Complex Networks I
Structural properties and models

M. Ángeles Serrano
ICREA Research Professor
Dpt. de Física de la Matèria Condensada
UB Institute of Complex Systems UBICS
Universitat de Barcelona
Complexity Science Hub Vienna CSH

Màster de Modelització Computacional Atomística i Multiescala en Física, Química i Bioquímica
Màster de Física dels Sistemes Complexos i Biofísica

COMPLEX NETWORKS

Preface

Motivation

Two blocks

Complex networks I. Structural properties and network models

Complex networks II. Percolation and dynamical processes

Extra bonus: Network geometry

Evaluation

- Report of assignments in TFG format + codes

  content and presentation will be evaluated

COMPLEX NETWORKS

CALENDAR

	22 Proves d'avaluació	23 Sant Jordi	24 Proves d'avaluació	25 Reavaluació: FENE	26 Reavaluació: FESCiB	10:00h -12:00h 15:00h - 17:00h 17:00h - 19:00h
	29 F. S. Econòmics i Socials Matèria Tova	30 Xa MAS kes	1 Festiu	2 Neurociència	3	15:00h - 17:00h 17:00h - 19:00h
MAIG	6 F. S. Econòmics i Socials Matèria Tova	7 Xa MAS kes	8 F. S. Econòmics i Socials Matèria Tova	9 Neurociència	10 Xa MAS kes	15:00h - 17:00h 17:00h - 19:00h
	13 F. S. Econòmics i Socials Matèria Tova	14 Xa MAS kes	15 F. S. Econòmics i Socials Matèria Tova	16 Neurociència	17 Xa MAS kes	15:00h - 17:00h 17:00h - 19:00h
	20 Festiu	21 Xa MB kes	22 F. S. Econòmics i Socials Matèria Tova	23 Neurociència	24 Xa MB kes	15:00h - 17:00h 17:00h - 19:00h
	27 F. S. Econòmics i Socials Matèria Tova	28 Xa MB kes	29 F. S. Econòmics i Socials Matèria Tova	30 Neurociència	31 Xa MB kes	15:00h - 17:00h 17:00h - 19:00h
JUNY	3 F. S. Econòmics i Socials Matèria Tova	4	5	6	7	15:00h - 17:00h 17:00h - 19:00h
	10 F. S. Econòmics i Socials Matèria Tova	11 Xa MB es	12 F. S. Econòmics i Socials Matèria Tova	13 Neurociència	14 Xa MB kes	15:00h - 17:00h 17:00h - 19:00h
	17 F. S. Econòmics i Socials Matèria Tova	18 Xa MAS es	19 F. S. Econòmics i Socials Matèria Tova	20	21 Xa MAS kes	15:00h - 17:00h 17:00h - 19:00h
	24 Festiu	25 Xa MB kes	26 Matèria Tova	27 Inici proves d'avluació	28 Fi proves d'avaluació	15:00h - 17:00h 17:00h - 19:00h

REFERENCES

Online

Curs introductori a les xarxes complexes,
Prof. M. Boguñá http://complex.ffn.ub.es/~mbogunya/videos.php

Books

[B1] *A first course in network science,* F. Menczer, S. Fortunato and Clayton A. Davis, Cambridge University Press (2020); *Network Science,* A.-L Barabási with Márton Pósfai, Cambridge University Press (2016); *Networks: An Introduction,* M. E. J. Newman, Oxford University Press (2010)

[B2] *Dynamical Processes on Complex Networks,* A. Barrat, M. Barthélémy, and A. Vespignani, Cambridge University Press (2012)

Reviews

[R1] *The shortest path to network geometry: A Practical Guide to Basic Models and Applications,* M. Á. Serrano & M. Boguñá Elements in Structure and Dynamics of Complex Networks, Cambridge: Cambridge University Press (2022)

[R2] *Critical phenomena in complex networks*, S. N. Dorogovtsev, A. V. Goltsev, and J. F. F. Mendes, *Rev. Mod. Phys.* **80**, 1275 (2008)

Popular science

[P1] J. Ladyman, K. Wiesner 2020, *What is a complex system?* Yale University Press

[P2] A.-L. Barabási 2014, *Linked: How Everything Is Connected to Everything Else and What It Means,* Basic Books

[P3] D. J. Watts 1999, *Small Worlds: The Dynamics of Networks between Order and Randomness,* Princeton University Press

[P4] R. Solé 2009, *Redes Complejas. Del genoma a Internet,* Tusquets Editores

MOTIVATION

COMPLEX

- From dictionary.com:

! com·plex 1. composed of many interconnected parts; compound; composite: a complex highway system. 2. characterized by a very complicated or involved arrangement of parts, units, etc.: *complex machinery*. 3. so complicated or intricate as to be hard to understand or deal with: *a complex problem*.

! com·pli·cate –verb (used with object) 1. to make complex, intricate, involved, or difficult: *His recovery from the operation was complicated by an allergic reaction.* –adjective 2. complex; involved.

COMPLEX

- "**The wiring on an aircraft is complicated**. To figure out where everything goes would take a long time. But if you studied it for long enough, you could know with (near) certainty what each electrical circuit does and how to control it. The system is ultimately knowable. If understanding it is important, the effort to study it and make a detailed diagram of it would be worthwhile.

So **complicated** = not simple, but ultimately knowable from the study of the individual parts.

- Now, **put a crew and passengers in that aircraft** and try to figure out what will happen on the flight. Suddenly we go from complicated to **complex**. You could study the lives of all these people for years, but you could never know all there is to know about how they will interact. You could make some guesses, but you can never know for sure. And the effort to study all the elements in more and more detail will never give you that certainty.

So **complex** = not simple and never fully knowable from the study of the individual parts. Just too many variables interact.

Examples of Complex Systems

- human societies
- cells
- organisms
- brain
- humans
- ant colonies
- ecosystems
- animal societies
- disease ecologies
- social insects
- geophysical systems
- the world wide web
- the Internet
- power grids, road networks
- economic systems
- weather systems
- modern physics: quantum solid state physics, complex fluids

WHAT IS A COMPLEX SYSTEM?

Conditions + Emergent features

- **Numerosity;
many interactions among many components**
- **No central control**
- **Non-equilibrium; open and driven**
- **Diversity/Heterogeneity**
- **Feedback**
- **Long range interactions**

- **Self-organization;
spontaneous patterns and order**
- **Structures between complete order and disorder**
- **Nested structure and modularity**
- **Robustness/Vulnerability**
- **Nonlinearity**
- **History and memory**
- **Adaptive behavior**

Emergence

The Wikipedia on Emergence

“In philosophy, systems theory and science, emergence is the way complex systems and patterns arise out of a multiplicity of relatively simple interactions. Emergence is central to the theory of complex systems.”

The concept is rather old

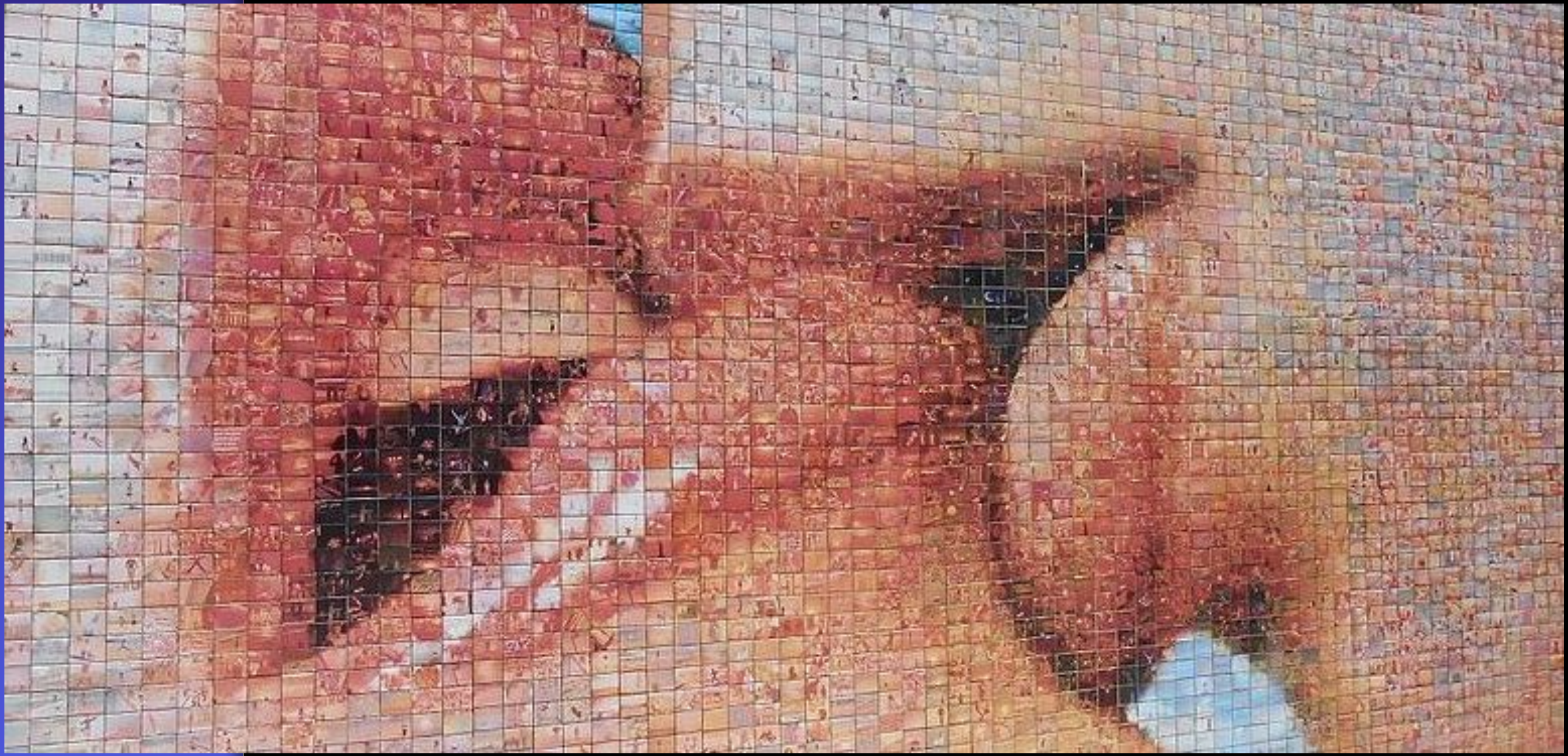
“The whole is more than the sum of its parts”
Lao Tse, Tao Te Ching, VI BC; Aristotle, Metaphysics, IV BC

But the term was coined for the first time by the pioneer psychologist **G. H. Lewes**
Lewes, G. H. (1875), *Problems of Life and Mind (First Series)* **2**, London: Trübner, ISBN 1-4255-5578-0

There's no love in a carbon atom, No hurricane in a water molecule, No financial collapse in a dollar bill.
-- Peter Dodds

Complex Systems

Complex Networks



Photomosaic

El món neix en cada besada
Joan Fontcuberta y Antoni Cumella, 2014
Plaça Isidre Nonell, Ciutat Vella, Barcelona

Two types of emergence

Weak emergence We can say that a high-level phenomenon is *weakly* emergent with respect to a low-level domain when the high-level phenomenon arises from the low-level domain, but truths concerning that phenomenon are *unexpected* given the principles governing the low-level domain. No new fundamental laws or properties are needed: everything will still be a consequence of physics but may still require the introduction of further levels of explanation above the physical level in order to make these phenomena maximally comprehensible to us.

Strong emergence We can say that a high-level phenomenon is *strongly emergent* with respect to a low-level domain when the high-level phenomenon arises from the low-level domain, but truths concerning that phenomenon are *not deducible* even in principle from truths in the low-level domain strong emergence. If it exists, has radical consequences. Are there strongly emergent phenomena? Consciousness? Fundamental physical laws need to be supplemented with further fundamental laws to ground the connection between low-level properties and high-level properties.

Philosophers thought already a lot about that......

Strong and weak emergence, David J. Chalmers , Philosophy Program , Research School of Social Sciences . Australian National University

Emergence Weak or Strong?

P. W. Anderson in *More is Different*

"The ability to reduce everything to simple fundamental laws does not imply the ability to start from those laws and reconstruct the universe... The constructionist hypothesis breaks down when confronted with the twin difficulties of scale and complexity... At each level of complexity entirely new properties appear... Psychology is not applied biology, nor is biology applied chemistry... We can now see that the whole becomes not merely more, but very different from the sum of its parts."

Mark A. Bedau in *Weak Emergence*

"Although strong emergence is logically possible, it is uncomfortably like magic. How does an irreducible but supervenient downward causal power arise, since by definition it cannot be due to the aggregation of the micro-level potentialities? Such causal powers would be quite unlike anything within our scientific ken. This not only indicates how they will discomfort reasonable forms of materialism. Their mysteriousness will only heighten the traditional worry that emergence entails illegitimately getting something from nothing."

In short:

- The concepts of Complexity and Emergence are not even well defined
- There is not a unified Theory of Complex Systems

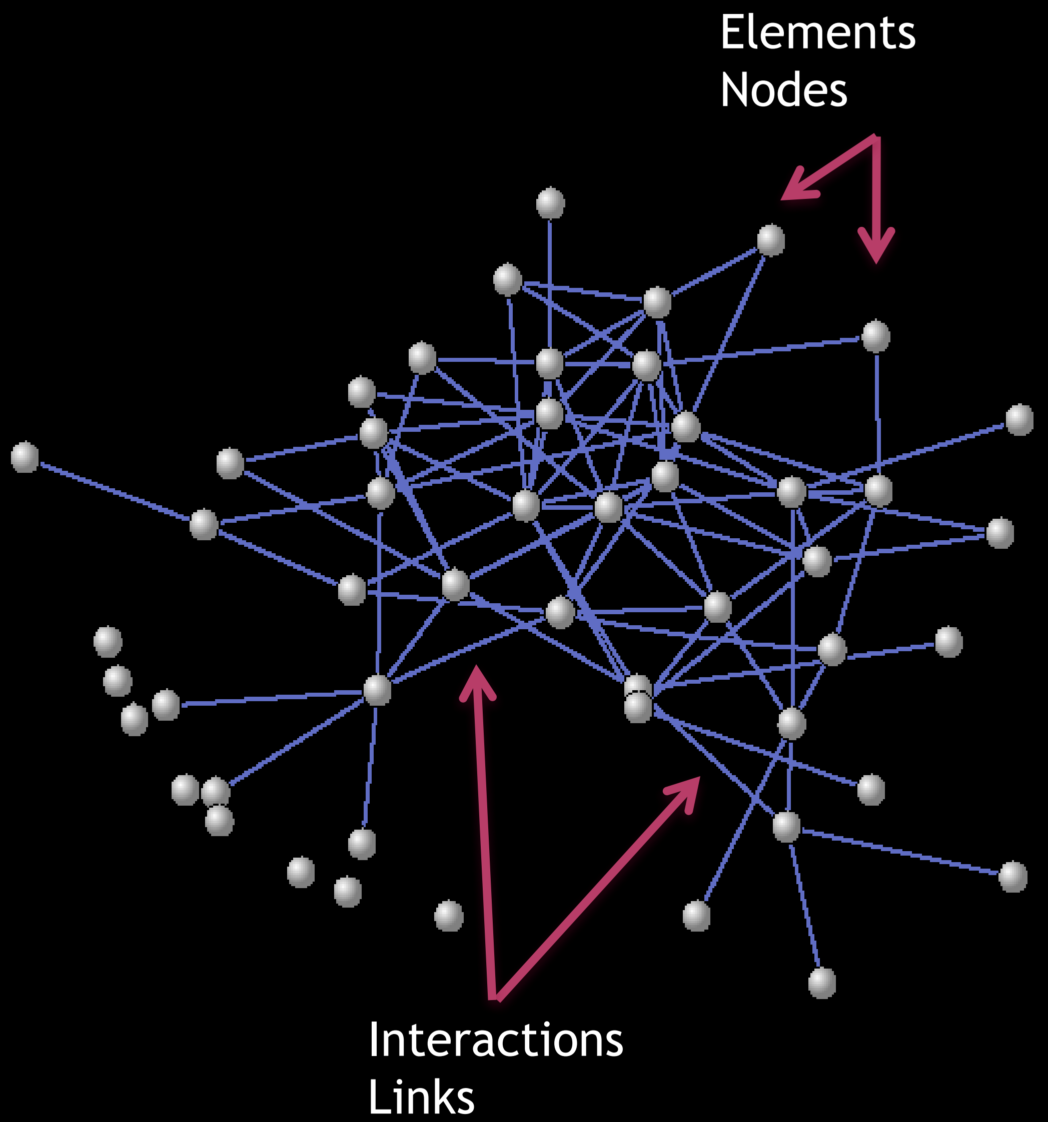
And yet, there are multitude of Complex Systems around...so let's perseverate.

Complex systems

Frameworks to study complex systems

Information Theory
is the mathematical study of the quantification and communication of information.

Network Science



Complex systems

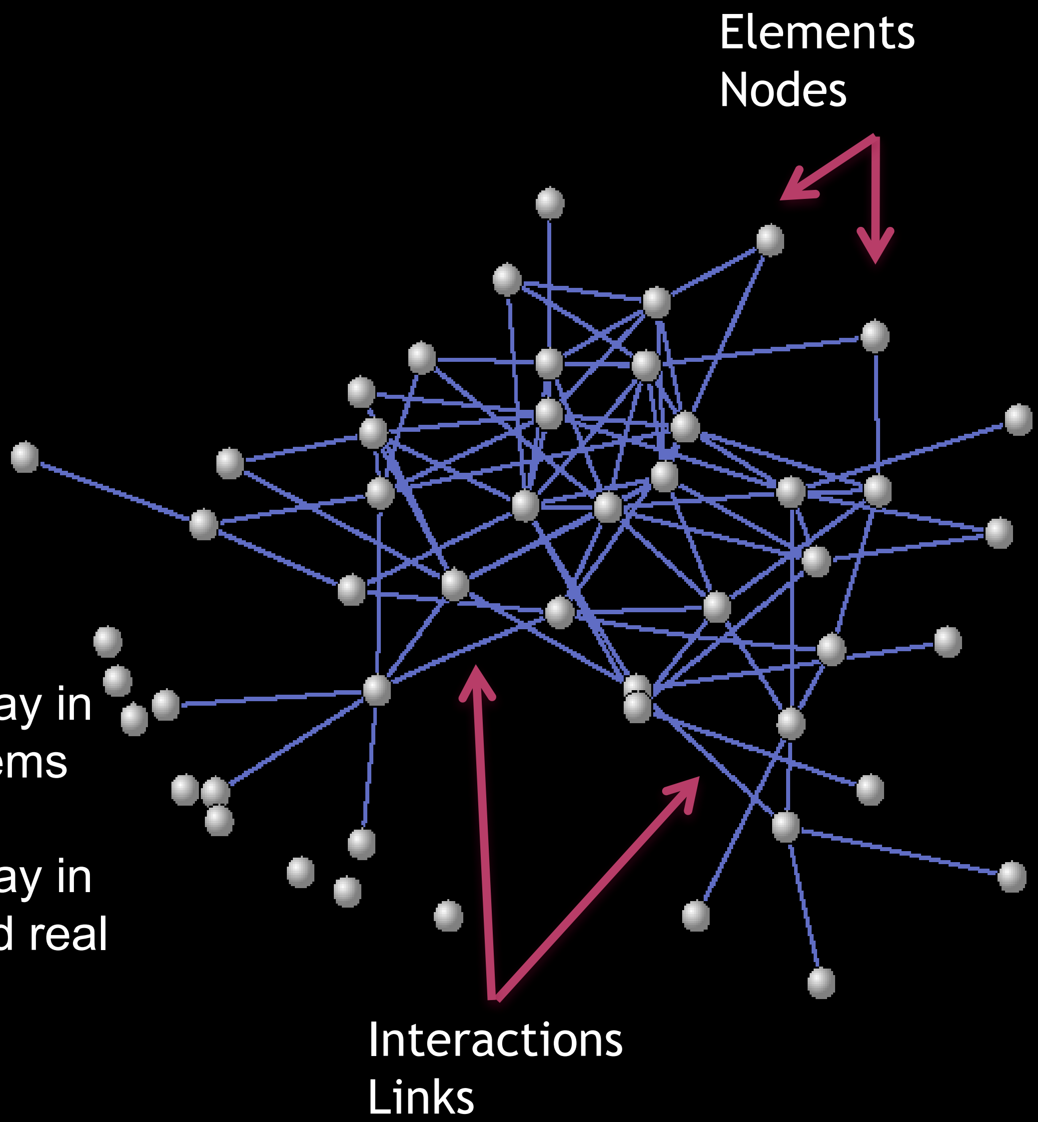
Frameworks to study complex systems

Networks
A change of paradigm

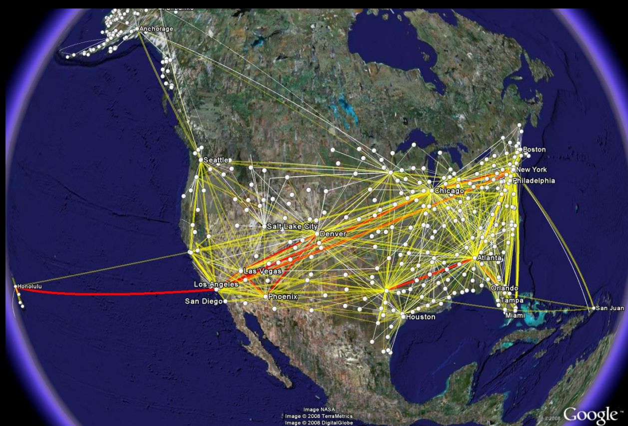
- Networks are everywhere

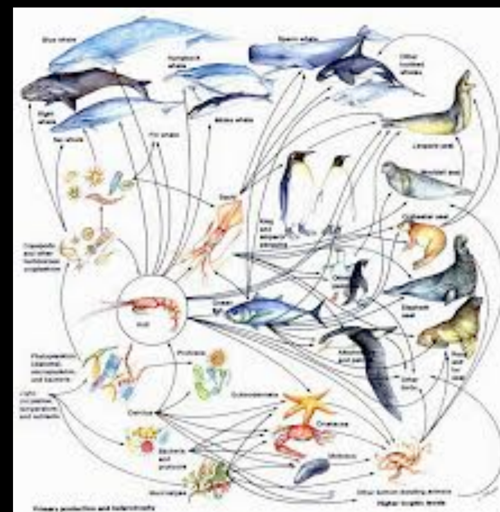
- Networks are changing the way in which we solve complex problems

