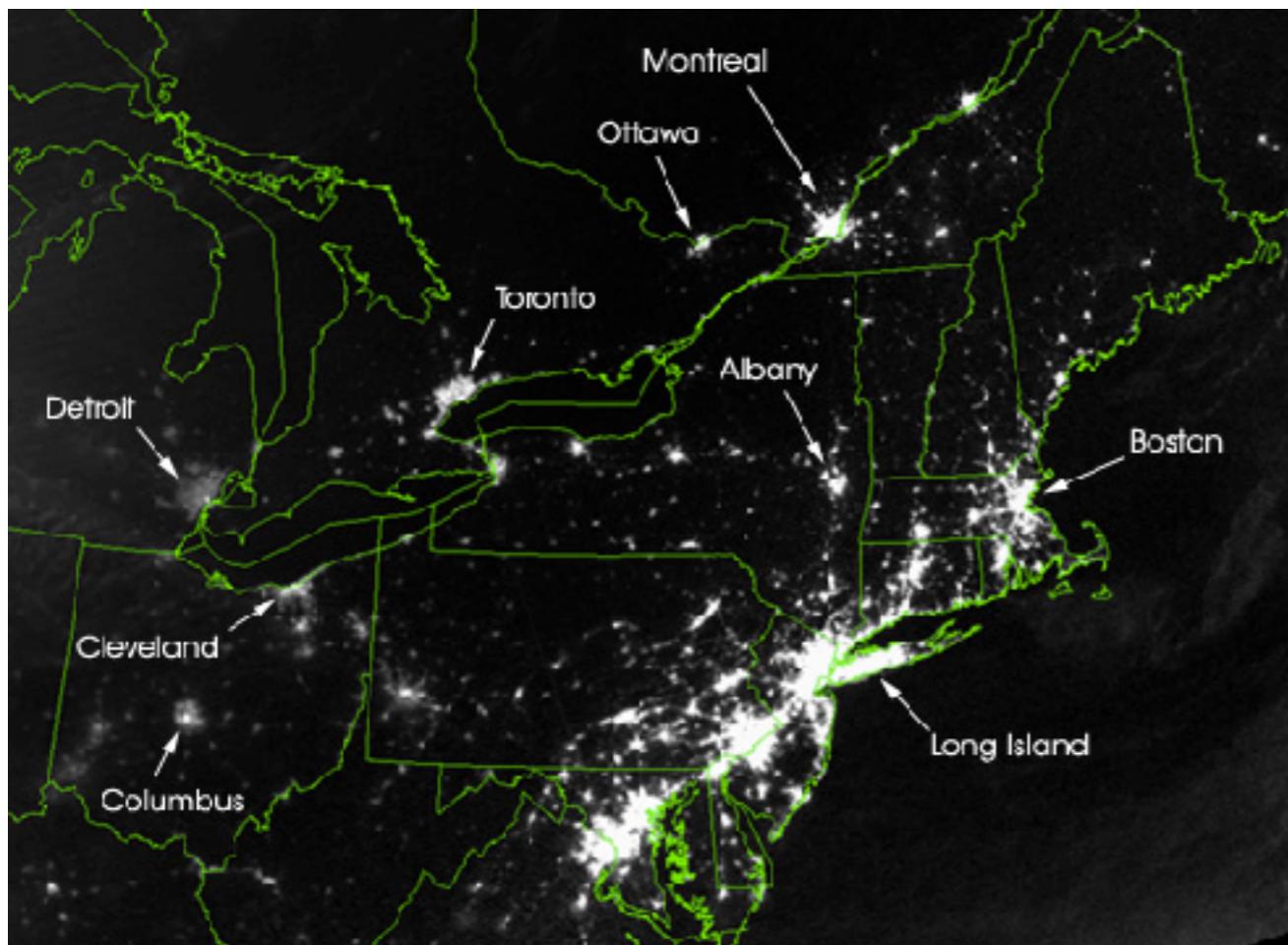
Network of Text Documents
20newsgroups

Why are networks important?

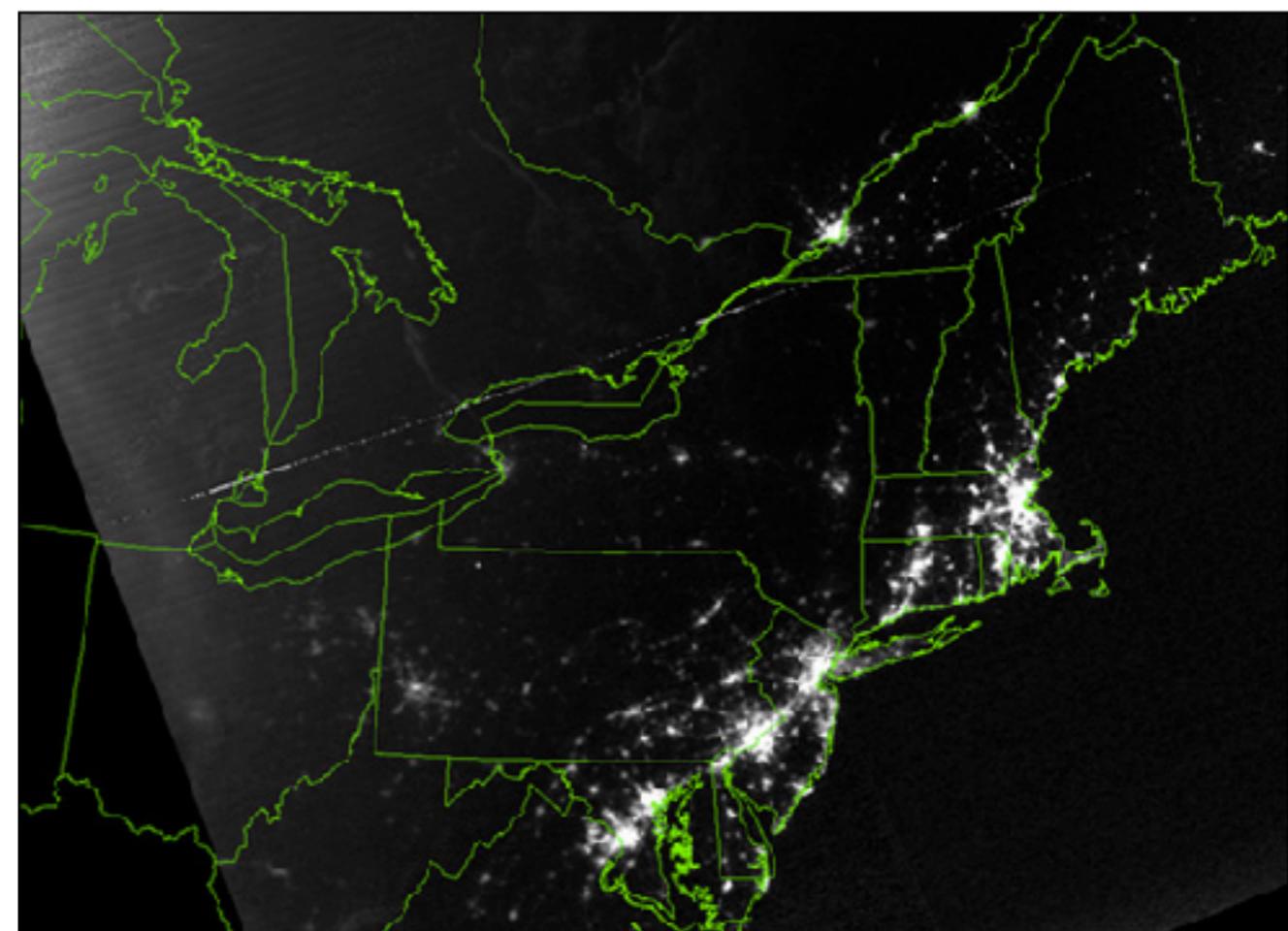
- Networks improve all data science tasks, for a small computational price!
- Essential data lie on networks:
 - (1) *Social networks* (Facebook, Twitter)
 - (2) *Biological networks* (genes, brain connectivity)
 - (3) *Communication networks* (Internet, wireless, traffic)



Blackout, Aug 15th, 2003



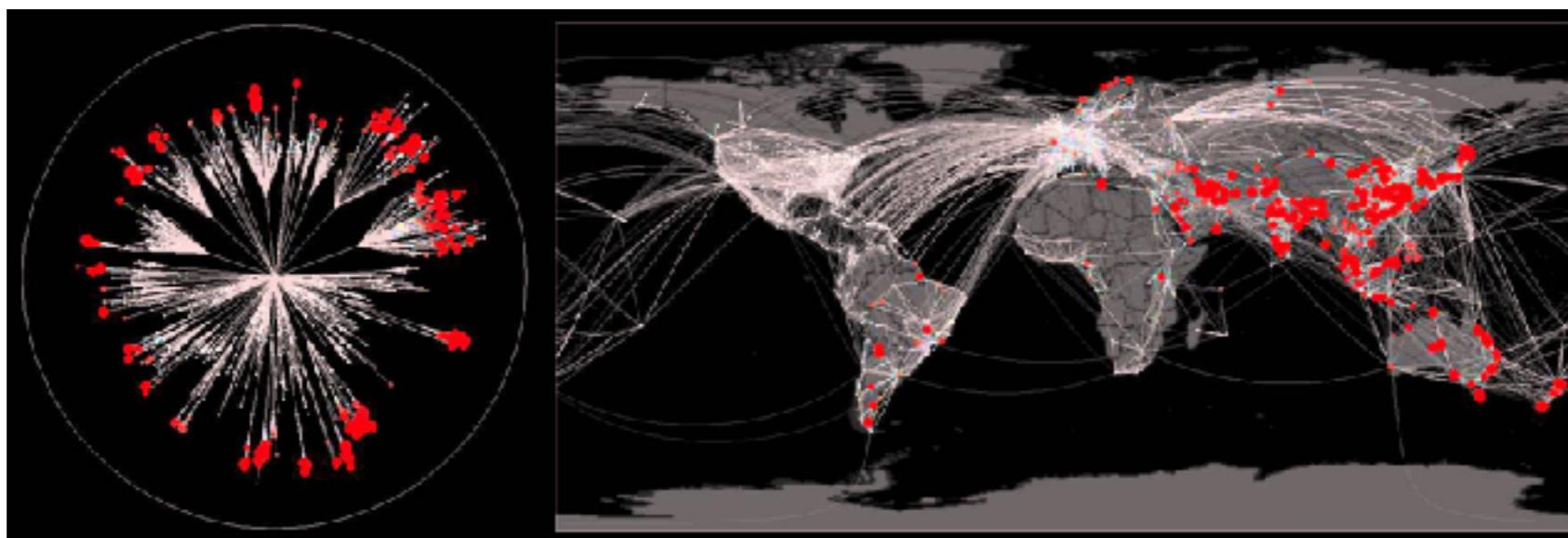
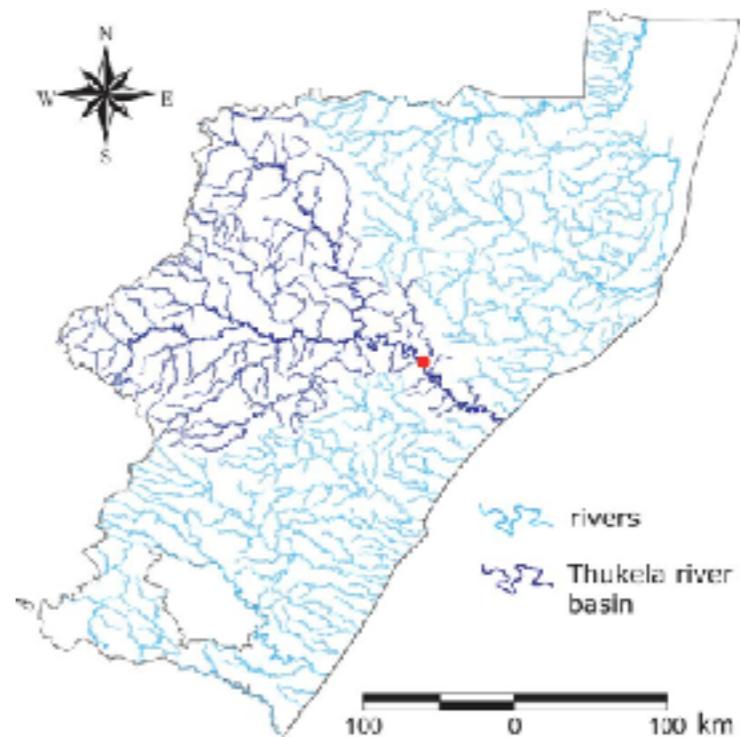
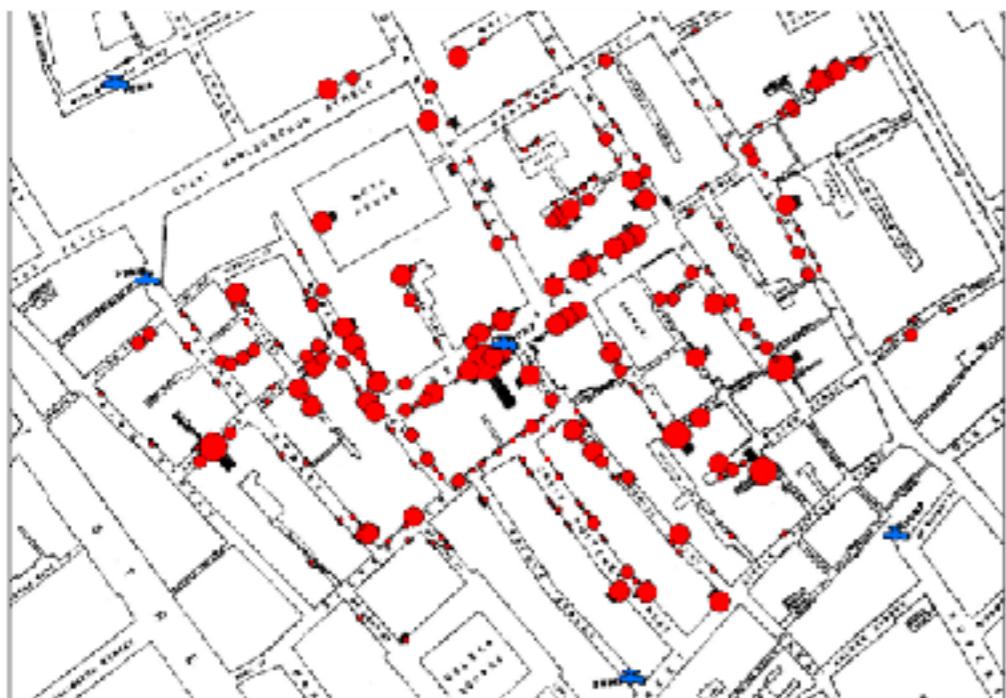
August 14, 2003: 9:29pm EDT
20 hours before