- Networks are changing the way in which we study and understand real complex systems



twitter
facebook





· Internet · Airports ·
· International Trade ·
· Food Webs ·
· Online/Offline Social ·
· Brain · Cells ·

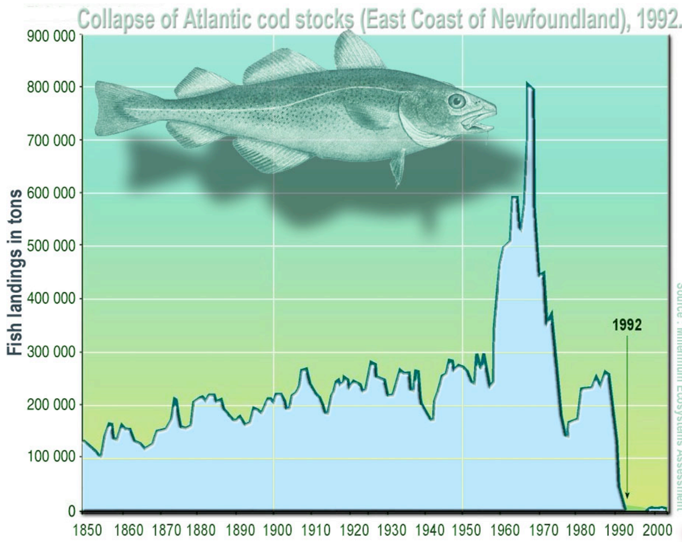
Networks everywhere

Networks as a powerful tool to solve problems

00:00 / 00:41

Networks as a powerful tool to understand reality

In the early 90's, the cod population in the northwestern Atlantic abruptly collapsed, generating an economic crisis in the Canadian fishery industry

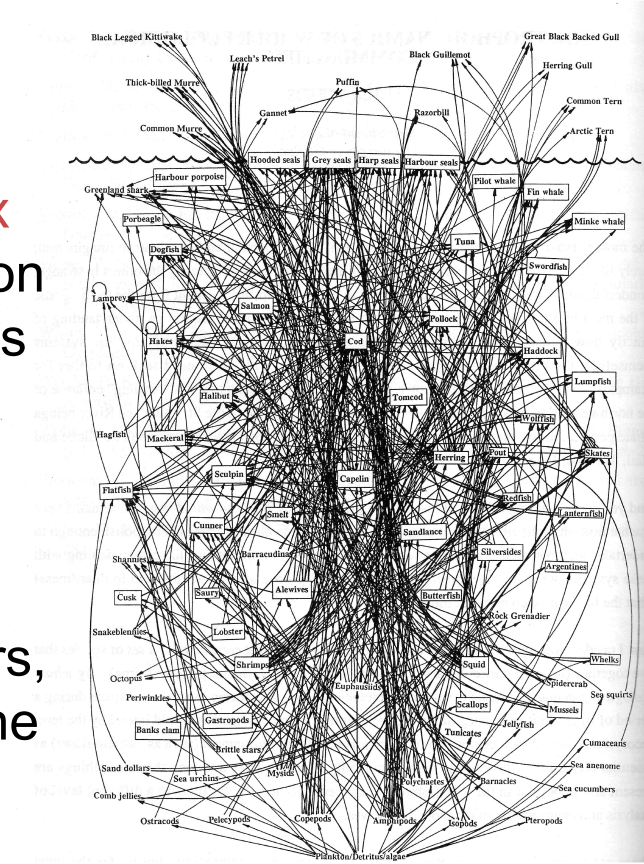


The Canadian government financed expeditions to hunt seals in Greenland, claiming that they, as cod's predators, were responsible for the collapse

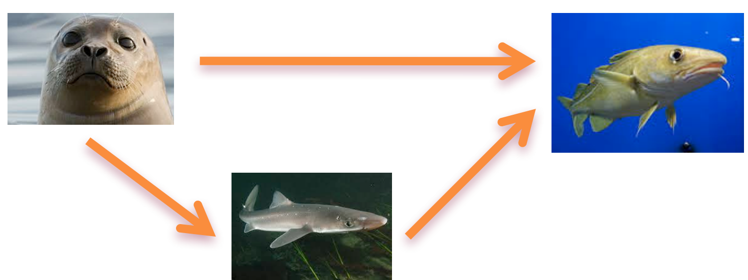
Nearly half a million seals were exterminated during the 90's, but the cod population continued to fall…

Networks as a powerful tool to understand reality

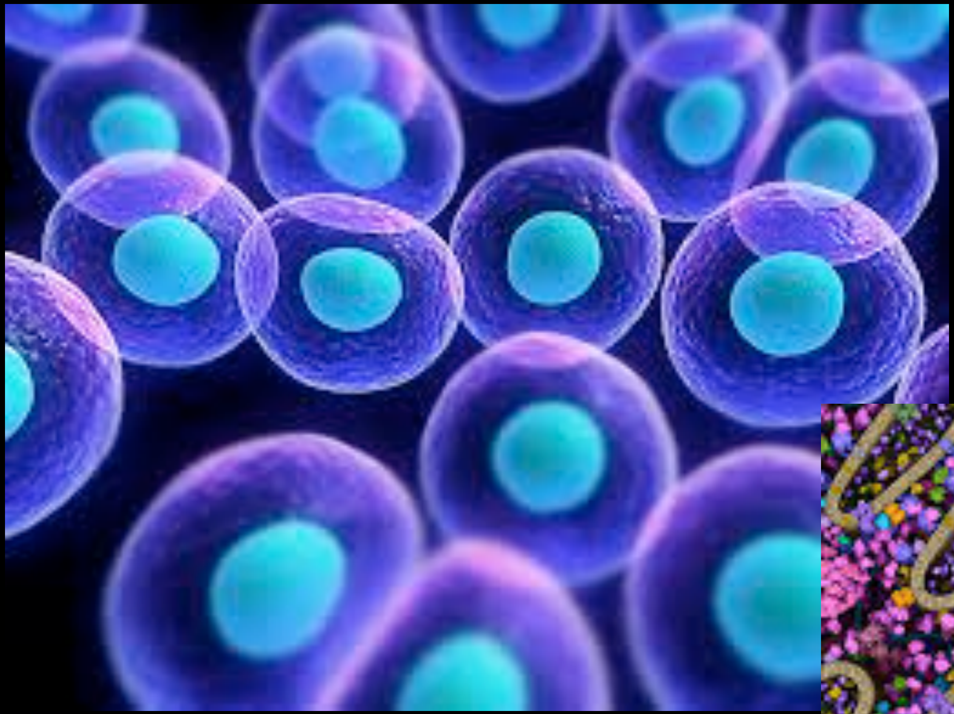
The food web of the northwest Atlantic ecosystem is a complex network with more than 10 million different chains connecting seals to cod



Moreover, between the 150 species preyed on by seals, there are several cod's predators, suggesting that a reduction in the population of seals can result in an increased pressure on cods

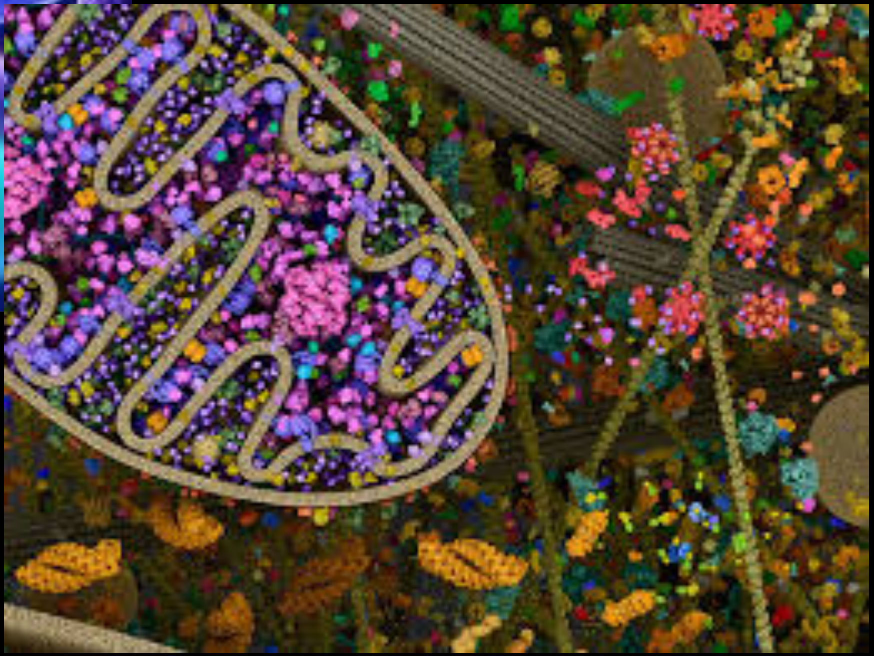


Networks as a powerful tool to understand reality
From Molecular Biology to Systems Biology



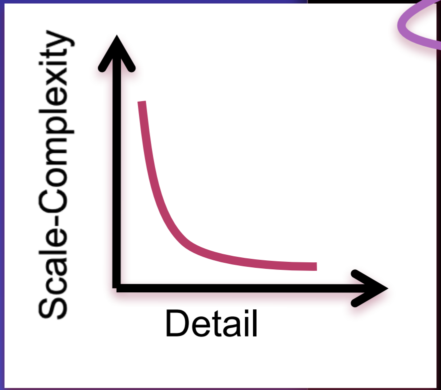
· Cells ·

· Crowded soup of genes, proteins, metabolites, … ·



Complex Systems

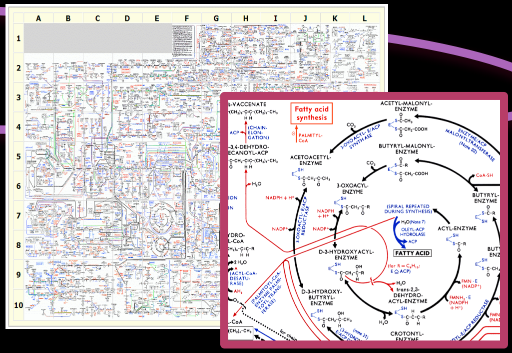
Complex Networks



Trade-off between small-scale detailed descriptions of parts of the system and the general panorama of the whole

A change of paradigm: From Molecular Biology to Systems Biology

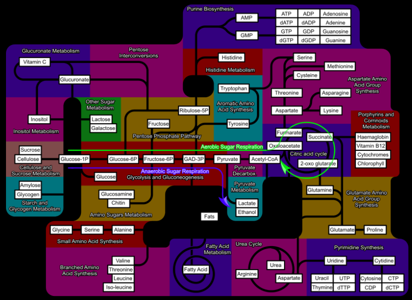
Complete metabolic network



Cellular metabolism is the set of interconnected chemical reactions that occur in cells in order **to maintain life**

Systems Biology

Biochemical pathways

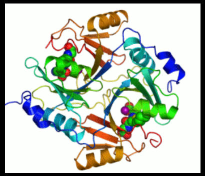


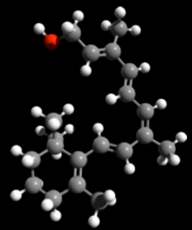
Biochemistry
Metabolic Control Analysis

Individual reactions (enzyme catalysis);Biomolecules (metabolites, enzymes...)

Chemistry, Physics...

$A+2B \rightarrow 3C$



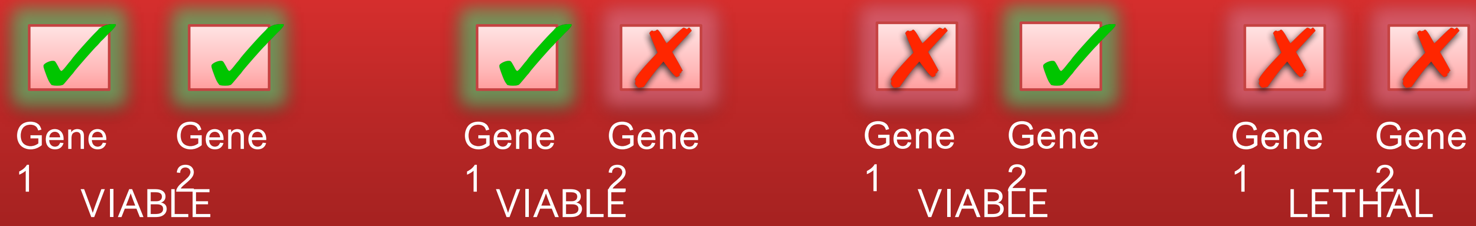


A change of paradigm: Synthetic Lethality

Systems level problems

In 1922, American geneticist Calvin Bridges was crossing fruit flies (*Drosophila melanogaster).* He noticed that mutations in certain non-essential genes were lethal only in combination

The term "synthetic lethality" was coined some 20 years later by Theodore Dobzhansky (evolutionary biology) who observed the same phenomenon in *Drosophila pseudoobscura*

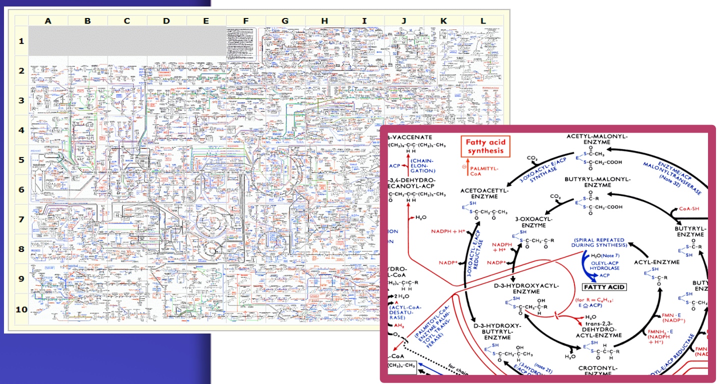


Synthetic lethality is an interesting phenomenon related to the complex organization of molecular interactions in the cell

This area has attracted a lot of attention because of the prospect of a new generation of anti-cancer drugs and drug resistance therapies based on SL targets

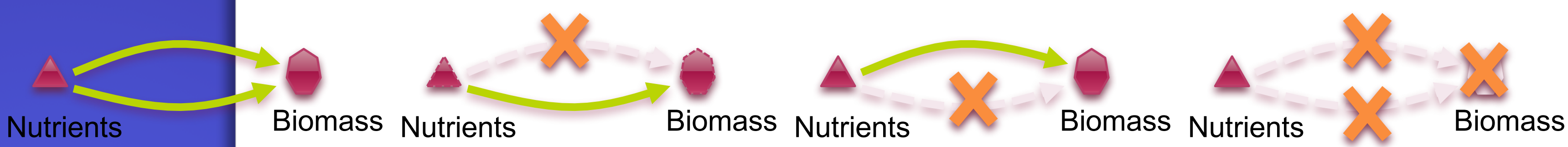
Synthetic lethality in metabolic networks

Güell O, Sagués F, Serrano M Á (2014) Essential Plasticity and Redundancy of Metabolism Unveiled by Synthetic Lethality Analysis. PLoS Comput Biol 10(5): e1003637-

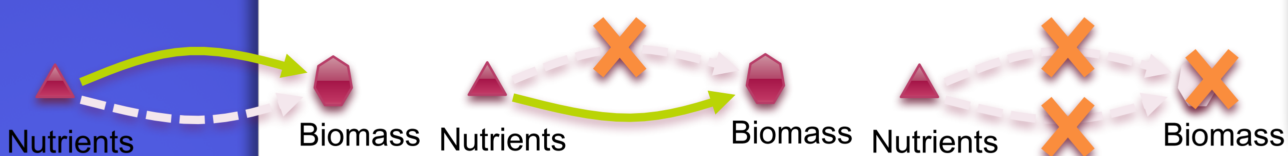


High quality genome-scale reconstructions

Redundancy Synthetic Lethality – Parallel use



Plasticity Synthetic Lethality – Backup mechanism



most common form of SL in metabolic networks

Networks as a powerful tool to understand reality
Economic systems

Trade-off between small-scale detailed descriptions of parts of the system and the general panorama of the whole

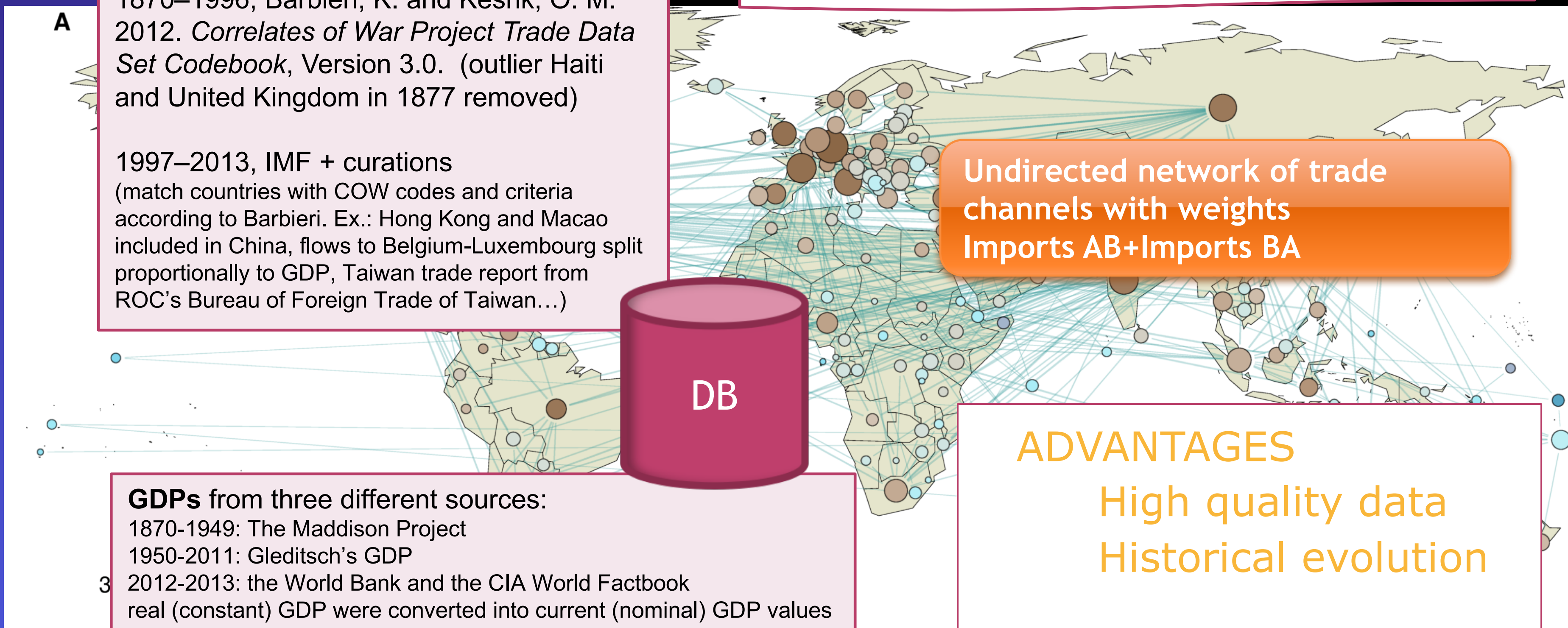
Traditional economics	**Complex systems approach**
• Closed, static, equilibrium	• Open, dynamic, non-equilibrium
• Complete information	• Incomplete partial information
• Rational agents, homogeneity, deductive calculations, no learning or adaptation	• Randomness, heterogeneous agents, inductive rules of thumb, learning, adaptation
• Interactions through markets; bilateral relationships	• Complex networks of interactions
•Micro and macro-economics remain separate disciplines	• Macro patterns are emergent results of micro-level behaviors and interactions; many levels of organization and interaction