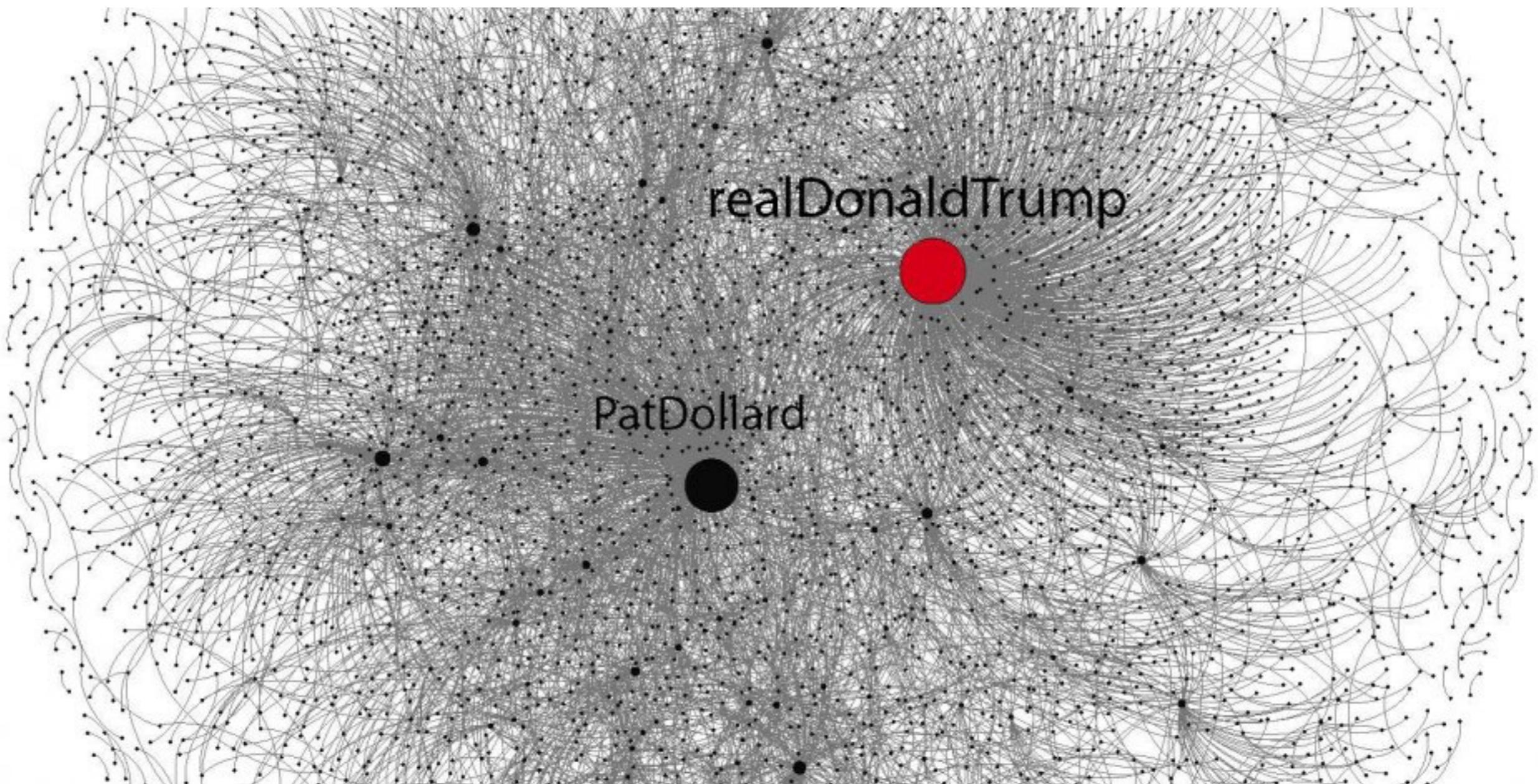
August 15, 2003: 9:14pm EDT
7 hours after

How does network structure influence behaviour and robustness of complex systems?

Diseases spread, then and now



Rumors spread, too

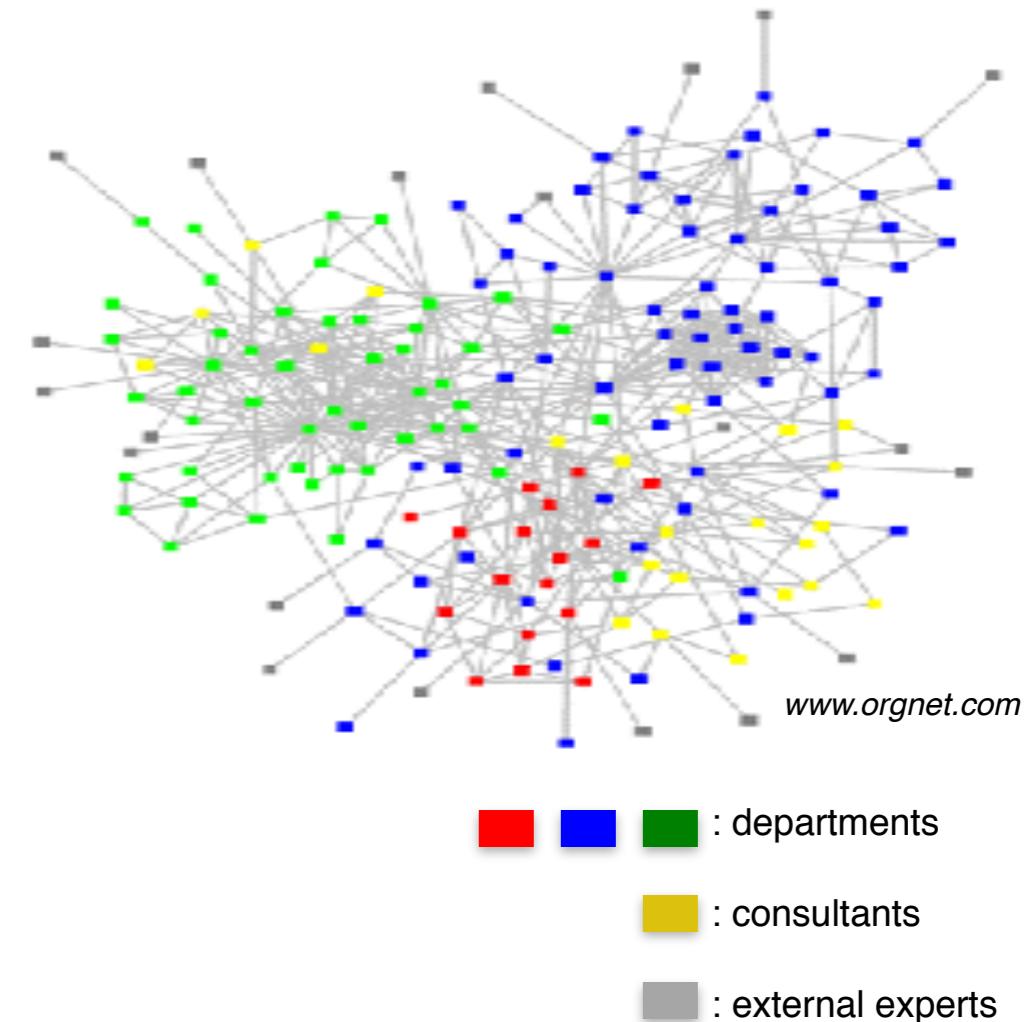


Complex systems and networks

- Behind each complex system there is a network, that defines the interactions between the components!