Complex Systems

Complex Networks

Networks as a powerful tool to understand reality
International trade

Country A	Country B	Imports AB	Exports AB	Exports BA	Import BA



1870–1996, Barbieri, K. and Keshk, O. M. 2012. *Correlates of War Project Trade Data Set Codebook*, Version 3.0. (outlier Haiti and United Kingdom in 1877 removed)

1997–2013, IMF + curations
(match countries with COW codes and criteria according to Barbieri. Ex.: Hong Kong and Macao included in China, flows to Belgium-Luxembourg split proportionally to GDP, Taiwan trade report from ROC's Bureau of Foreign Trade of Taiwan…)

Undirected network of trade channels with weights Imports AB+Imports BA

GDPs from three different sources:
1870-1949: The Maddison Project
1950-2011: Gleditsch's GDP
2012-2013: the World Bank and the CIA World Factbook
real (constant) GDP were converted into current (nominal) GDP values using a deflator

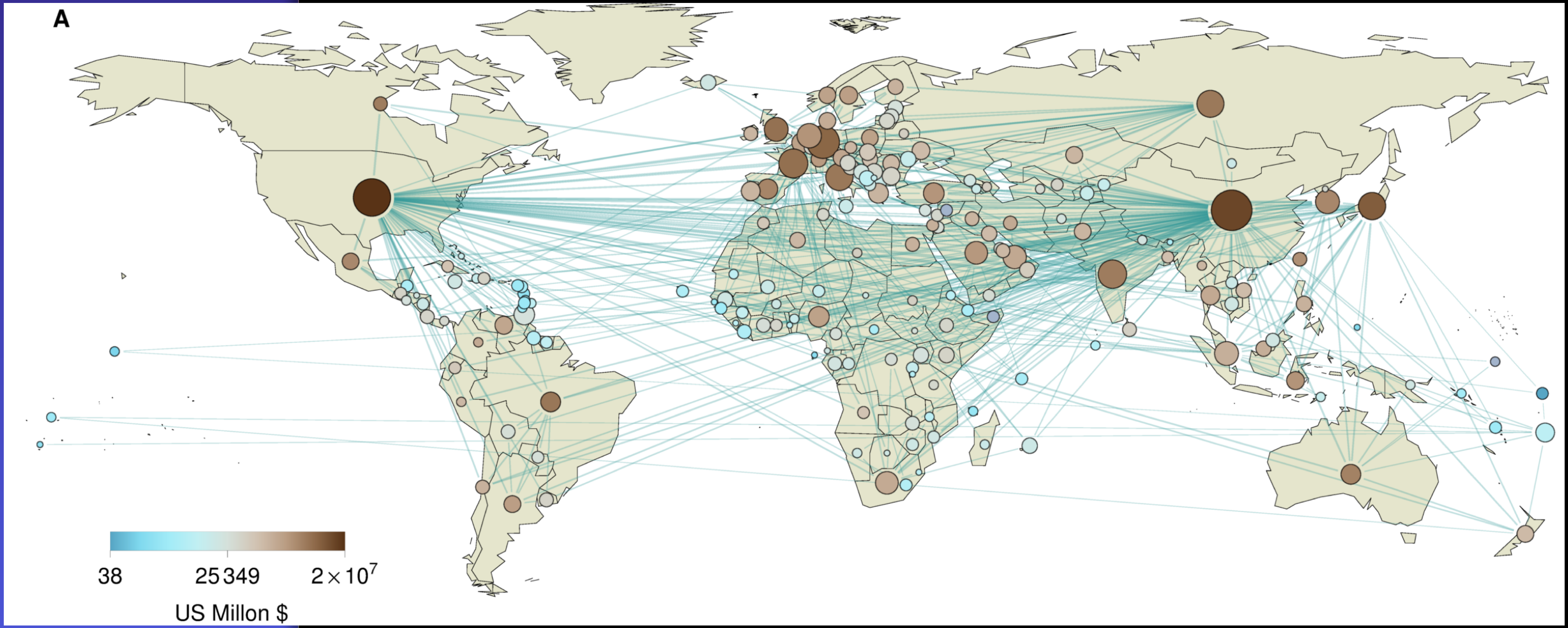
ADVANTAGES
- High quality data
- Historical evolution

CHALLENGES
- Very dense
- Pretty small network

Complex Systems

Complex Networks

Networks as a powerful tool to understand reality
International trade



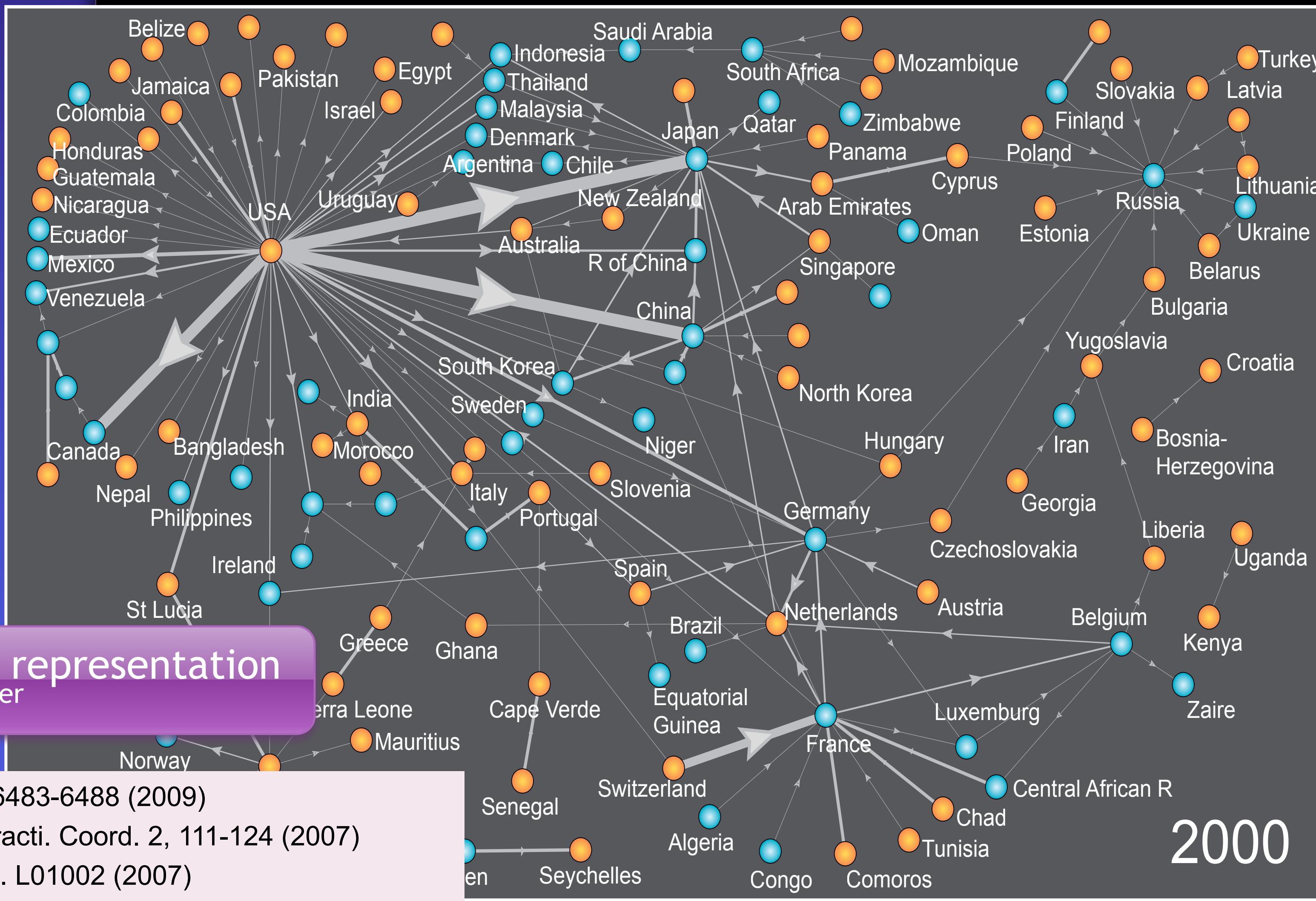
Exchange of goods and services across international borders of countries in the world

- Trade represents a significant share of gross domestic product (GDP)
- Trade has been at the basis of globalization processes though history

Complex Systems

Complex Networks

A change of paradigm: International trade



network representation
disparity filter

PNAS 106, 6483-6488 (2009)
J. Econ. Interacti. Coord. 2, 111-124 (2007)
J. Stat. Mech. L01002 (2007)
Phys. Rev. E 68, 015101 (R) (2003)

Complex Systems

Complex Networks

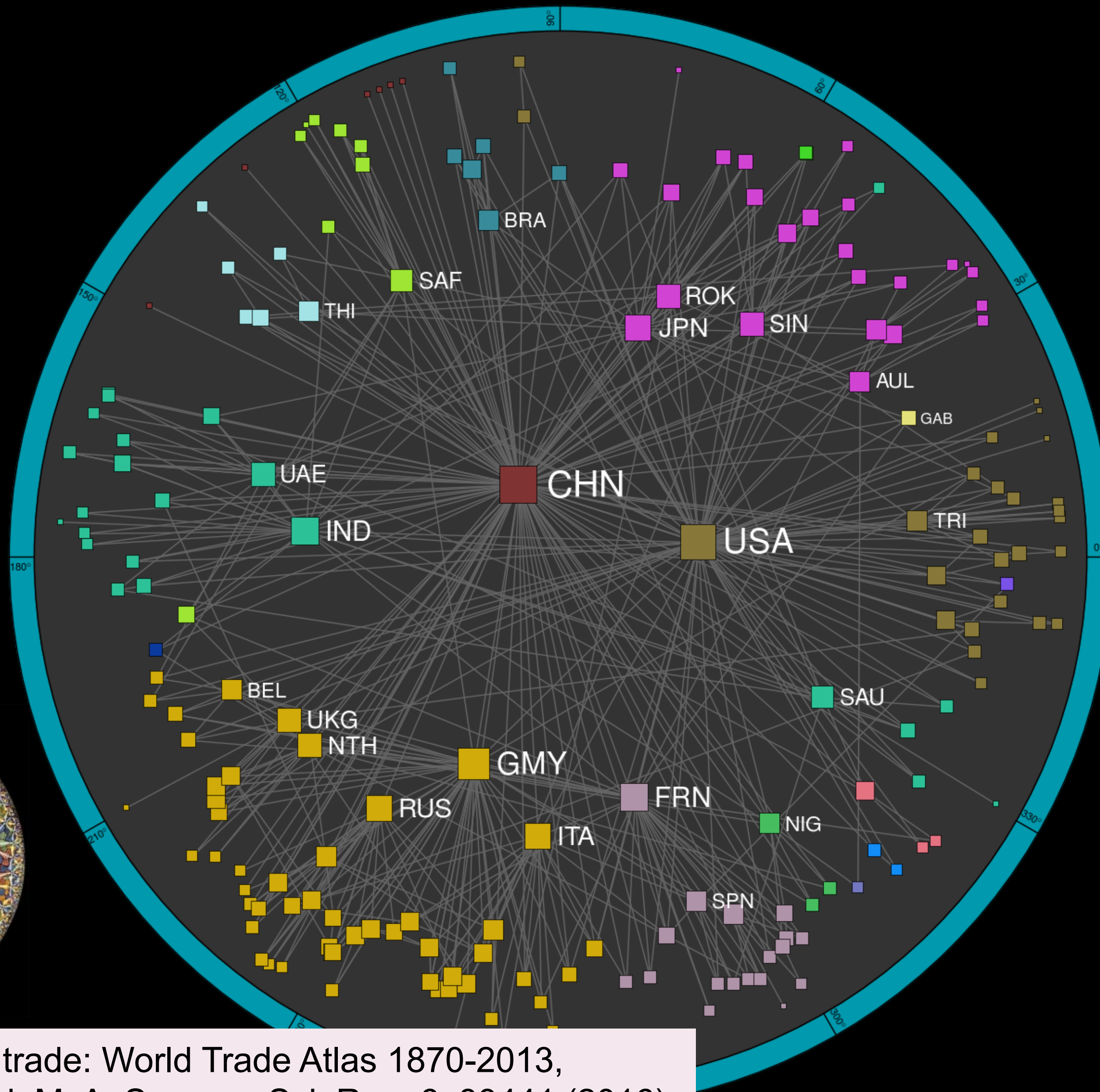
WTW Atlas 1870-2013

Maps in trade space

Trade space has an hyperbolic geometry

Large economies occupy central positions; small economies are displaced to the periphery

Circle Limit III, 1959
M. C. Escher



The hidden geometry of international trade: World Trade Atlas 1870-2013,
G. García-Pérez, M. Boguñá, A. Allard, M. A. Serrano, Sci. Rep. 6, 33441 (2016)

Complex Systems

Complex Networks

WTW Atlas 1870-2013

GLOBALIZATION INEQUALITIES

Since World War I, differences in trade distances have been growing, meaning that small economies find more difficulties to interact directly with each other, while large economies are increasing their chances of becoming connected worldwide

Signatures of globalization, localization, hierarchization

The hidden geometry of international trade: World Trade Atlas 1870-2013,
G. García-Pérez, M. Boguñá, A. Allard, M. A. Serrano, Sci. Rep. 6, 33441 (2016)

WTW Atlas 1870-2013

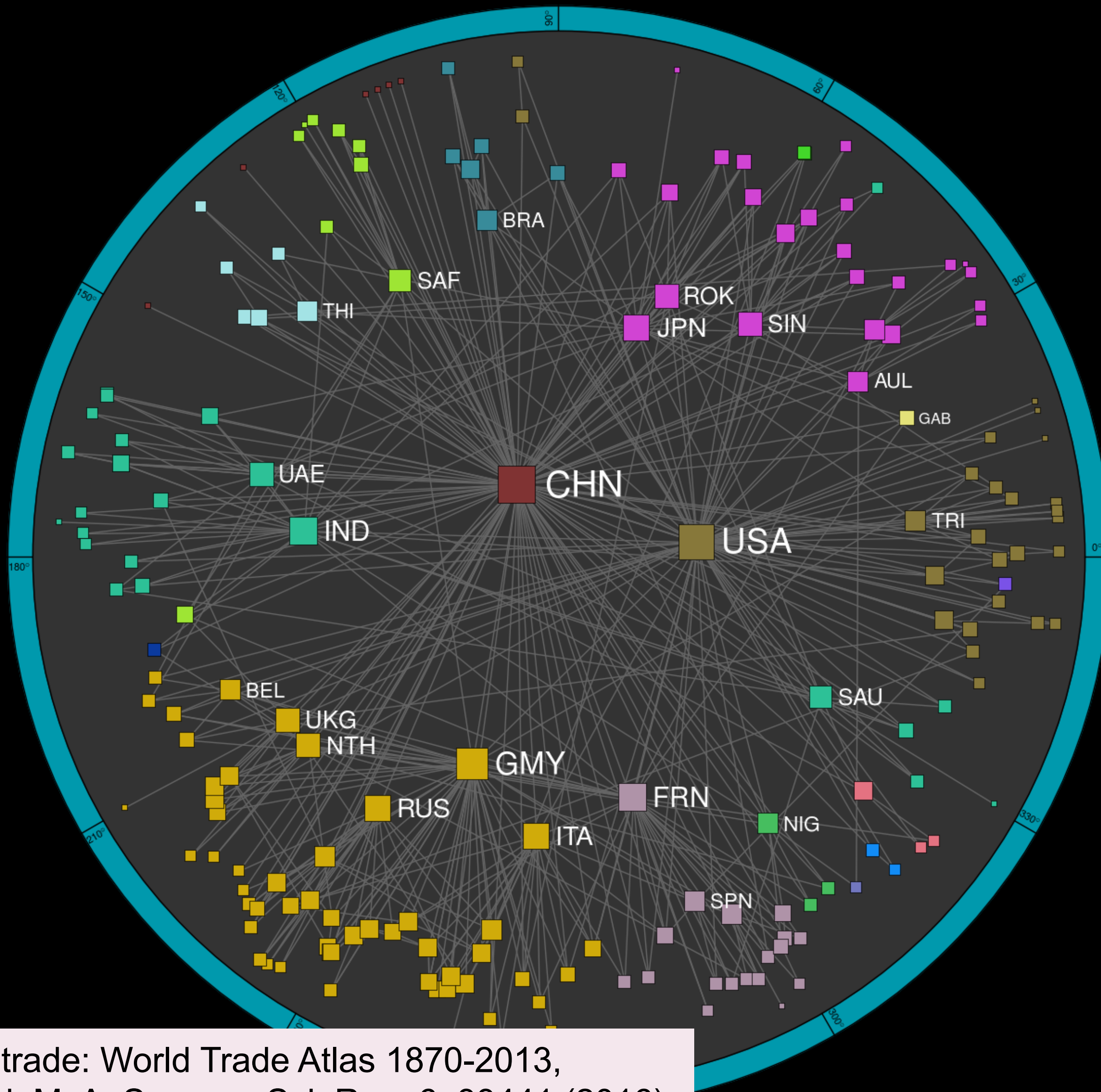
LOCALIZATION

Countries in the trade space organize in well-defined communities

The overlap between PTAs and natural communities is not perfect

EC Treaty: augmented internal barriers to trade
APTA: decreased internal barriers to trade

This process of localization happens at the same time and is entangled with globalization and hierarchization



The hidden geometry of international trade: World Trade Atlas 1870-2013,
G. García-Pérez, M. Boguñá, A. Allard, M. A. Serrano, Sci. Rep. 6, 33441 (2016)

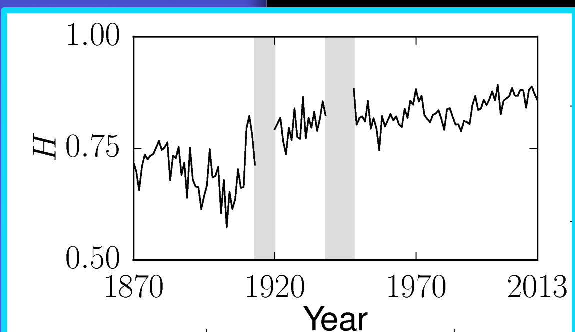
Complex Systems

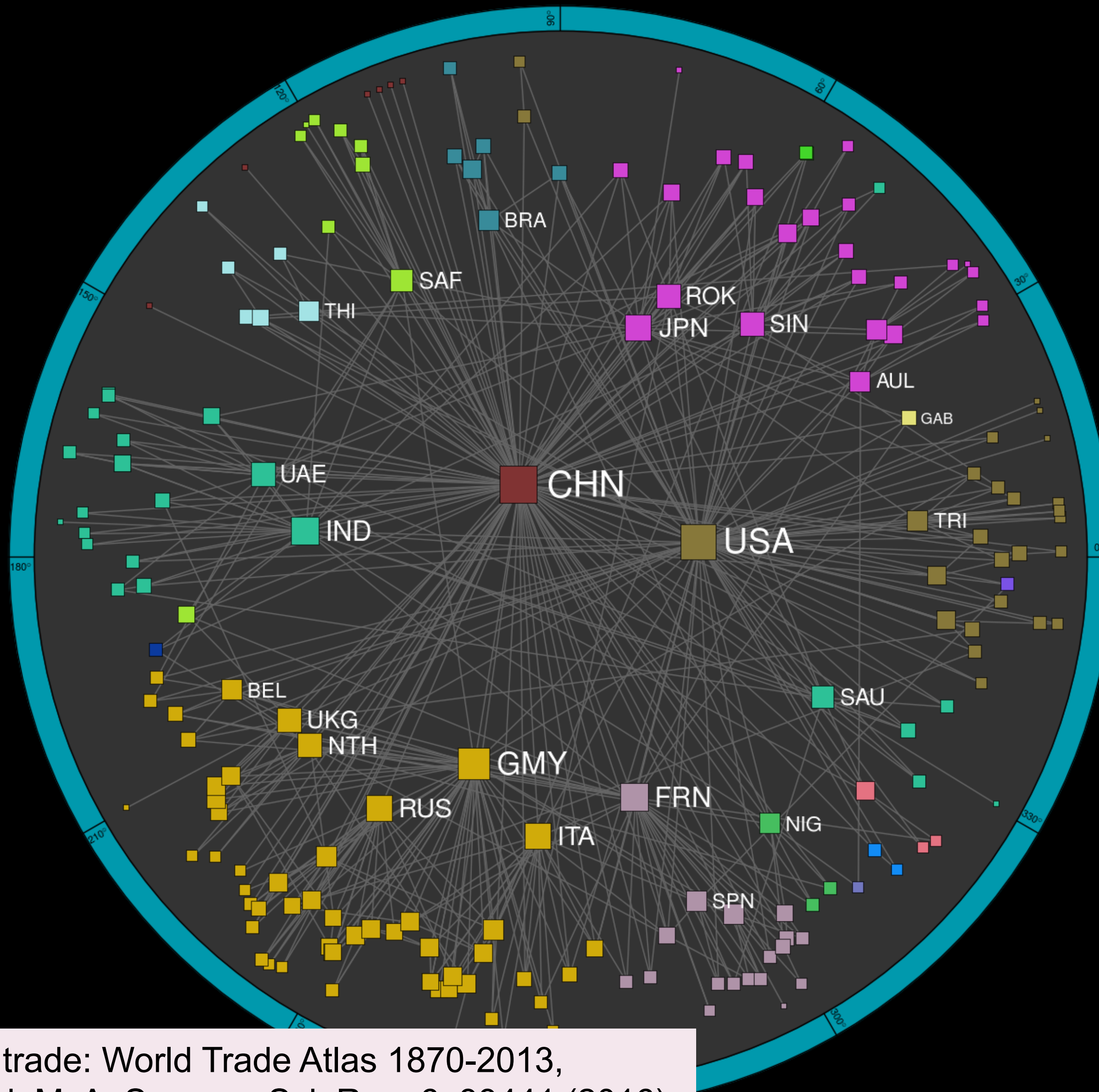
Complex Networks

WTW Atlas 1870-2013

HIERARCHIZATION

formation of intermediate hubs that dominate well-delimited angular regions as the number of layers in the hierarchy grows





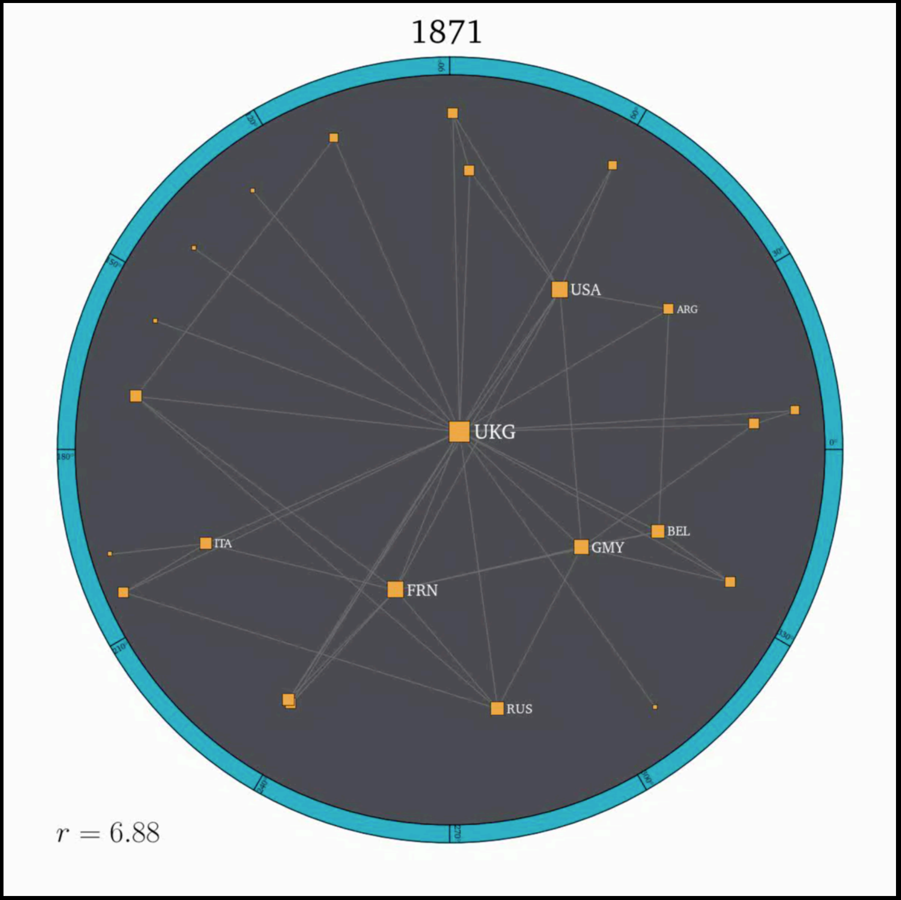
The hidden geometry of international trade: World Trade Atlas 1870-2013,
G. García-Pérez, M. Boguñá, A. Allard, M. A. Serrano, Sci. Rep. 6, 33441 (2016)

Complex Systems

Complex Networks

WTW Atlas 1870-2013

https://mappingcomplexity.net/nets2maps/www/WTA/index.html



Networks...

$^{\text{are}}$ changing the way in which we understand reality

$^{\text{are}}$ changing the way in which we solve complex problems

The network approach is crucial to propose judicious actions concerning some of the greatest open problems we are facing nowadays, like the efficient treatments against complex diseases, or the prediction and control of economic crises

Complex networks I.

Structural properties and network models

Complex Systems

Complex Networks

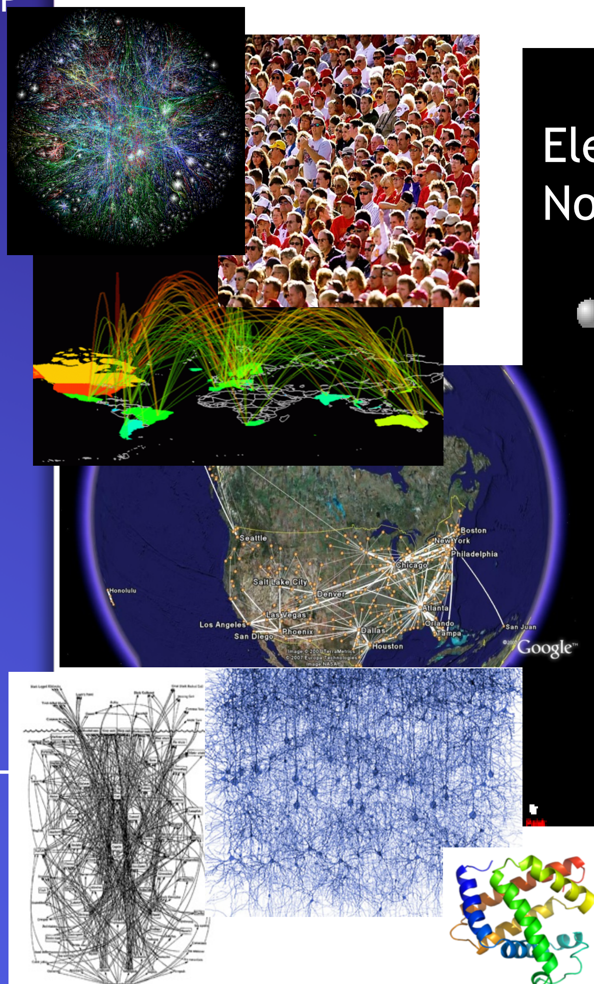
STRUCTURAL PROP
Global
Distances
SW
Degree distribution
ANND
Clustering
K-cores
Centrality
Communities

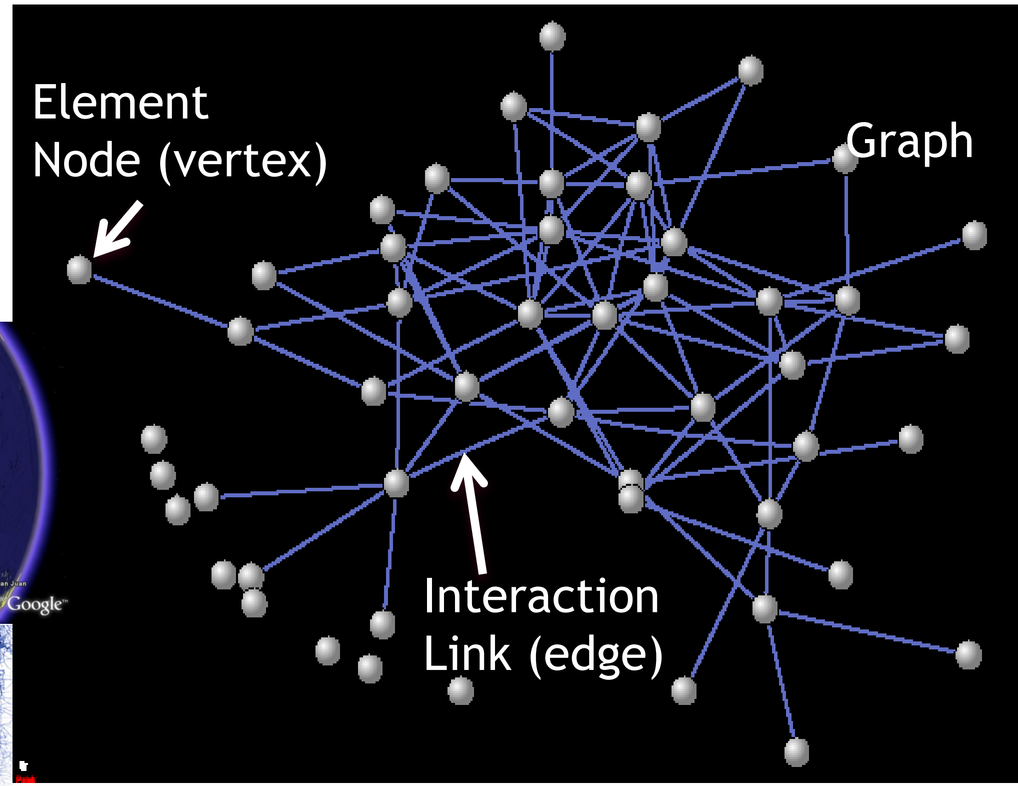
DESCRIPTIONS
Weighted
Directed
Bipartite
Multilayer

NETWORK MODEL
ER
SW
CM + Rewiring
BA
S1/H2
Null models

COMPLEX NETWORKS

Abstract representation of a discrete complex system





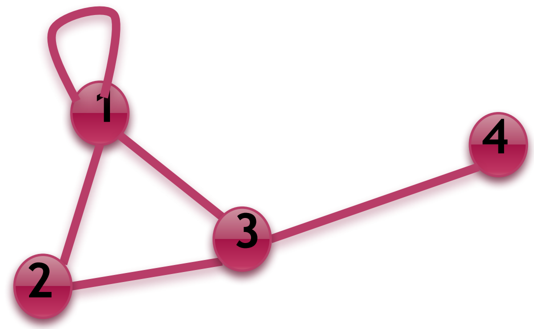
Between order and disorder randomness

COMPLEX NETWORKS

Fundamental representation

Adjacency matrix

$$\begin{pmatrix} 1 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$



Edge list (repeated or not)

1 1
2 1
3 1
2 3
3 4

Contains complete structural information,
but too detailed (large networks)…..

GLOBAL STATISTICS

N Number of nodes

E Number of edges

Complex networks are typically large in N

Complex networks are typically sparse (low E/N)

Node degree

Number of links per node

Notation: k_i = 0,1,2,. . . .

Average degree of a network

$$\langle k \rangle$$

Connection between number of edges E and average degree

$$2E = N\langle k \rangle$$

COMPLEX NETWORKS

ASSIGNMENT 1

Given a network in edge list format, make a simple but cool computer program that reads and stores it in memory, and prints the list of degrees, average degree <k>, and number of nodes N

Take a network from one of the public databases (at least 500 nodes, undirected)

ICON https://icon.colorado.edu/#!/networks

KONECT http://konect.cc/networks/

NDR https://networkrepository.com/

SNAP http://snap.stanford.edu/data/

Pajek http://vlado.fmf.uni-lj.si/pub/networks/data/default.htm

NetworkX
Network Analysis in Python

Gephi
makes graphs handy

WARNINGS: check format edge list repeated/not repeated; do not take into account multiple or self-connections; node label and node index

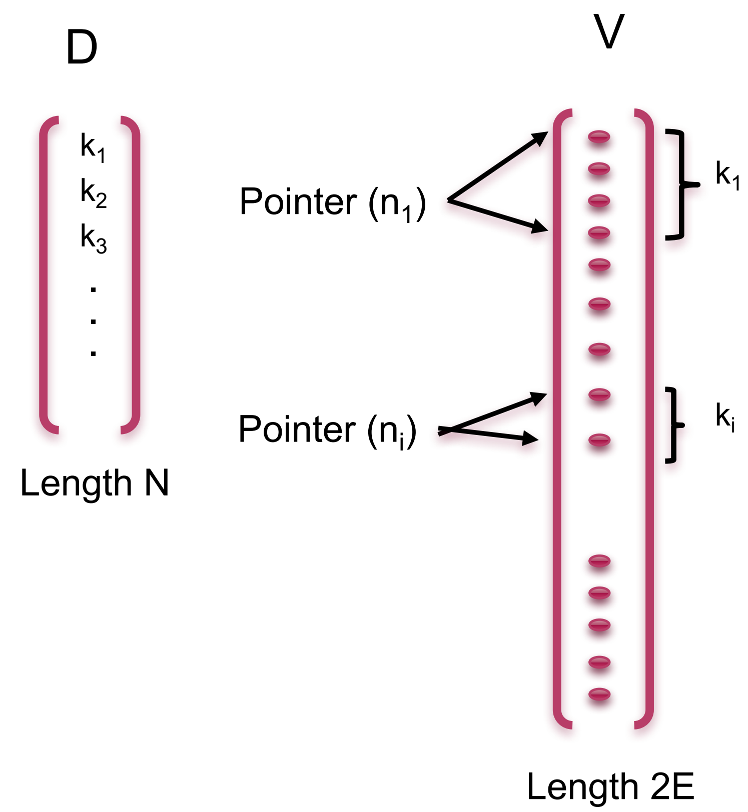
COMPLEX NETWORKS

ASSIGNMENT 1

Given a network in edge list format, make a simple but cool computer program that reads and stores it in memory, and prints the list of degrees, average degree <k>, and number of nodes N

- Create and initialize a vector V of length 2N to store the network and 2 pointers for each node, and a vector D of length N to store the degrees of the nodes

- Read the network file once. Update the current degree of nodes after reading each line



COMPLEX NETWORKS

ASSIGNMENT 1

Given a network in edge list format, make a simple but cool computer program that reads and stores it in memory, and prints the list of degrees

- Initialize the pointers according to degrees to reserve enough space in vector V to store the neighbors of each node. First pointer of node n is frozen and points to the first position in V that will store the neighbors of n. Second pointer is initialized to position of first pointer and incremented one position each time a neighbor is added

- Re-read the network file. Annotate in V the corresponding neighbors and update the second pointer of the corresponding nodes

- Check number of nodes and links and calculate average degree

- Print number of nodes, number of links, average degree, and the list of degrees in format “node degree”

DELIVERABLE: commented code + report

Complex Systems

Complex Networks

STRUCTURAL PROP
Global
Distances
SW
Degree distribution
ANND
Clustering
K-cores
Centrality
Communities

DESCRIPTIONS
Weighted
Directed
Bipartite
Multilayer

NETWORK MODELS
ER
SW
CM + Rewiring
BA
S1/H2
Null models

COMPLEX NETWORKS

COMPLEX NETWORKS I
- Connected components (PERCOLATION)
- Small world property (topological distance)
- Degree distribution
- Correlations (two-point (ANND) and three-point (clustering))
- K-cores
- Centrality (measures the importance of nodes)
- Community structure: modularity
- Motifs: very small recurring functional units in complex networks
- Hierarchy: a graded or ranked series
- Networks of networks: multiplex and multilevel networks
- Self-similarity: underlying geometries

Weights (Backbones) – Directionality – Bipartiveness - Multilayerty

Network models: ER, SW, CM, BA, S1/H2

COMPLEX NETWORKS II – PERCOLATION, DYNAMICS

COMPLEX NETWORKS III – NETWORK GEOMETRY

Complex Systems

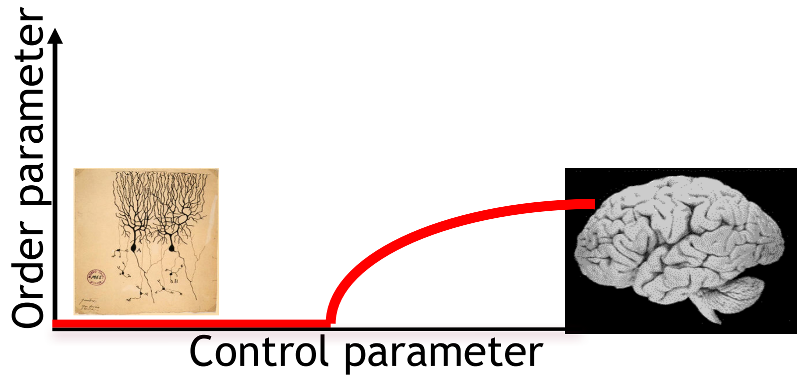
Complex Networks

STRUCTURAL PROP
Global
Distances
SW
Degree distribution
ANND
Clustering
K-cores
Centrality
Communities

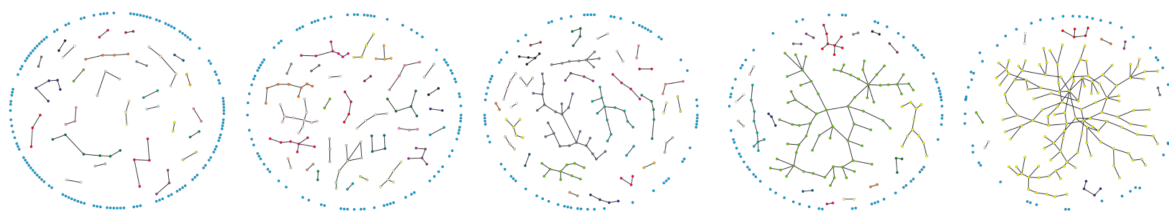
DESCRIPTIONS
Weighted
Directed
Bipartite
Multilayer

NETWORK MODELS
ER
SW
CM + Rewiring
BA
S1/H2
Null models

PERCOLATION



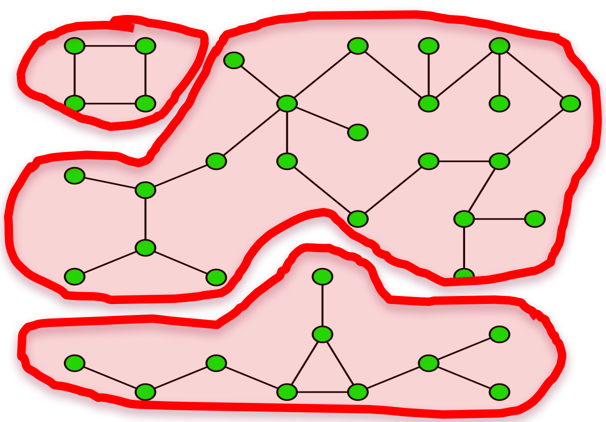
Giant connected component – complex networks are globally connected, what is essential to develop functionality



Percolation theory

Mathematics: structure of connected clusters in random graphs

Physics: fluid flow in random media



DISTANCES IN NETWORKS

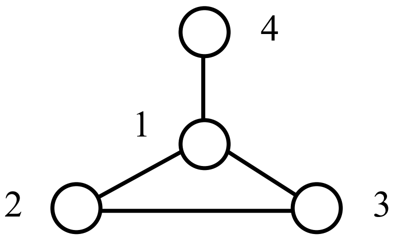
The distance between two nodes is measured as the number of links (unit length) in a shortest path connecting them

(a) shortest path length: Fewest number of steps between nodes i and j .