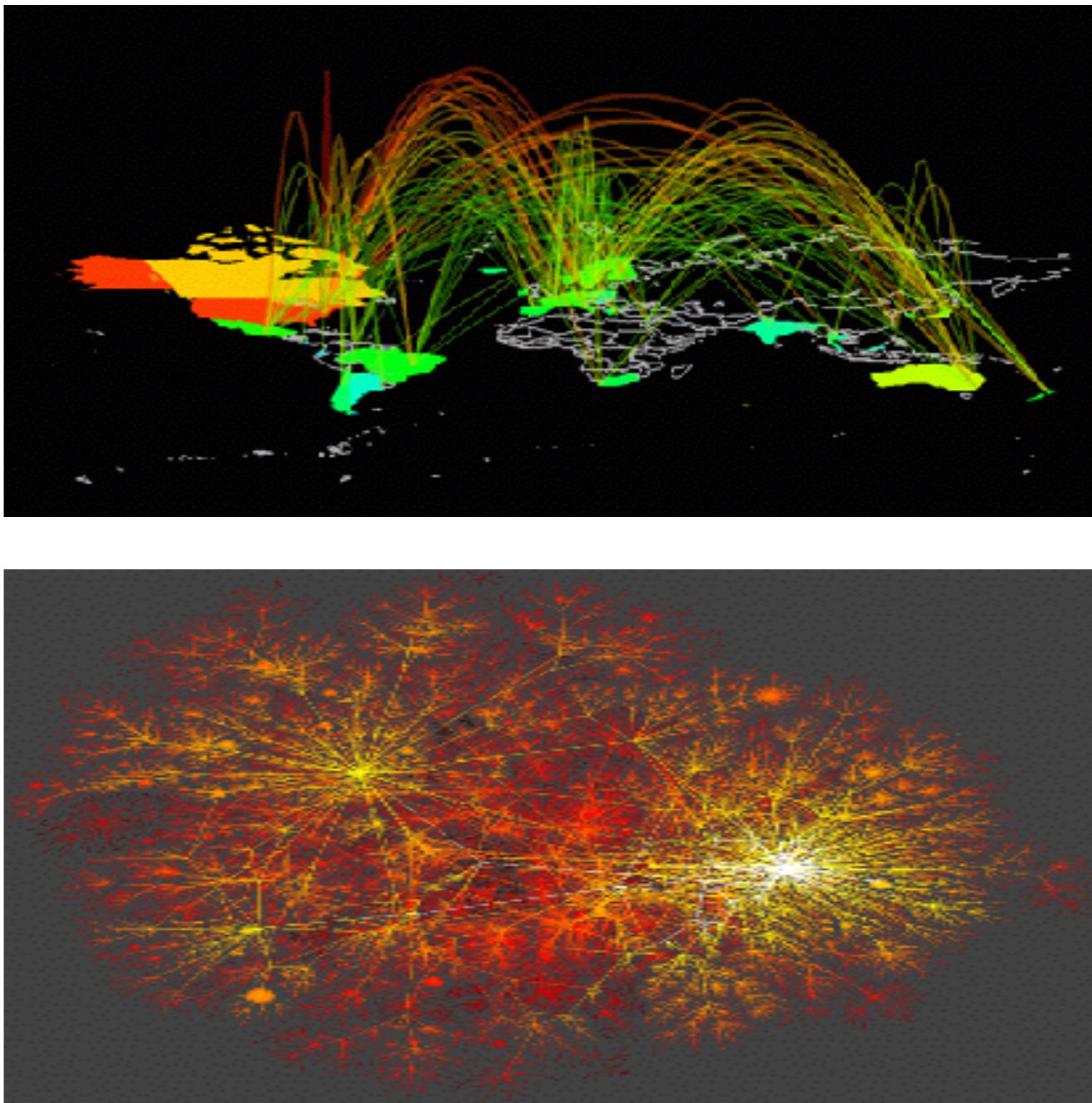
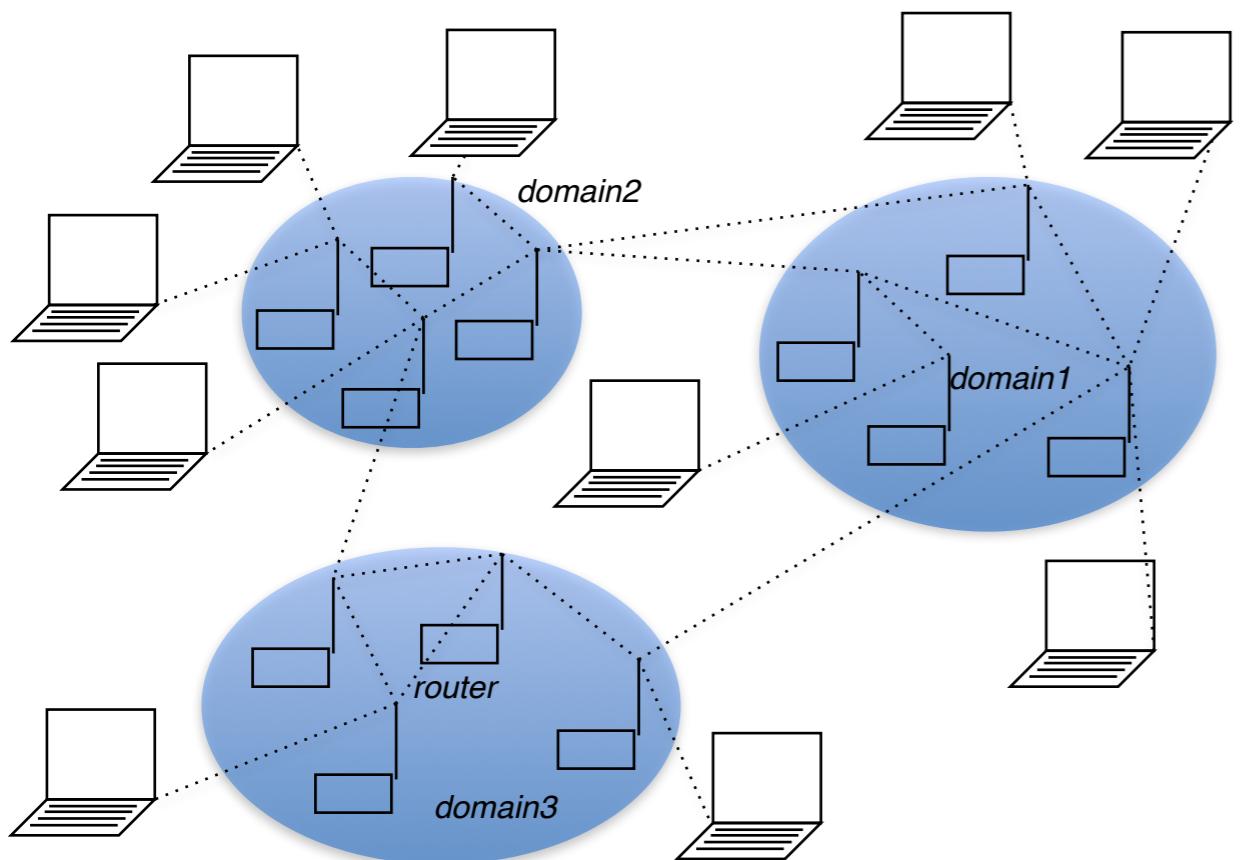


Keith Shepherd's "Sunday Best". <http://baseballart.com/2010/07/shades-of-greatness-a-story-that-needed-to-be-told/>