(b) average path length: Average shortest path length in the whole network.

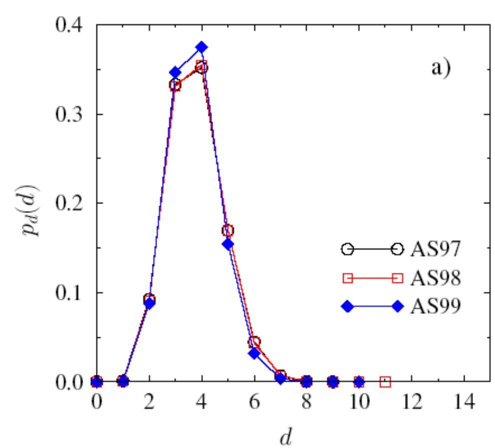
(c) Network diameter: Maximum shortest path length in the network.

ASPL and D computationally expensive



$d_{12} = d_{13} = d_{14} = d_{23} = 1$

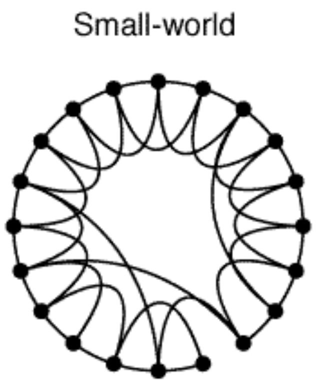
$d_{24} = d_{34} = 2$



Topological distances in networks are short and picked around a typical value (small-world property)

SMALL WORLD PROPERTY

Small-world property – networks are compact



each node in every pair is very close to each other

long range connections (shortcuts)

Collective dynamics of 'small-world' networks, D. J. Watts and S. H. Strogatz, Nature 393, 440-442(1998)

high clustering + small average path length

Six degrees of separation in social systems

Conjectured in 1929 by Frigyes Karinthy

Works in the 60's by Michael Gurevich, U.S. ; Kochen and de Sola Pool

Famous experiments by S. Milgram in 1967

Repeated in 2001, 2003 , right now (Internet-based: emails,…)

de Sola Pool, I., Kochen, M. (1978-1979)."Contacts and Influence." Social Networks 1(1): 5-51

S. Milgram, "The Small World Problem", Psychology Today, 1967, Vol. 2, 60-67

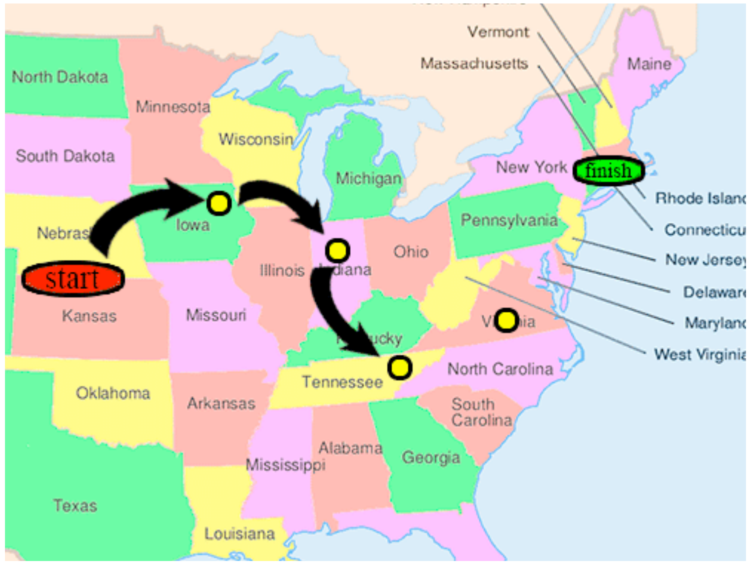
The maximum distance (diameter) grows slower than any polynomial

$$l(N) \sim \log N$$

The number of nodes grows exponentially with the diameter of the network

$$N \approx e^{\bar{d}}$$

MILGRAM'S EXPERIMENT



Probability that two randomly selected people would know each other???

Milgram chose individuals in the U.S. cities of Omaha, Nebraska, and Wichita, Kansas, to be the starting points and Boston, Massachusetts, to be the end point of a chain of correspondence (296 letters).

- 232 of 296 letters never reached the destination
- 64 of the letters eventually did reach the target contact. The average path length fell around five and a half or six.

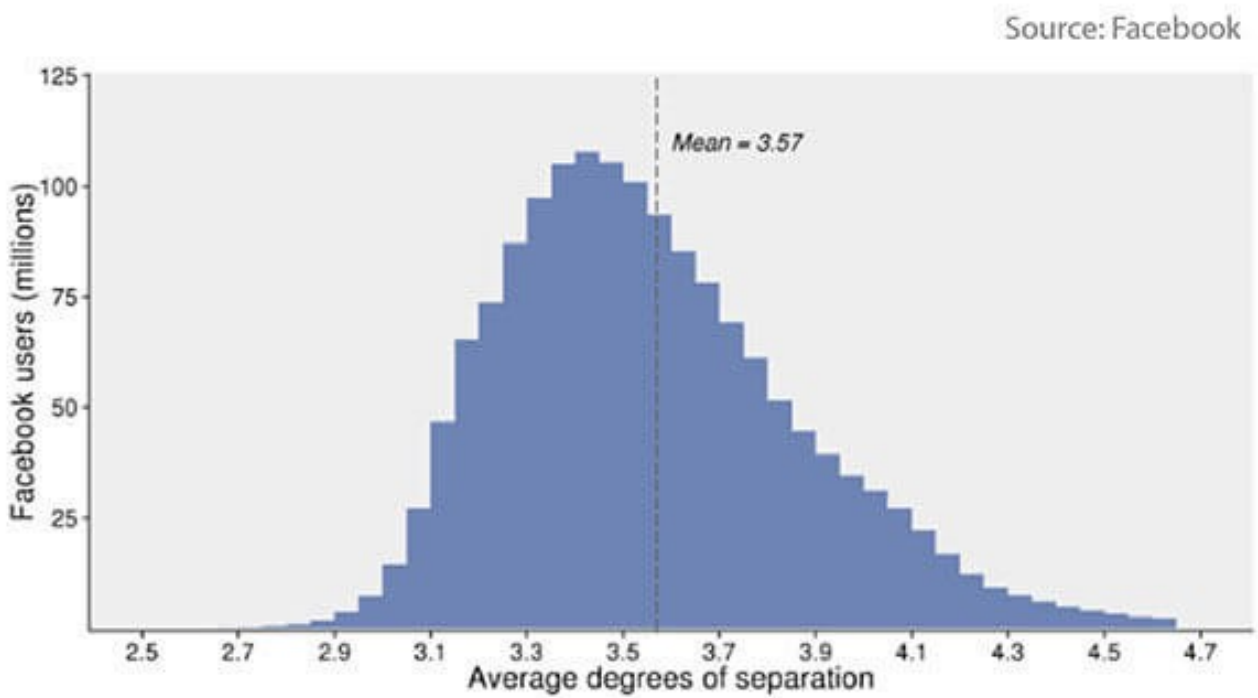
Conclusion:
people in the United States are separated by about six people on average

Do you know the contact person?

YES
Forward the letter directly to that person

NO
Think of a friend or relative who was more likely to know the target

SMALL WORLD PROPERTY



Each person in the world (at least among the 1.59 billion people active on Facebook) is connected to every other person by an average of three and a half other people. The average distance we observe is 4.57, corresponding to 3.57 intermediaries or "degrees of separation." (as of February 2016) https://research.fb.com/blog/2016/02/three-and-a-half-degrees-of-separation/

DEGREE DISTRIBUTION

Each node i has a degree k_i

(a) Degree distribution: the probability of a node having k neighbors $P(k)$

(b) Cumulative degree distribution: the probability of a node having at most degree k

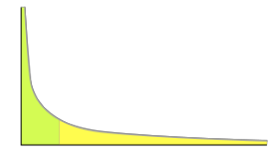
$$P(k_i \leq k)$$

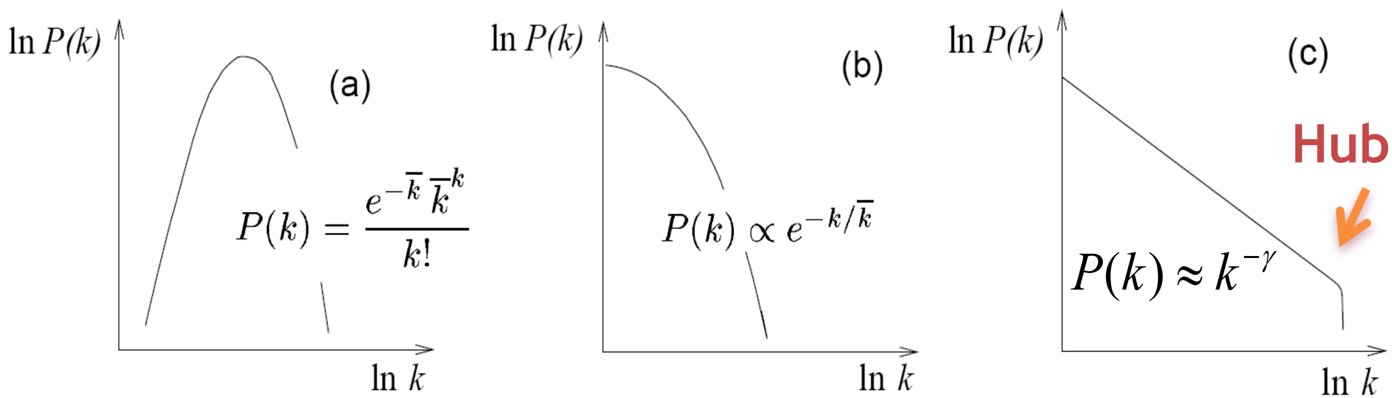
(c) Complementary cumulative degree distribution (1-CDF): probability of a node having a degree larger than k

$$P_c(k) = P(k_i > k)$$

DEGREE DISTRIBUTION

Scale-free- power law degree distribution





Poisson	Exponential	Scale-free
Classical random graphs	Growing random graph	BA model

Sometimes just fat-tailed or heterogeneous degree distribution

Complex Systems

Complex Networks

STRUCTURAL PROP
Global
Distances
SW
Degree distribution
ANND
Clustering
K-cores
Centrality
Communities

DESCRIPTIONS
Weighted
Directed
Bipartite
Multilayer

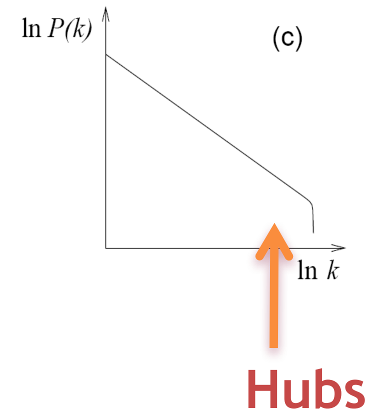
NETWORK MODELS
ER
SW
CM + Rewiring
BA
S1/H2
Null models

DEGREE DISTRIBUTION

Scale-free degree distributions - thermodynamic limit

Real Networks: Exponent typically between 2 and 3

$$P(k) = k^{-\gamma} / A$$



Normalization factor (γ greater than 1)

$$A = \sum k_i^{-\gamma}$$

Riemann zeta function with argument s=γ

$$\zeta(s) = \sum_{n=1}^{\infty} n^{-s} = \frac{1}{1^s} + \frac{1}{2^s} + \frac{1}{3^s} + \cdots$$

If k is taken as a continuous variable: $P(k) = \frac{\gamma - 1}{k_{\min}} \left(\frac{k}{k_{\min}} \right)^{-\gamma}$

First moment

$$\bar{k} = \sum_{k=0}^{\infty} k P(k)$$

In infinite networks, all moments of order greater or equal to γ-1 diverge

⟨k^2⟩→∞   k_c→∞   k_c~N^[1/(γ−1)]

Complex Systems

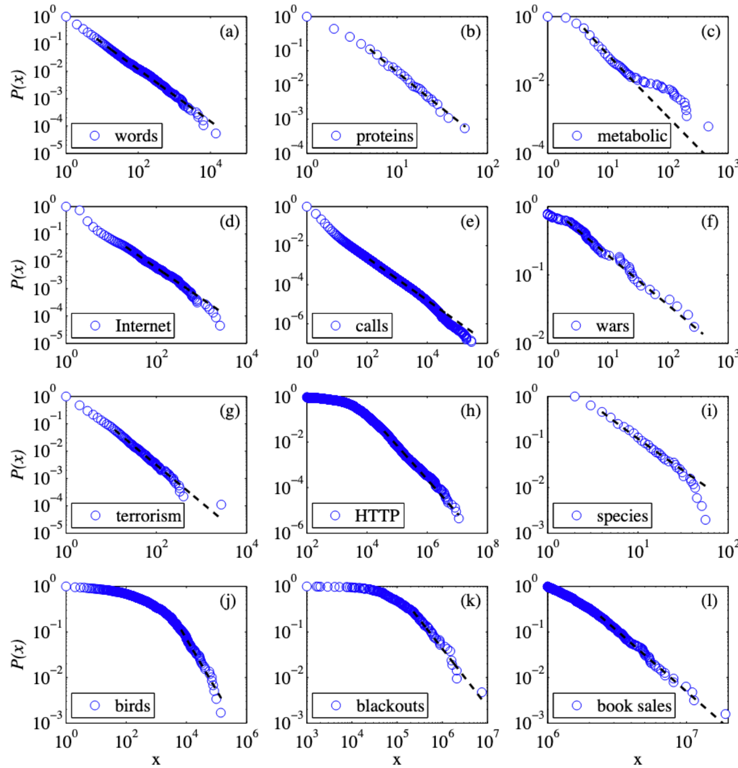
Complex Networks

STRUCTURAL PROP
Global
Distances
SW
Degree distribution
ANND
Clustering
K-cores
Centrality
Communities

DESCRIPTIONS
Weighted
Directed
Bipartite
Multilayer

NETWORK MODELS
ER
SW
CM + Rewiring
BA
S1/H2
Null models

DEGREE DISTRIBUTION



"*Power-law* distributions in empirical data"

https://aaronclauset.github.io/powerlaws/

FIG. 6.1. *The cumulative distribution functions $P(x)$ and their maximum likelihood power-law fits for the first twelve of our twenty-four empirical data sets. (a) The frequency of occurrence of unique words in the novel Moby Dick by Herman Melville. (b) The degree distribution of proteins in the protein interaction network of the yeast S. cerevisiae. (c) The degree distribution of metabolites in the metabolic network of the bacterium E. coli. (d) The degree distribution of autonomous systems (groups of computers under single administrative control) on the Internet. (e) The number of calls received by US customers of the long-distance telephone carrier AT&T. (f) The intensity of wars from 1816–1980 measured as the number of battle deaths per 10 000 of the combined populations of the warring nations. (g) The severity of terrorist attacks worldwide from February 1968 to June 2006, measured by number of deaths. (h) The number of bytes of data received in response to HTTP (web) requests from computers at a large research laboratory. (i) The number of species per genus of mammals during the late Quaternary period. (j) The frequency of sightings of bird species in the United States. (k) The number of customers affected by electrical blackouts in the United States. (l) The sales volume of bestselling books in the United States.*

A. *Clauset*, C.R. Shalizi, and M.E.J. *Newman*, SIAM Review 51(4), 661-703 (2009).

Correlations

M. A. Serrano, M. Boguñá, R. Pastor-Satorras and A. Vespignani. Correlations in complex networks, in Large Scale Structure and Dynamics of Complex Networks, G. Caldarelli and A. Vespignani editors, World Scientific (2007), pp. 35-66.

Two-point degree correlations:

P(k,k'), probability that a link selected at random connects two nodes of degrees k and k'

Three-point degree correlations:

P(k,k',k''), prob. that a randomly chosen triangle connects three vertices of degrees k, k', and k''

From a local perspective, P(k'|k) and P(k'',k'|k)

probability that a node of degree k is connected to a node of degree k'

probability that a node of degree k is simultaneously connected to two nodes of degree k' and k''

In particular, P(k'|k) obeys the detailed balance condition

(closure of the network, physical conservation of edges; the total number of edges pointing from vertices with degree k to vertices of degree k′ must be equal to the total number of edges that point from vertices with degree k′ to vertices of degree k)

$$kP(k' \,|\, k)P(k) = k'P(k \,|\, k')P(k') = \langle k \rangle P(k, k')$$

Correlations

Proof of the detailed balance condition

$$kP(k' \,|\, k)P(k) = k'P(k \,|\, k')P(k') = \langle k \rangle P(k, k')$$

$E_{kk'}$ symmetric matrix counting number of edges between degree classes k and k' (2 times for the diagonal

$$\sum_{k'} E_{kk'} = kN_k,$$

links of degree class k

$$P(k' \,|\, k) = \frac{E_{k'k}}{kN_k}$$

$$\sum_{k,k'} E_{kk'} = \langle k \rangle N = 2E,$$

total number of links

$$P(k, k') = \frac{E_{kk'}}{\langle k \rangle N}$$

$$P(k' \,|\, k) = \frac{\langle k \rangle P(k, k')}{kP(k)}$$

Symmetry of P(k,k')

DBC

M. A. Serrano, M. Boguñá, R. Pastor-Satorras and A. Vespignani. Correlations in complex networks, in Large Scale Structure and Dynamics of Complex Networks, G. Caldarelli and A. Vespignani editors, World Scientific (2007), pp. 35-66.

Complex Systems

Complex Networks

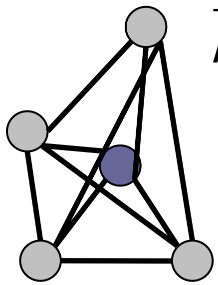
STRUCTURAL PROP
Global
Distances
SW
Degree distribution
ANND
Clustering
K-cores
Centrality
Communities

DESCRIPTIONS
Weighted
Directed
Bipartite
Multilayer

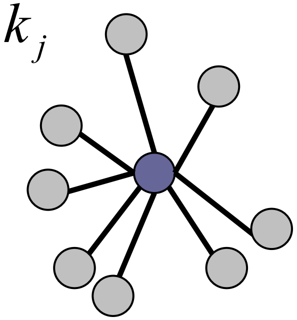
NETWORK MODELS
ER
SW
CM + Rewiring
BA
S1/H2
Null models

Average nearest neighbors degree

For empirical measurements: uniparametric projections



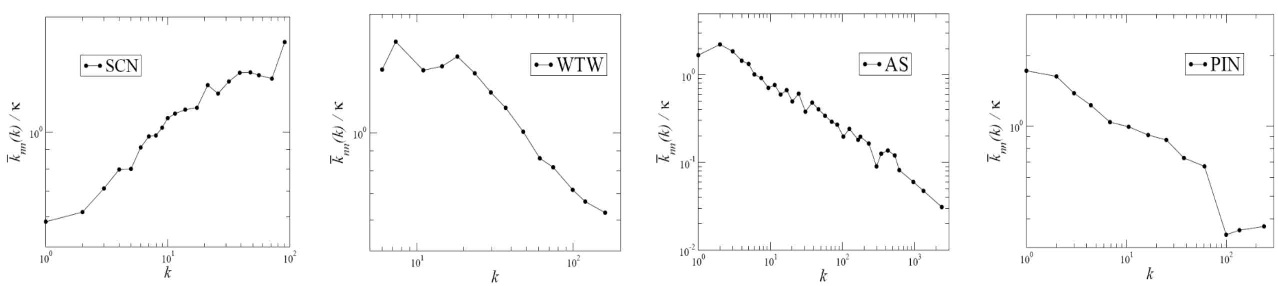
$$\bar{k}_{nn}(k) = \sum_{k'} k' P(k'|k) = \frac{1}{N_k} \sum_{i \in \nu(k)} \frac{1}{k_i} \sum_j a_{ij} k_j$$



It is a measure of the tendency of nodes to connect to peers in terms of degree

Assortative mixing

Disassortative mixing

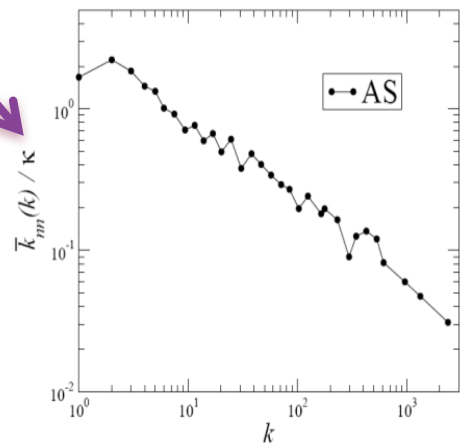


Average nearest neighbors degree

Empirical measurements

$$\bar{k}_{nn}(k) = \sum_{k'} k' P(k'|k)$$

$$\frac{\langle k^2 \rangle}{\langle k \rangle}$$



The function ANND/κ has average 1 (*over a certain distribution that depnds on the degree distribution of the network)* independently of the network and the pattern of correlations and so it can be used to compare two point correlations in different networks

Correction that makes comparable the ANND functions of different real networks

Boguñá, M; Pastor-Satorras, R; Vespignani, A PHYSICAL REVIEW LETTERS 90, 28701 (2003)

Complex Systems

Complex Networks

STRUCTURAL PROP
Global
Distances
SW
Degree distribution
ANND
Clustering
K-cores
Centrality
Communities

DESCRIPTIONS
Weighted
Directed
Bipartite
Multilayer

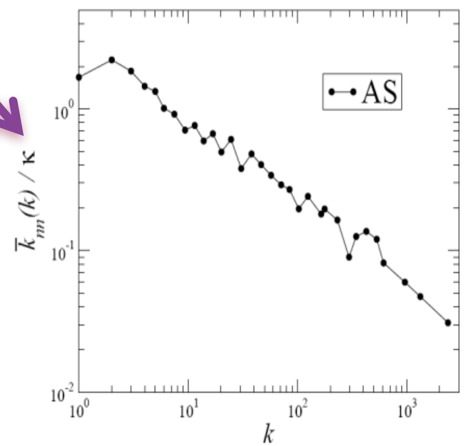
NETWORK MODELS
ER
SW
CM + Rewiring
BA
S1/H2
Null models

Average nearest neighbors degree

Empirical measurements

$$\bar{k}_{nn}(k) = \sum_{k'} k' P(k'|k)$$

$$\frac{\langle k^2 \rangle}{\langle k \rangle}$$



Correction that makes comparable the ANND functions of different real networks

Proof from detailed balance:

$$kP(k'|k)P(k) = k'P(k|k')P(k')$$

multiplying by a k factor both terms and summing over k′ and k

$$\langle k^2 \rangle = \sum_{k'} k' P(k') \sum_k kP(k|k')$$

with the normalization conditions $\sum_k P(k) = \sum_{k'} P(k'|k) = 1$.

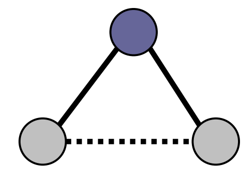
Divide by <k> both sides such that k’P(k’) is well normalized, this quantity is the probability of arriving to a node of degree k when selecting a link at random

$$\frac{<k^2>}{<k>} = \sum_{k'} \frac{k'P(k')}{<k>} \overline{k_{nn}}(k')$$

Boguñá, M; Pastor-Satorras, R; Vespignani, A PHYSICAL REVIEW LETTERS 90, 28701 (2003)

Clustering

It is a measure of the likelihood that two neighbors of a given vertex are neighbors themselves



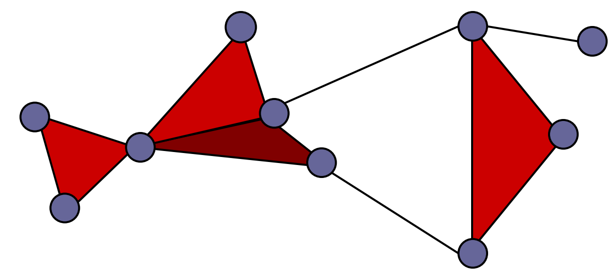
- Global measures

$$C_\Delta = \frac{3 \times (\text{number of triangles})}{(\text{number of connected triples})}$$

It is a measure of the number of triangles in a network

- Local measures

$$c_i = \frac{2T_i}{k_i(k_i - 1)} \qquad C = \frac{\sum_i c_i}{N}$$



Clustering coefficient

Projections of the three-point distribution function P(k,k',k'')

Complex Systems

Complex Networks

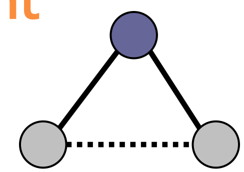
STRUCTURAL PROP
Global
Distances
SW
Degree distribution
ANND
Clustering
K-cores
Centrality
Communities

DESCRIPTIONS
Weighted
Directed
Bipartite
Multilayer

NETWORK MODELS
ER
SW
CM + Rewiring
BA
S1/H2
Null models

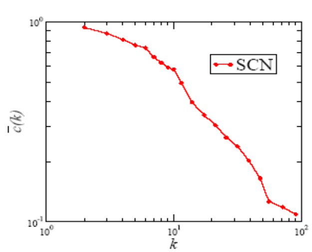
Clustering

- **Degree dependent Clustering coefficient**
  average of local clustering over degree classes

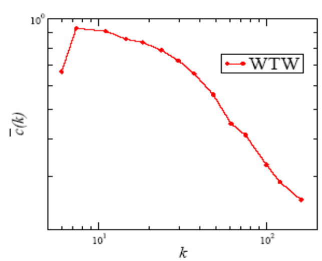


$$\bar{c}(k) = \frac{1}{N_k} \sum_{i \in \Upsilon(k)} c_i = \frac{1}{k(k-1)N_k} \sum_{i \in \Upsilon(k)} 2T_i \qquad \bar{c} = \sum_k P(k)\bar{c}(k)$$

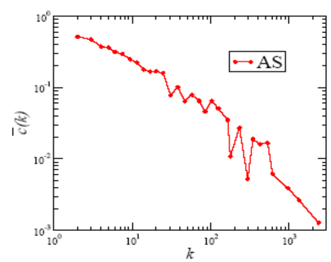
$$\bar{c}(k) = \sum_{k'k''} P(k'k''|k) r^{k}_{k'k''} = \frac{1}{N_k} \sum_{i \in \nu(k)} \frac{1}{k_i(k_i - 1)} \sum_{jl} a_{ij} a_{il} a_{jl}$$



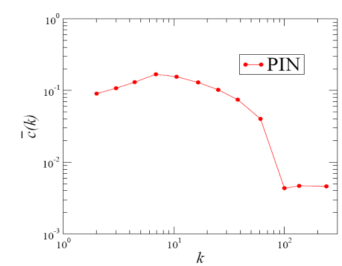
A. Science Collaboration Network cond-mat 1995-1998



B. World Trade Web of commercial channels among countries



C. The Internet at the Autonomous System level



D. Protein interaction network

Clustering

Also important:

Weak vs Strong clustering $\bar{c}(k) < \dfrac{1}{k-1}$

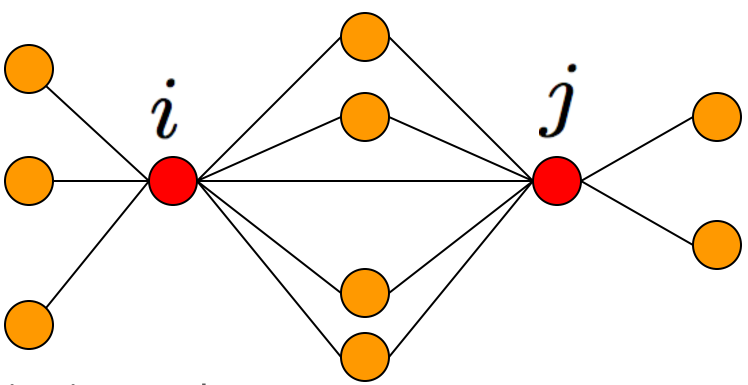
Triangles are disjoint

Triangles coalesce

The locally tree like assumption is a good approximation for networks with weak clustering, and enables analytical approaches like the generating function formalism to estimate percolation thresholds, sizes of giant components...

Edge multiplicity m_{ij}

Number of triangles passing through edge i-j



M. A. Serrano and M. Boguñá, Clustering in complex networks. I. General formalism, Physical Review E 74, 056114 (2006)

COMPLEX NETWORKS

ASSIGNMENT 2

Given a network, make a computer program that calculates its degree distribution (direct and complementary cumulative), its average nearest neighbors degree, and its clustering

Use the same network as in ASSIGNMENT 1

- Read and store the network including degrees (ASSIGNMENT 1)

- Calculate P(k) and cc P(k): read degrees and calculate the normalized histograms (1 loop) of degree occurrences (1 vector) and accumulated degree occurrences (1 loop from 0 to current degree, 1 vector)

- Calculate k_nn(k): for each node i (1 loop), increment the corresponding entry in k_nn(k) (1 vector) with the local contribution of all neighbors of i (degree of neighbor normalized by degree of node and number of nodes in degree class k). Include the normalization.

COMPLEX NETWORKS

ASSIGNMENT 2

Given a network, make a computer program that calculates its degree distribution (direct and complementary cumulative), its average nearest neighbors degree, and its clustering

- Calculate the average clustering coefficient $\bar{c}(k)$: for each node i (1 loop) increment the corresponding entry in $\bar{c}(k)$ (1 vector) with the local contribution of all neighbors of i (number of triangles normalized by the maximum possible number of triangles given the degree of node i and by number of nodes in degree class k)

To calculate the number of triangles attached to a node, check connections between all possible pairs of neighbors of the node (3 nested loops)

DELIVERABLE: commented code + report

K-cores

Concept to study the tree-likeness of a network

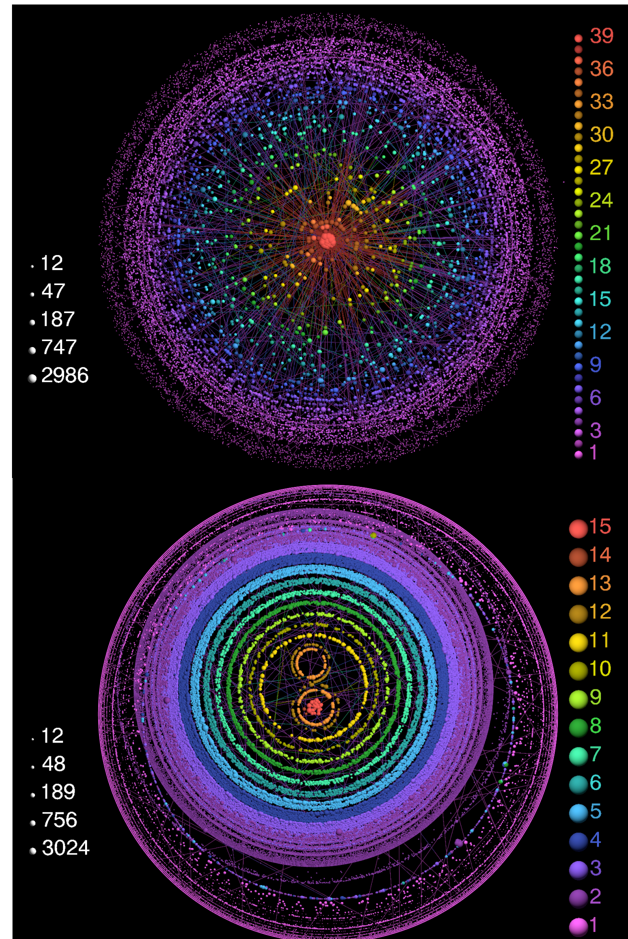
k-core: subgraph in which every node has at least k connections with other neighbors in the subgraph

The k-cores are obtained in a recursive way:

i. remove from the original graph all nodes (and their connections) with degree less than k

ii. remove from the remaining graph all nodes (and their connections) with degree less than k

iii. repeat ii) until no further removal is possible



LANET-VI visualization

Complex Systems

Complex Networks

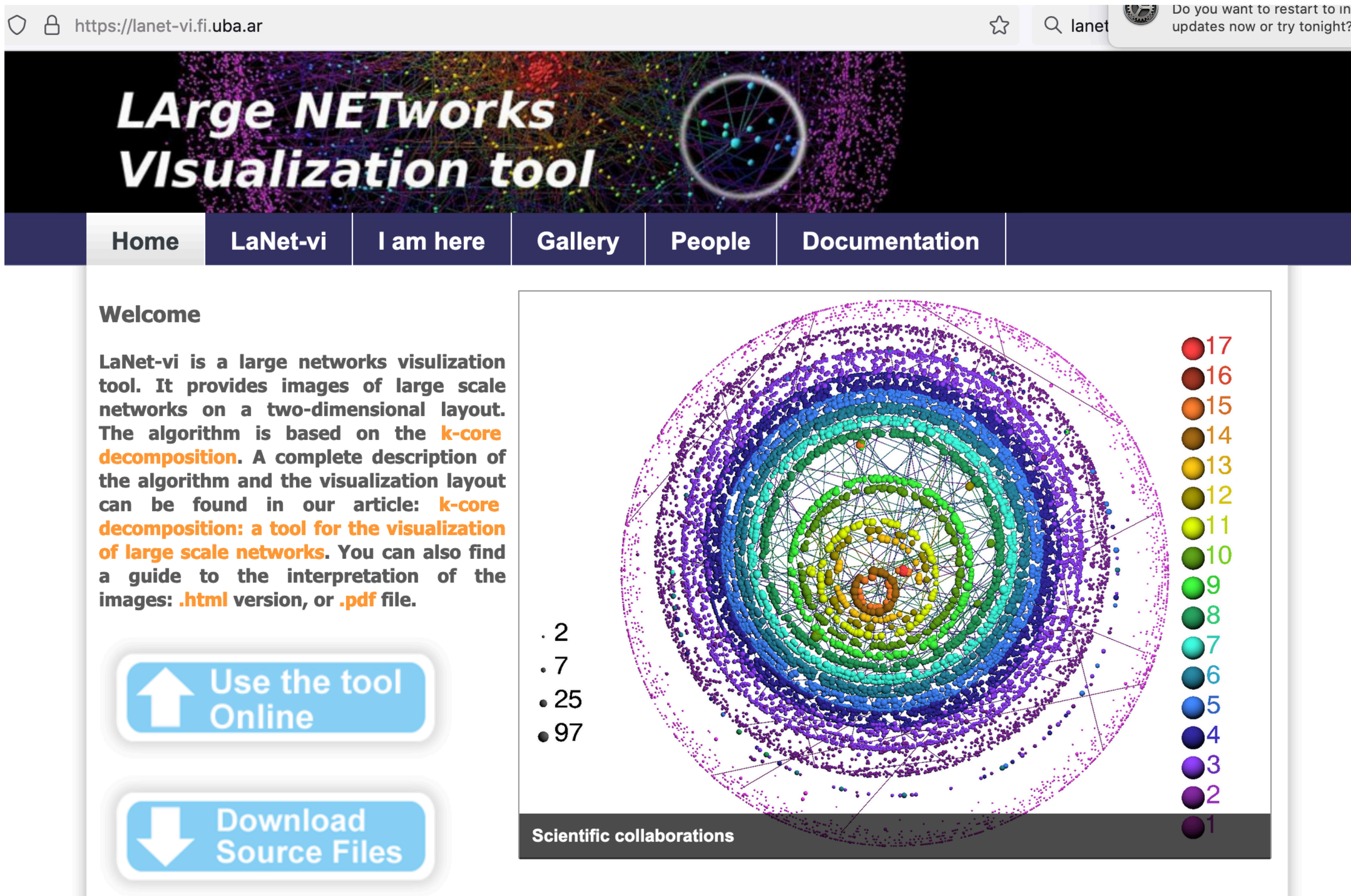
STRUCTURAL PROP
Global
Distances
SW
Degree distribution
ANND
Clustering
K-cores
Centrality
Communities

DESCRIPTIONS
Weighted
Directed
Bipartite
Multilayer

NETWORK MODELS
ER
SW
CM + Rewiring
BA
S1/H2
Null models

K-cores

LANET-VI tool



Complex Systems

Complex Networks

STRUCTURAL PROP
Global
Distances
SW
Degree distribution
ANND
Clustering
K-cores
Centrality
Communities

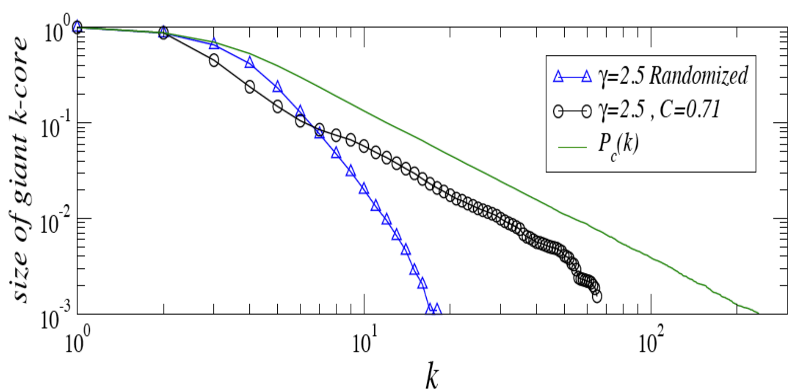
DESCRIPTIONS
Weighted
Directed
Bipartite
Multilayer

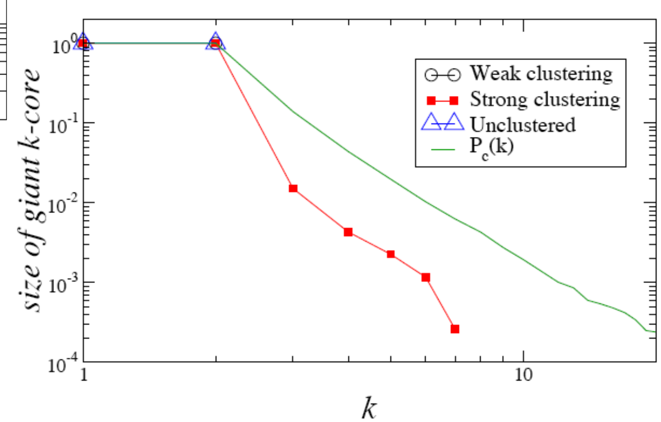
NETWORK MODELS
ER
SW
CM + Rewiring
BA
S1/H2
Null models

K-cores

The k-core organization of a network gives idea of its sparsity/tree-likeness

Stronger clustering produces a deeper hierarchy of k-cores

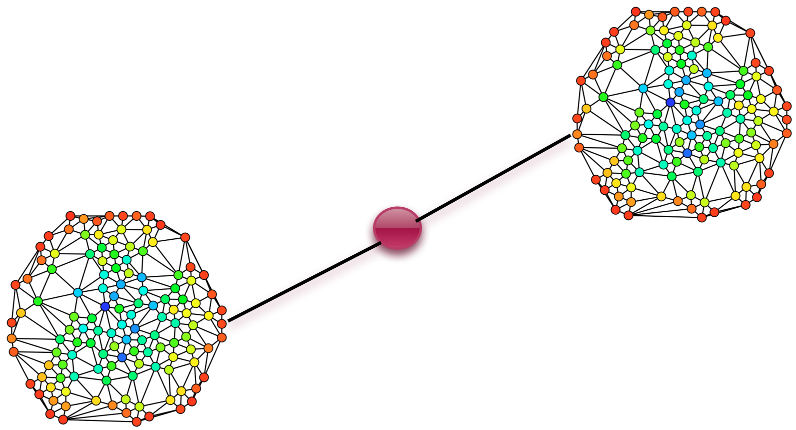




Centrality

Structural measures of a node's 'importance'

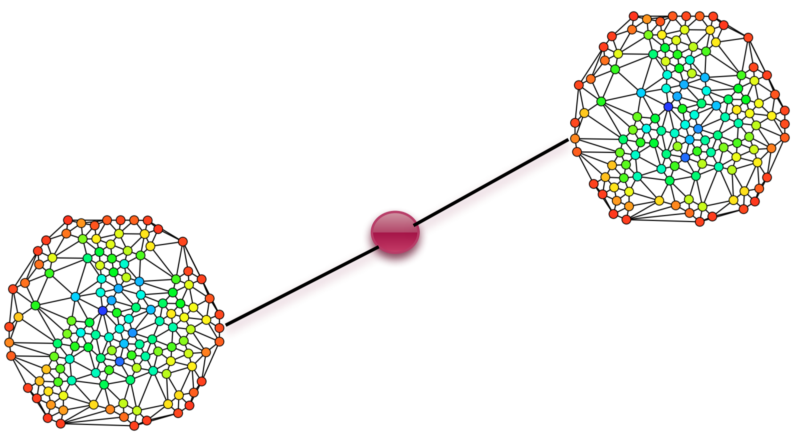
- **Degree centrality:** k



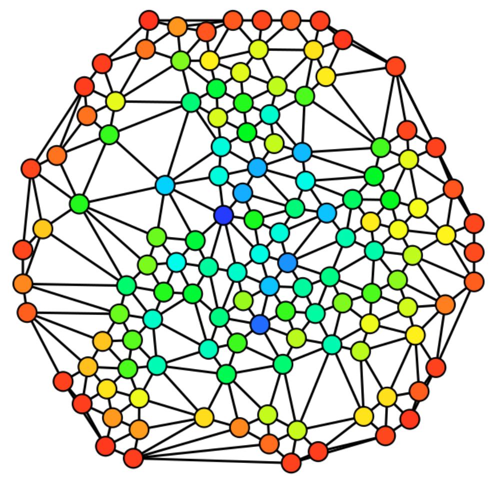
Centrality

Structural measures of a node's 'importance'

- **Degree centrality:** k



Node betweenneess



- **Node/Edge betweenness:** fraction of all shortest paths between any two vertices that pass through that node/edge

Centrality

Structural measures of a node's 'importance'

- **Eigenvector centrality**. Eigenvector with the greatest eigenvalue (all the entries non-negative) of the adjacency matrix.