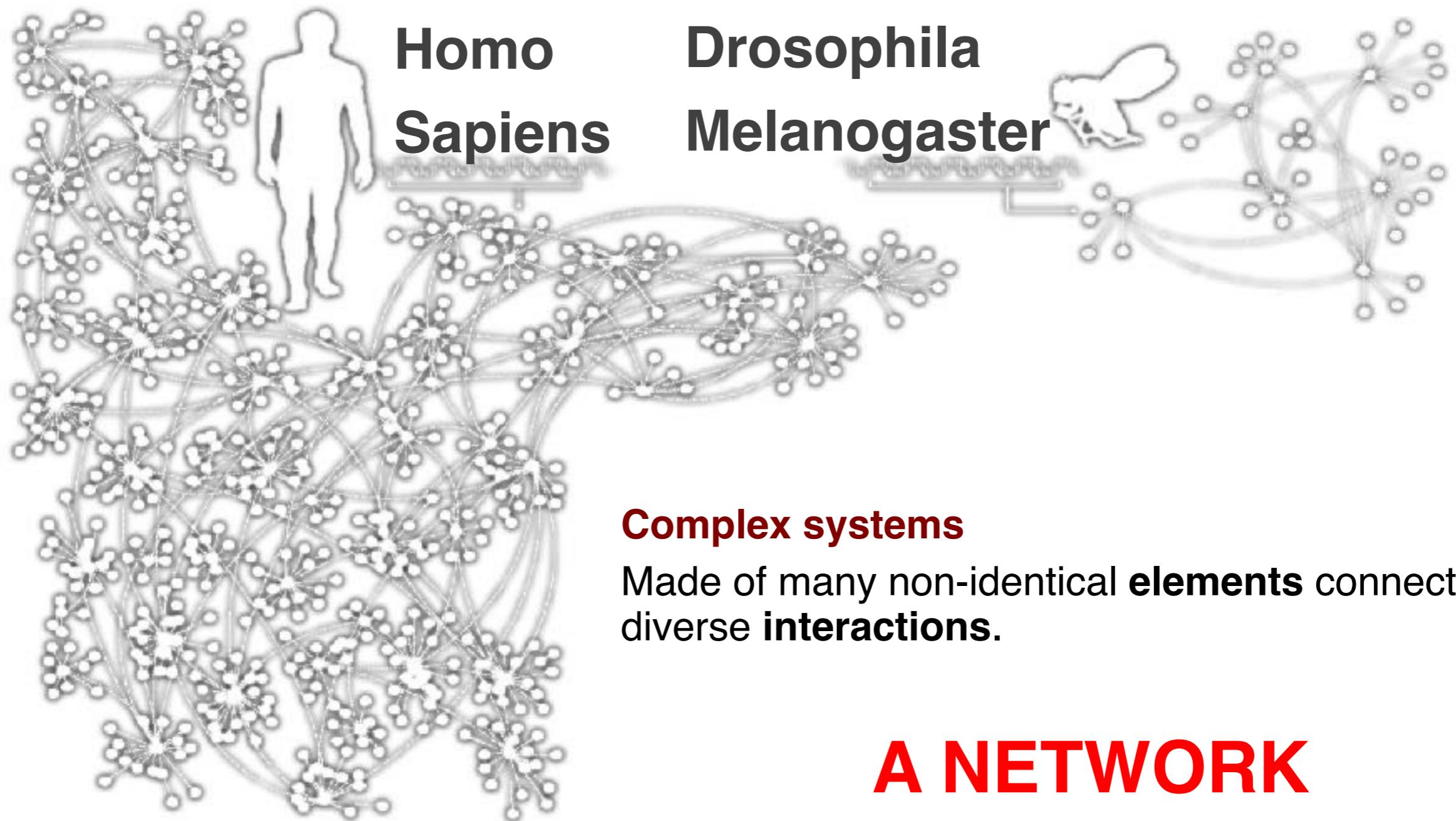


- We will never understand complex system unless we map out and understand the networks behind them!