The score of a node is proportional to the sum of the scores of all its neighbors. A node that is linked to many nodes with high score receives a high score itself.

$$\mathbf{A}\mathbf{x} = \lambda\mathbf{x} \qquad x_i = \frac{1}{\lambda}\sum_j a_{ij}x_j$$

Variant: Google's PageRank algorithm to rank web pages. The underlying assumption is that more important websites are likely to receive more links from other important websites.

PageRank= stationary state of an unbiased random walk in the network (normalized EC)+random jumps

normalized EC in undirected networks $\quad x_i = \sum_j \frac{1}{k_j} a_{ij}x_j = k_i$

Complex Systems

Complex Networks

STRUCTURAL PROP
Global
Distances
SW
Degree distribution
ANND
Clustering
K-cores
Centrality
Communities

DESCRIPTIONS
Weighted
Directed
Bipartite
Multilayer

NETWORK MODELS
ER
SW
CM + Rewiring
BA
S1/H2
Null models

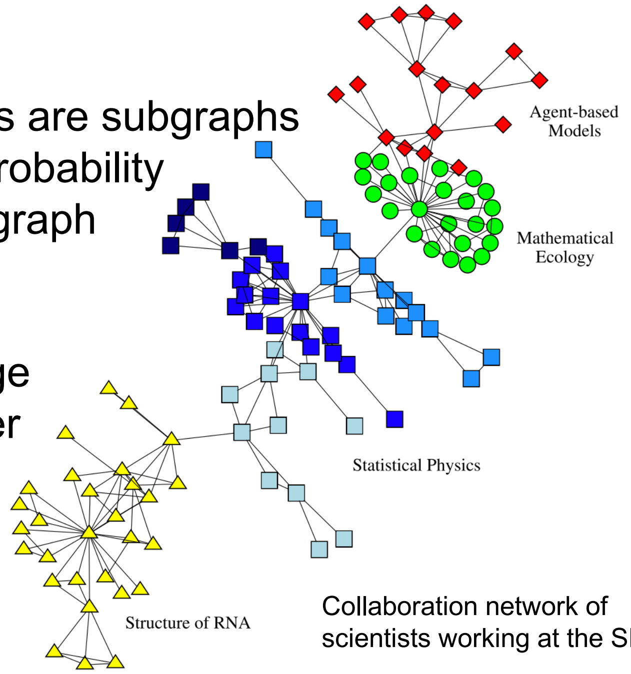
Communities in networks

Nodes can be organized into groups called communities, clusters, or modules. Identifying communities is a major question in NS

Definition of communities?

A traditional definition is that communities are sets of nodes more densely interconnected.

A more modern definition: communities are subgraphs each of whose vertices has a higher probability to be linked to every vertex of the subgraph than to any other vertex of the graph (*strong community*),
or such that such that the average edge probability of each vertex with the other members of the group exceeds the average edge probability of the vertex with the vertices of any other group.



Collaboration network of scientists working at the SFI.

S. Fortunato. D. Hric, ***Community detection in networks: a user guide***, Physics Reports **659**, 1-44 (2016)
S. Fortunato, ***Community detection in graphs***, Physics Reports **486**, 75-174 (2010)

Complex Systems

Complex Networks

STRUCTURAL PROP
Global
Distances
SW
Degree distribution
ANND
Clustering
K-cores
Centrality
Communities

DESCRIPTIONS
Weighted
Directed
Bipartite
Multilayer

NETWORK MODELS
ER
SW
CM + Rewiring
BA
S1/H2
Null models

Communities in networks

Many traditional community detection methods are based on Newman and Girvan modularity

$$Q = \frac{1}{2m} \sum_{ij} \left(A_{ij} - \frac{k_i k_j}{2m} \right) \delta(C_i, C_j).$$

M.E.J. Newman, M. Girvan, Phys. Rev. E 69 (2) (2004) 026113.

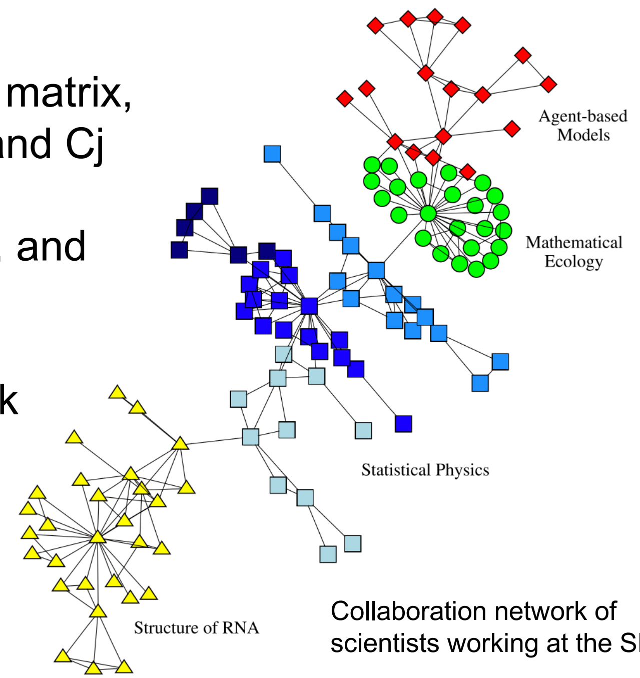
- m is the number of edges
- Aij is the element of the adjacency matrix,
- the Kronecker delta at the end Ci and Cj

indicates the communities of i and j

- Pij = kikj/2m is the null model term, and

corresponds to the expected number of edges joining vertices i and j in a randomized version of the network preserving degrees.

Therefore, modularity measures how different the original graph is from such randomizations.



Collaboration network of scientists working at the SFI.

The concept was inspired by the idea that by randomizing the network structure communities are destroyed...

Complex Systems

Complex Networks

STRUCTURAL PROP
Global
Distances
SW
Degree distribution
ANND
Clustering
K-cores
Centrality
Communities

DESCRIPTIONS
Weighted
Directed
Bipartite
Multilayer

NETWORK MODELS
ER
SW
CM + Rewiring
BA
S1/H2
Null models

Communities in networks

Method	**Approach**	**Reference**
Louvain	Multilevel modularity	Blondel et al. (2008) [2]
Fast greedy	Modularity optimization	Clauset et al. (2004) [4]
Spectral	Vector partitioning	Newman and Girvan (2004) [18]
Spin glass	Energy model	Reichardt and Bornholdt (2006) [21]
DCSBM	Stochastic blockmodels	Karrer and Newman (2011) [13]
Walktrap	Dynamic distance	Pons and Latapy (2005) [19]
Conclude	Dynamic distance	Meo et al. (2014) [16]
Edge betweenness	Edge centrality detection	Girvan and Newman (2002) [10]
Infomap	Information compression	Rosvall and Bergstrom (2008) [22]
Label propagation	Topological closeness	Raghavan et al. (2007) [20]
SLPA	Topological closeness	Xie and Szymanski (2011) [25]

Table 1 : A summary of community detection methods used to study community structure.

and many more….

Weighted networks

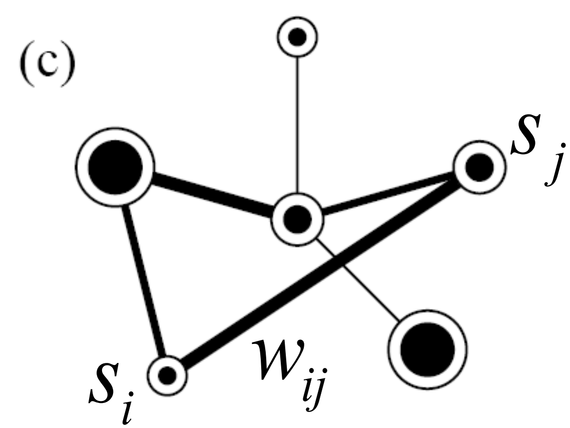
Links are not binary (just present or absent) but are characterized by different intensities, w

The sum of all weights incident on a node is its strength, s

$$s_i = \sum_j w_{ij}$$

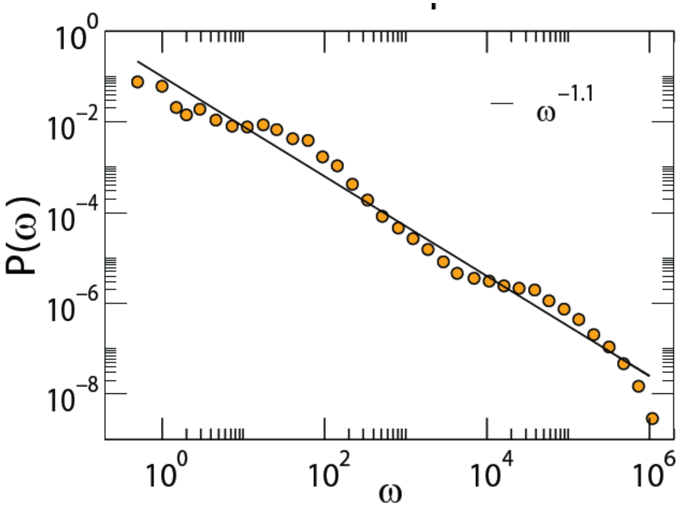
The degree distribution has to be complemented with the distribution of weights associated to links and the distribution of strengths associated to nodes

$$P(k) \to P(k), P(w), P(s)$$



Weighted networks

In many real networks, the global distribution of weights P(w) is very heterogeneous spanning several orders of magnitude



Relationship between the strength of a node and its degree

$$s(k) \sim k^{\beta}$$

Weights are locally correlated and non-trivially coupled to topology (β>1)

$$s_i = \langle w \rangle \Sigma_j \, a_{ij} = \langle w \rangle k_i$$

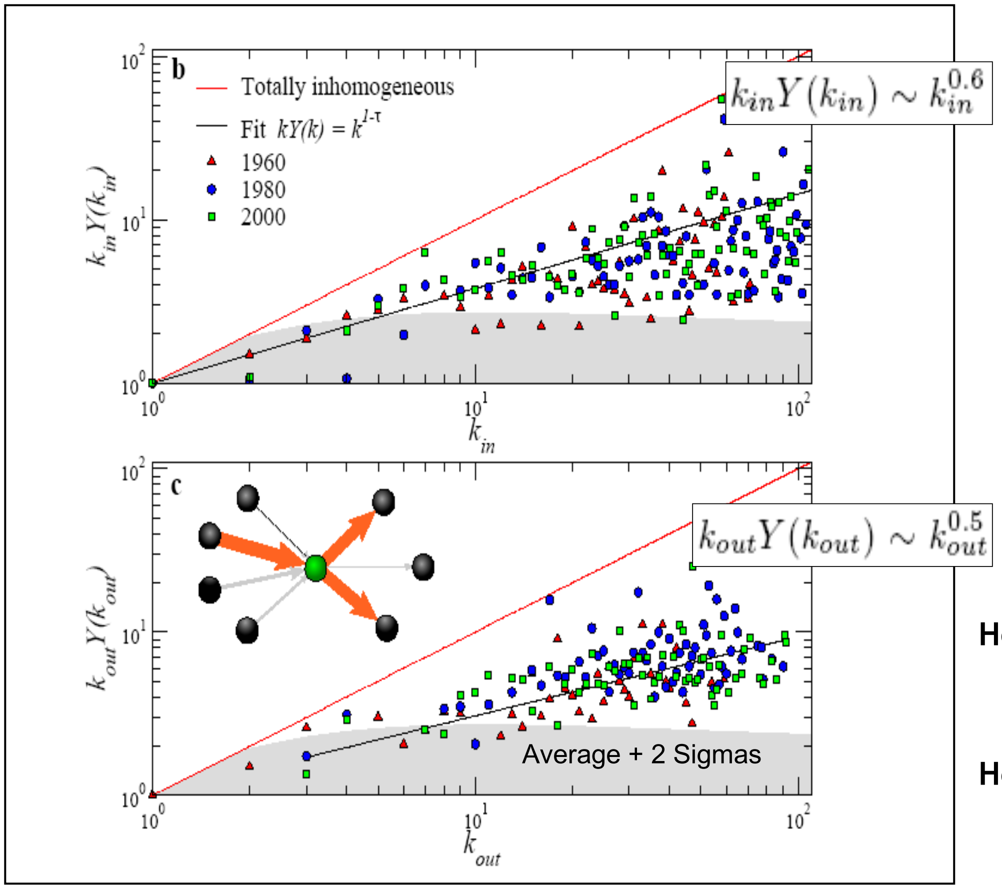
Absence of correlations between weights and degrees (β=1); weights do not add information

A. Barrat, M. Barthélemy, R. Pastor-Satorras, A. Vespignani. PNAS 101, 3747-3752 (2009).

Weighted networks

In many real networks, the local distribution of weights is also very heterogeneous

M. A. Serrano, M. Boguñá, A. Vespignani. Extracting the multiscale backbone of complex weighted networks. PNAS 106, 6483-6488 (2009).



$$Y_i(k) = \sum_{j=1,k} \left(\frac{w_{ij}}{s_i} \right)^2$$

Homogeneity $kY_i(k) \propto 1$

Heterogeneity $kY_i(k) \propto k$

Complex Systems

Complex Networks

STRUCTURAL PROP
Global
Distances
SW
Degree distribution
ANND
Clustering
K-cores
Centrality
Communities

DESCRIPTIO
Weighted
Directed
Bipartite
Multilayer

NETWORK
ER
CM + Rewirir
WS
BA
Structural cor
Null models

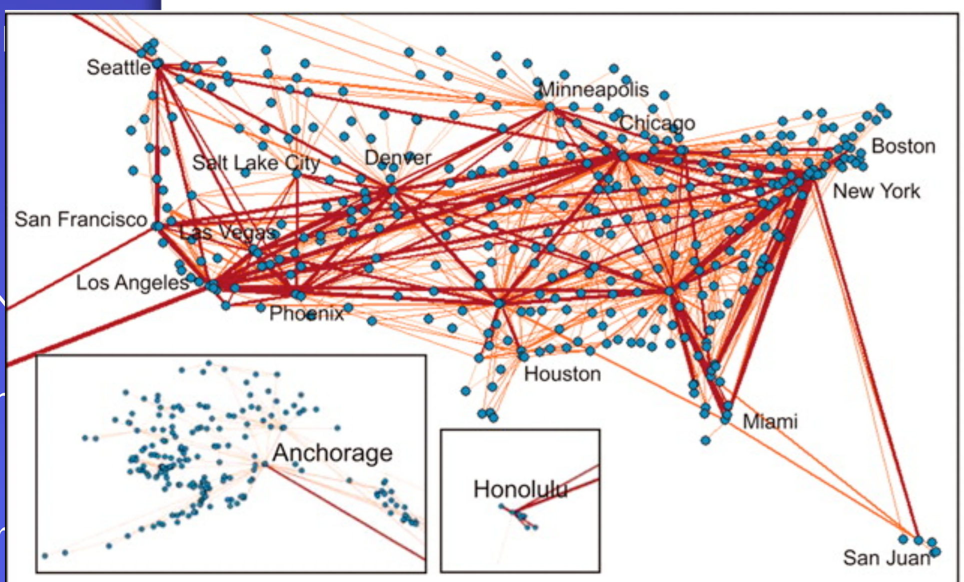
Weighted networks

Weight heterogeneities can be exploited to filter weighted networks, that are typically very dense

M. A. Serrano, M. Boguñá, A. Vespignani. Extracting the multiscale backbone of complex weighted networks. PNAS 106, 6483-6488 (2009).

The DP filter selects the links of a node with significant weights not expected by random fluctuations, those which concentrate most of the strength of the node

$$Y_i(k) = \sum_{j=1,k} \left(\frac{w_{ij}}{s_i} \right)^2$$



For a given normalized weight p_{ij}, the probability of having normalized weight larger or equal to p_{ij} in the null model, in which weights are locally distributed at random, is given by $\alpha_{ij} = (1 - p_{ij})^{k-1}$.

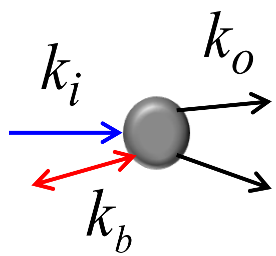
If α_{ij} is larger than a significance level α (between 0 and 1), the link will be filtered out. Changing α we can progressively remove irrelevant links thus effectively extracting the backbone of the weighted network.

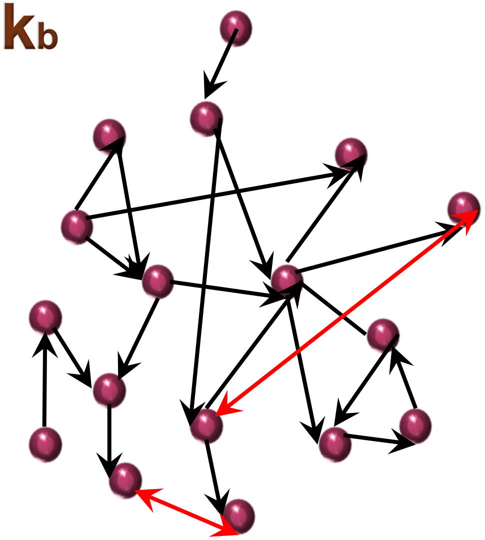
Directed networks

Links have an associated directionality, fro i to j, from j to i, or are bidirectional

$$P(k) \rightarrow P(k_i, k_o, k_b)$$

k → kin and kout and kb





Many times, bidirectional links are decoupled, and just $P(k_i)$ and $P(k_o)$ are considered (usually, one distribution is fat-tailed and the other is homogeneous...)

$$k_b P(\mathbf{k}) P_b(\mathbf{k}'|\mathbf{k}) = k_b' P(\mathbf{k}') P_b(\mathbf{k}|\mathbf{k}')$$

detailed balance conditions

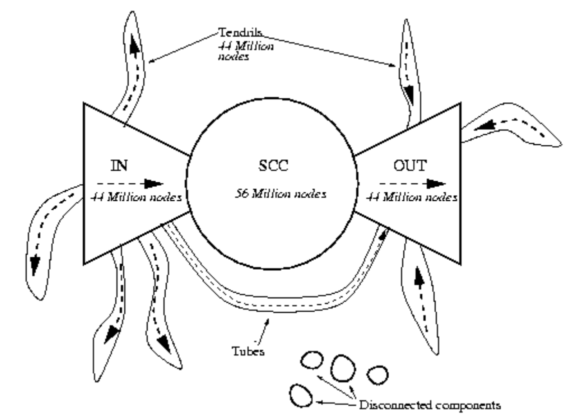
$$k_o P(\mathbf{k}) P_o(\mathbf{k}'|\mathbf{k}) = k_i' P(\mathbf{k}') P_i(\mathbf{k}|\mathbf{k}').$$

M. Boguñá and M. A. Serrano. Generalized percolation in random directed networks, Physical Review E 72, 016106 (2005).

Directed networks

Global connectivity structure – nodes – BOW TIE

- SCC (strongly connected component):
  - can reach all nodes from any other by following *directed* edges
- IN
  - can reach SCC from any node in 'IN' component by following directed edges
- OUT
  - can reach any node in 'OUT' component from SCC
- Tendrils and tubes
  - connect to IN and/or OUT components but not SCC
- Disconnected
  - isolated components



The bow-tie structure of the Web graph
A. Broder *et al.*, Computer Networks, 33:309–320, 2000.