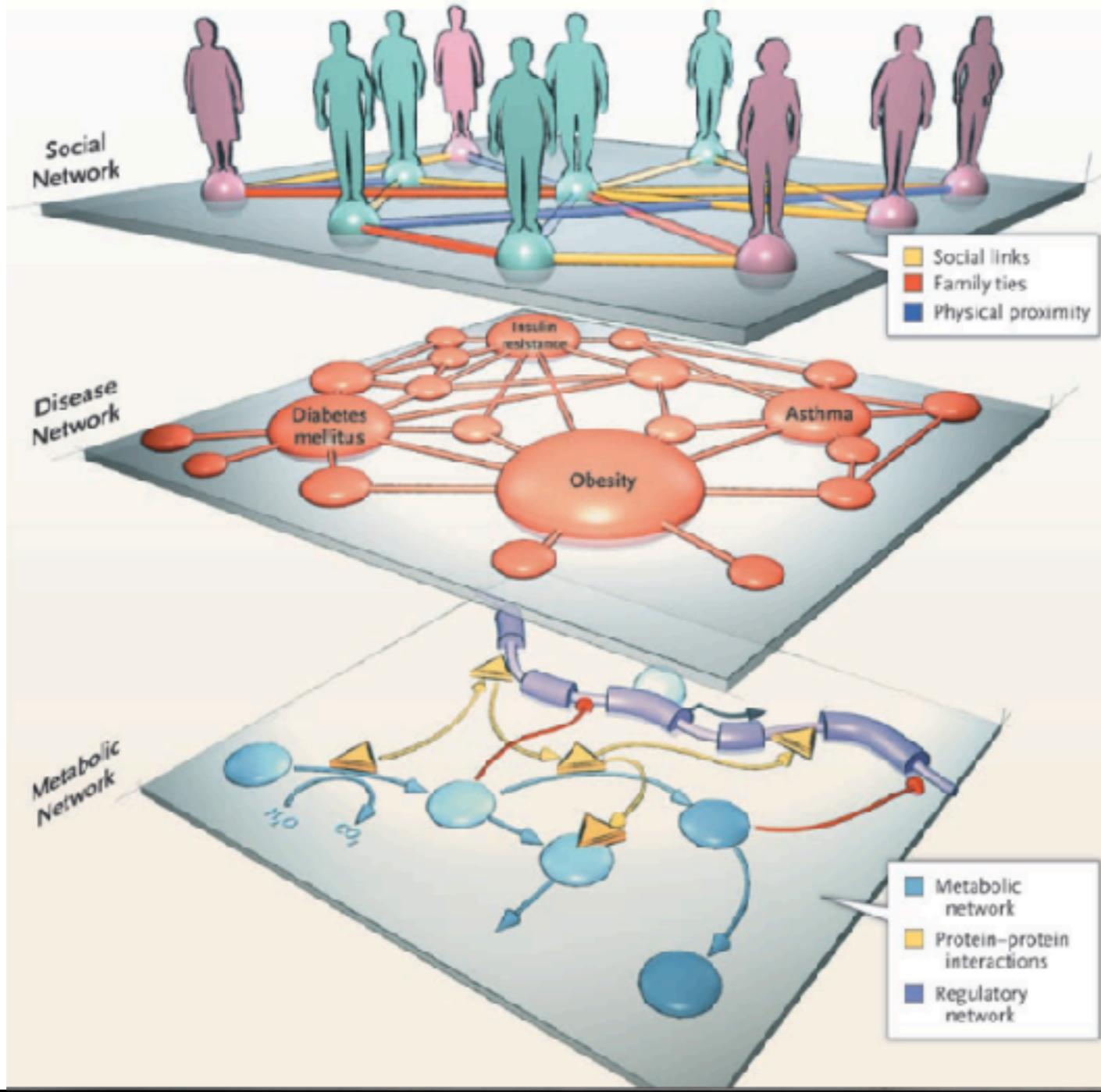
Communication networks: Internet



Biological networks: human genes



Model complex interactions



High economic impact



Licensed under CC Attribution-NonCommercial-ShareAlike 2.0 Germany | Ludwig Datzeke | <http://flickr.com/photos/ludwig-datzeke/>

Google

Market Cap(2010 Jan 1):
\$189 billion

Cisco Systems

networking gear Market cap
(Jan 1, 2019):
\$112 billion

Facebook market cap: *\$50 billion*

[www.bizjournals.com/austin/news/2010/11/15/
facebook... - Cached](http://www.bizjournals.com/austin/news/2010/11/15/facebook...-Cached)

The times they are a-changing

March 2006

ExxonMobil

362.5B



348.5B



Microsoft

279.0B



230.9B



225.9B

March 2016



621.8B

Alphabet

530.2B



Microsoft

444.9B

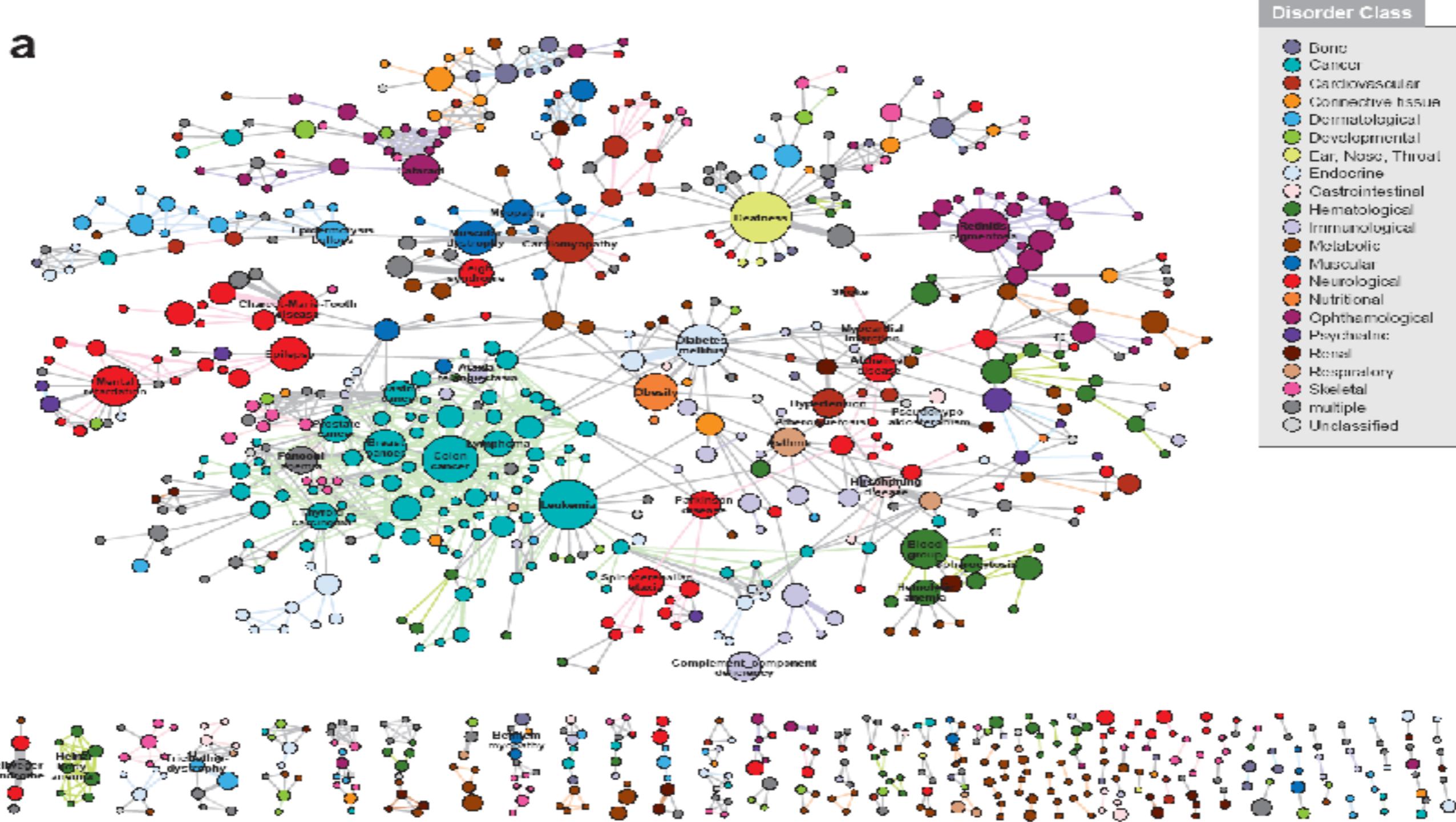
ExxonMobil

352.2B

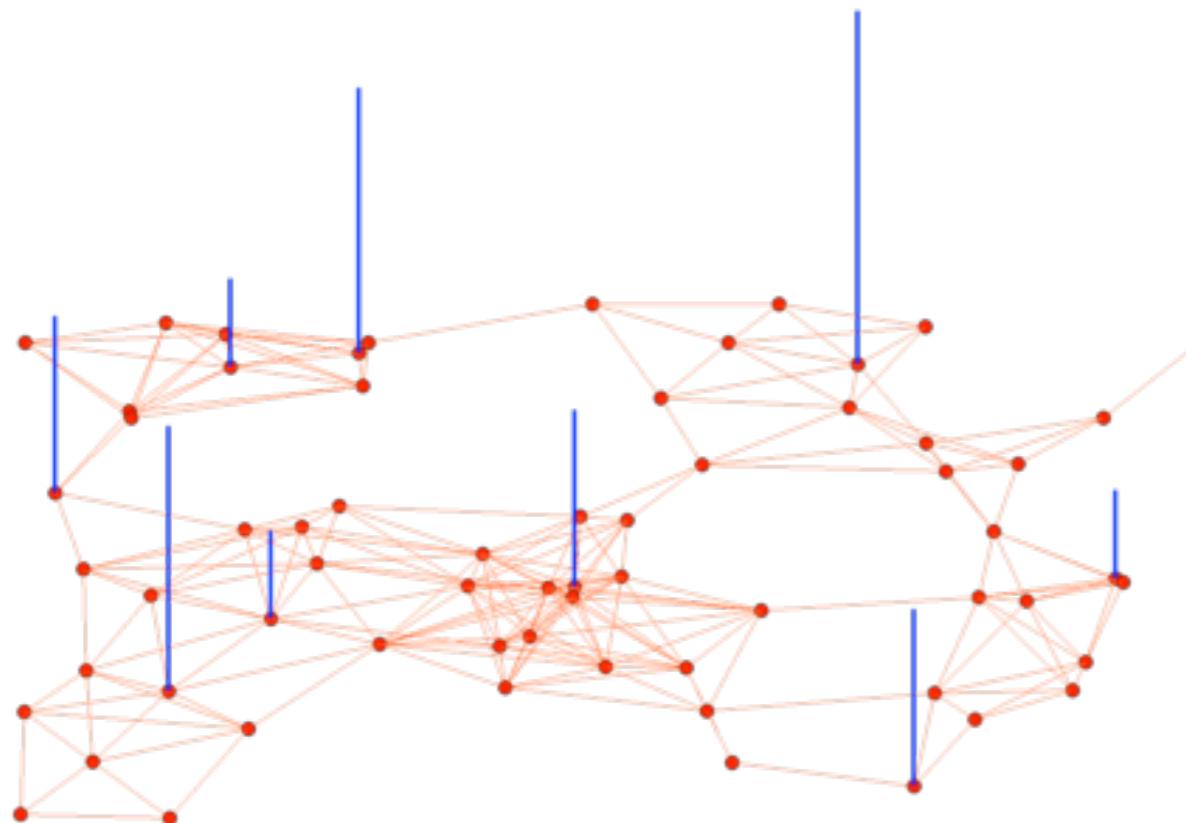


333.5B

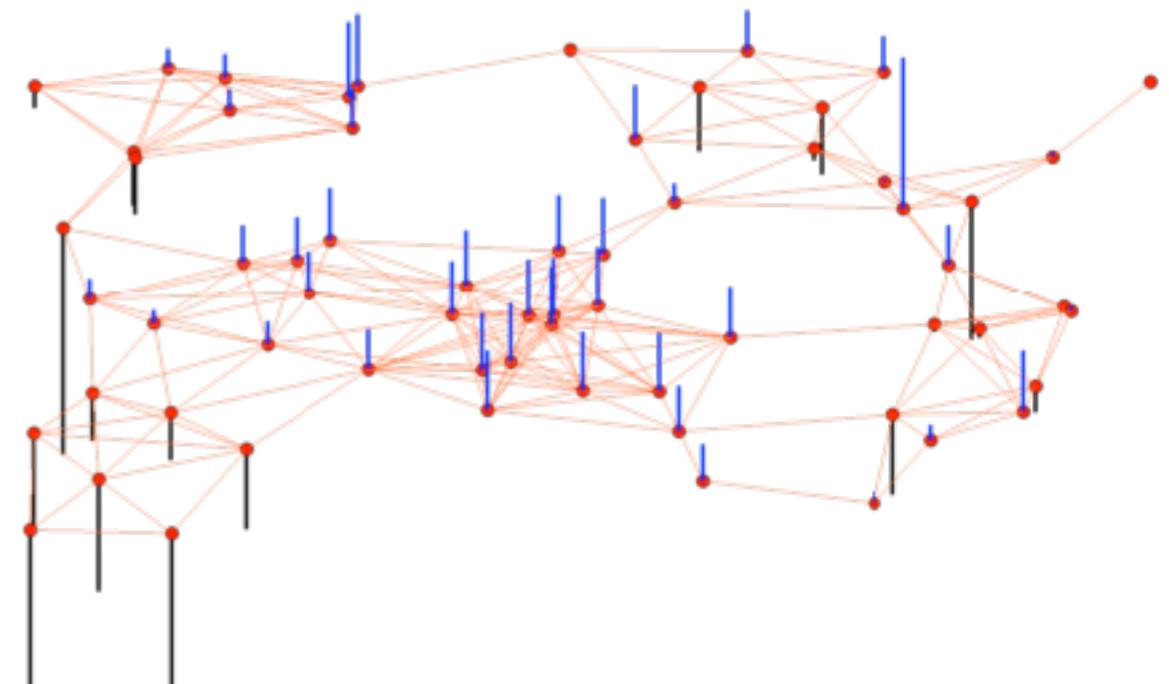
High human impact



New trend: info over networks



Semi-Supervised Learning



Analysis / Information Extraction

Properties of network data science

Interdisciplinary

Empirical

Quantitative and Mathematical

Computational