Systems characterized by transport phenomena (of matter, energy, information….

Complex Systems

Complex Networks

STRUCTURAL PROP
Global
Distances
SW
Degree distribution
ANND
Clustering
K-cores
Centrality
Communities

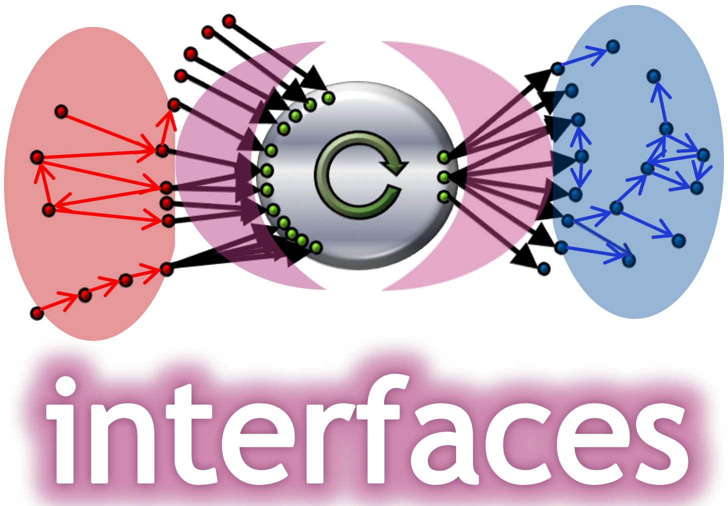
DESCRIPTIONS
Weighted
Directed
Bipartite
Multilayer

NETWORK MODELS
ER
SW
CM + Rewiring
BA
S1/H2
Null models

Directed networks

Global connectivity structure – edges – L BOW TIE

- The connected components are computed in terms of edges instead of nodes.
- The IN and OUT peripheral components are connected to the core SCC through interfaces, so that the L-Bow tie structure is characterized by five main components instead of three.
- The specific conformation of the interfaces could bring new light into the discussion of how structure is interwoven with functionality in transport networks.

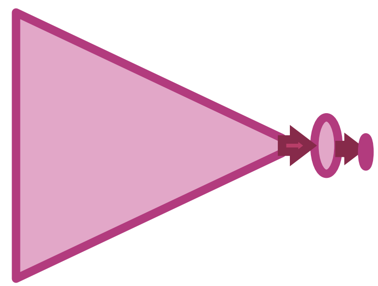


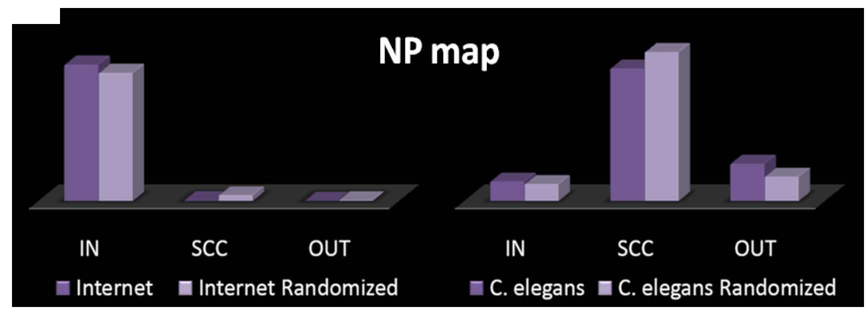
Phys. Rev. E 76, 056121 (2007)

Directed networks

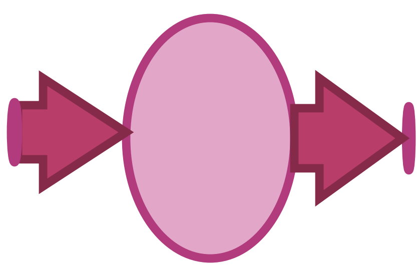
Global connectivity structure

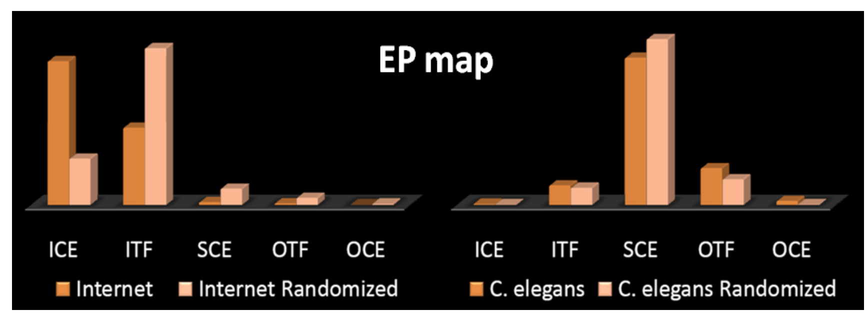
Internet customer-provider AS relationships (connections are payment for services provided)





Synaptic neuronal structure of C. Elegans (connections are chemical synapses)





M. A. Serrano and P. De Los Rios. Structural efficiency of percolated landscapes in flow networks. PLoS ONE 3(11): e3654 (2008).

BIPARTITE NETWORKS

Example: metabolic networks

(a)

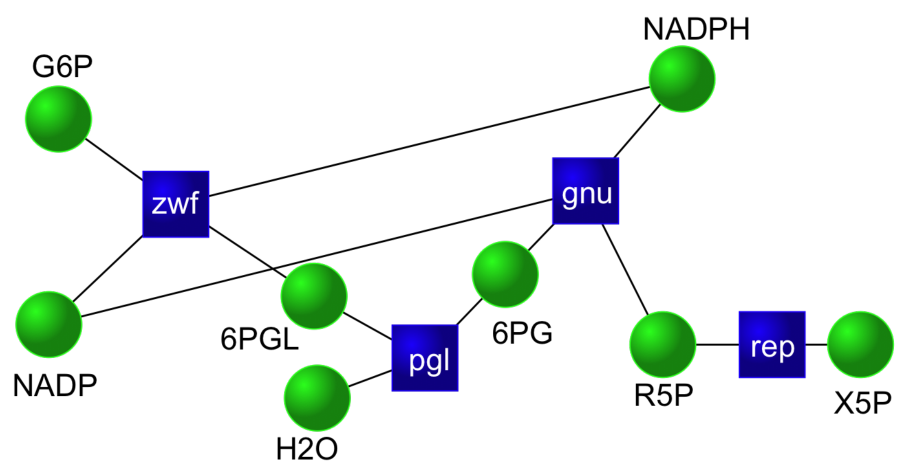
1 glucose 6-phosphate (G6P) + 1 NADP$^+$ $\xrightarrow{zwf}$ 1 6-phosphoglucono δ-lactone (6PGL) + 1 NADPH

1 6-phosphoglucono δ-lactone + 1 H_2O $\xrightarrow{pgl}$ 1 6-phosphogluconate (6PG)

1 6-phosphogluconate + 1 NADP$^+$ $\xrightarrow{gnd}$ 1 ribulose 5-phosphate (R5P) + 1 NADPH

1 ribulose 5-phosphate $\xleftrightarrow{rpe}$ 1 xylulose 5-phosphate (X5P)

four coupled equations in the pentose-phosphate pathway of E. coli



Complex Systems

Complex Networks

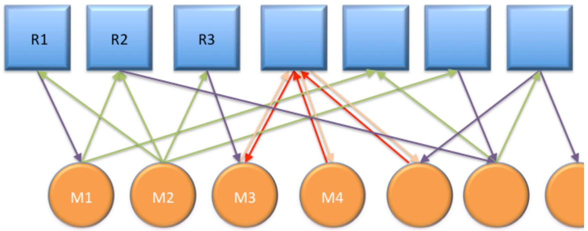
STRUCTURAL PROP
Global
Distances
SW
Degree distribution
ANND
Clustering
K-cores
Centrality
Communities

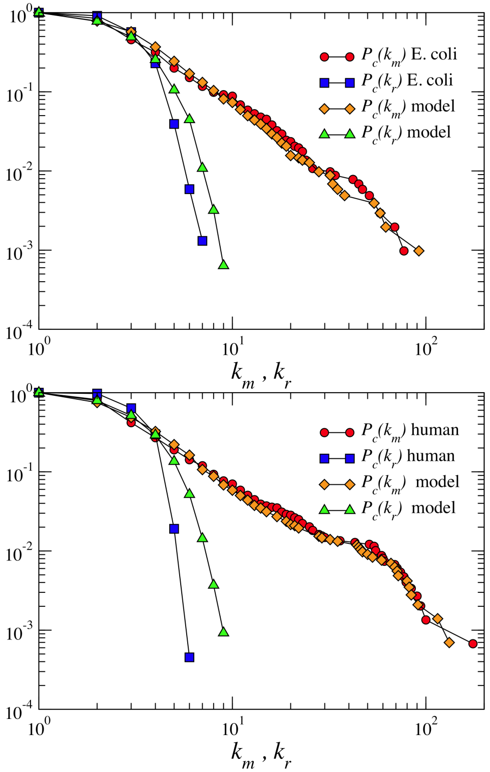
DESCRIPTIONS
Weighted
Directed
Bipartite
Multilayer

NETWORK MODELS
ER
SW
CM + Rewiring
BA
S1/H2
Null models

BIPARTITE NETWORKS

Bipartite networks: two types of nodes without connections in the same class





Real bipartite networks are typically characterized by a scale-free degree distribution for nodes in one class and a bounded/homogeneous distribution for nodes in the other

M. A. Serrano, M. Boguñá, F. Sagués, Uncovering the hidden geometry behind metabolic networks. Molecular BioSystems, 8, 843-850 (2012).

Complex Systems

Complex Networks

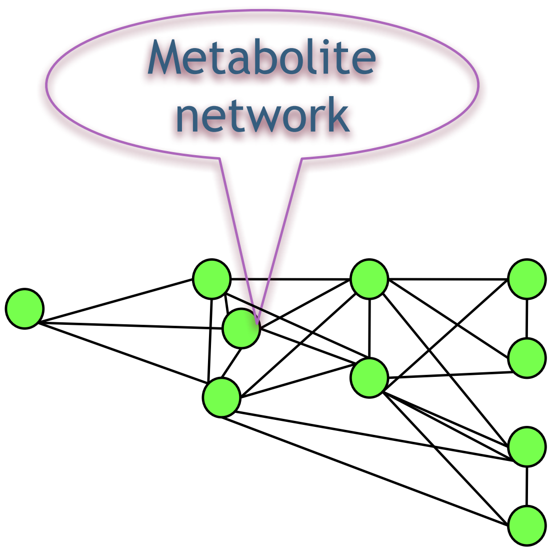
STRUCTURAL PROP
Global
Distances
SW
Degree distribution
ANND
Clustering
K-cores
Centrality
Communities

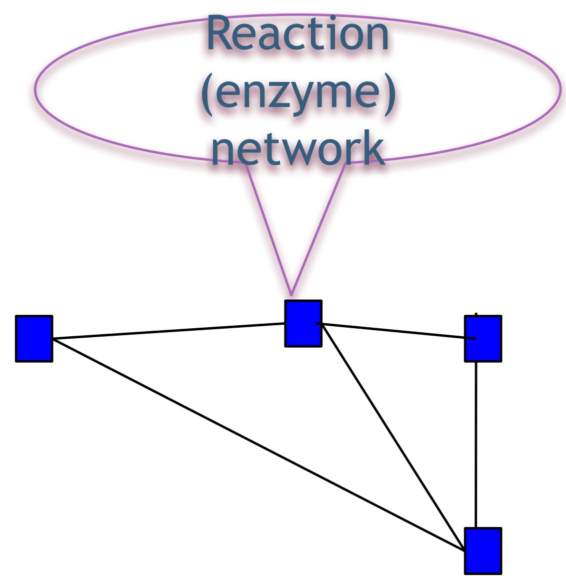
DESCRIPTIONS
Weighted
Directed
Bipartite
Multilayer

NETWORK MODELS
ER
SW
CM + Rewiring
BA
S1/H2
Null models

BIPARTITE NETWORKS

One-mode projections (weights can be added – number of neighbors shared)





Two metabolites connected if they share a reaction

Two reactions/enzymes connected if they have common metabolites

Complex Systems

Complex Networks

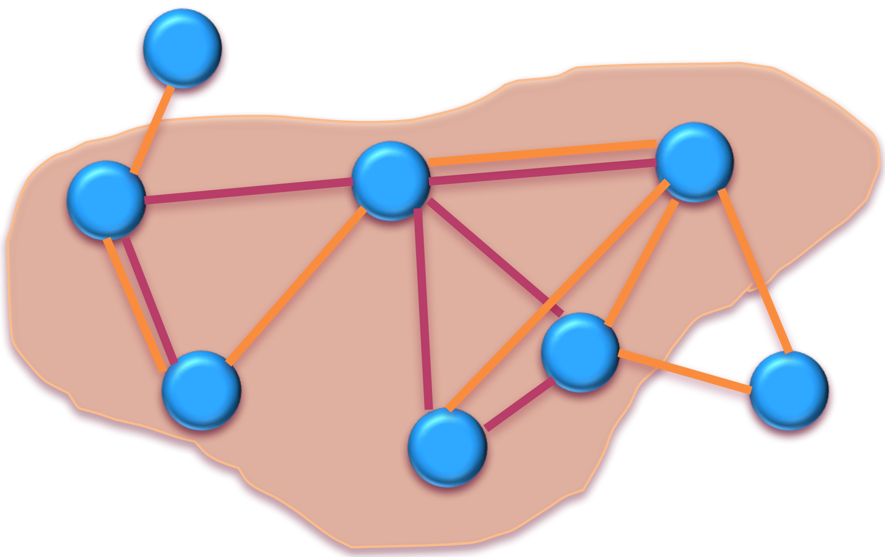
STRUCTURAL PROP
Global
Distances
SW
Degree distribution
ANND
Clustering
K-cores
Centrality
Communities

DESCRIPTIONS
Weighted
Directed
Bipartite
Multilayer

NETWORK MODELS
ER
SW
CM + Rewiring
BA
S1/H2
Null models

MULTILAYER NETWORKS

Multiplex



Complex Systems

Complex Networks

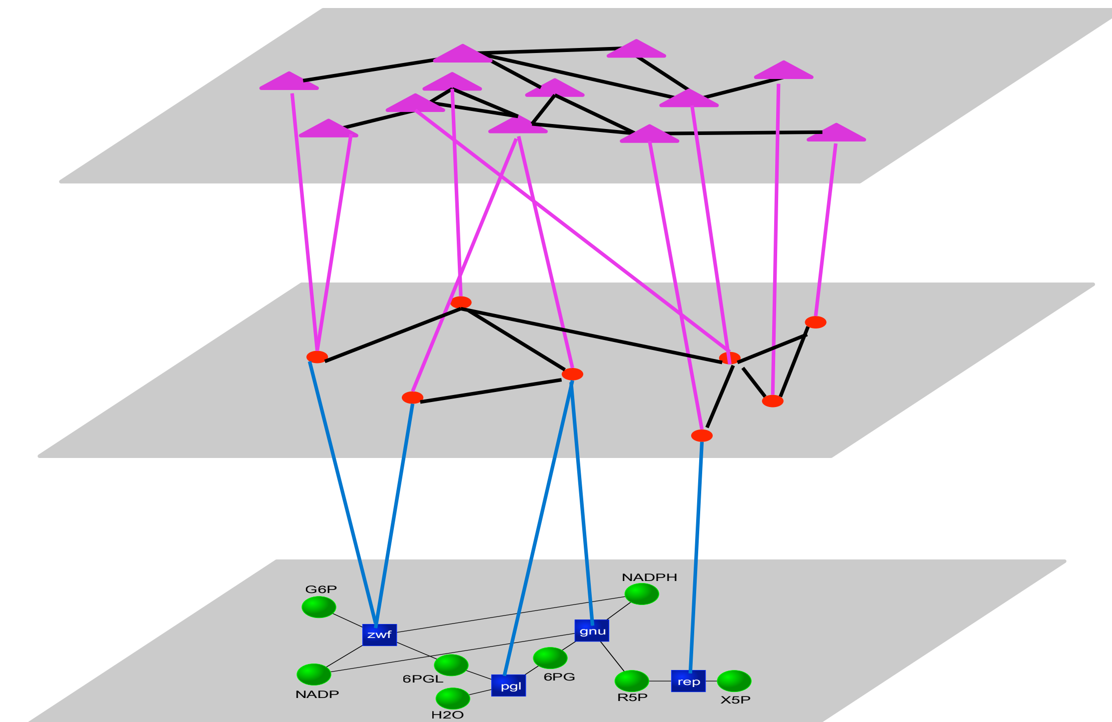
STRUCTURAL PROP
Global
Distances
SW
Degree distribution
ANND
Clustering
K-cores
Centrality
Communities

DESCRIPTIONS
Weighted
Directed
Bipartite
Multilayer

NETWORK MODELS
ER
SW
CM + Rewiring
BA
S1/H2
Null models

MULTILAYER NETWORKS

Interconnected



Complex Systems

Complex Networks

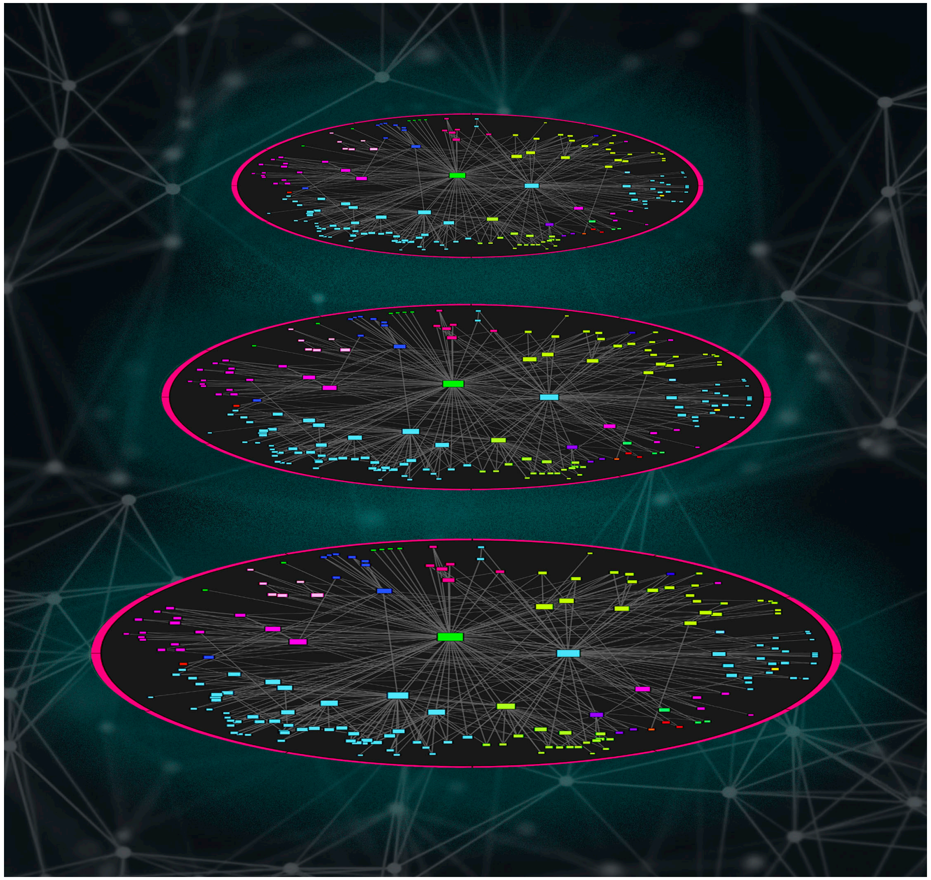
STRUCTURAL PROP
Global
Distances
SW
Degree distribution
ANND
Clustering
K-cores
Centrality
Communities

DESCRIPTIONS
Weighted
Directed
Bipartite
Multilayer

NETWORK MODELS
ER
SW
CM + Rewiring
BA
S1/H2
Null models

MULTILAYER NETWORKS

Multiscale



NETWORK MODELS

Set of rules to interconnect a number of elements in order to reproduce specific features (but always SWs)

- **Equilibrium network models**

  The number of nodes is fixed to N

  - Classical random graphs, Erdös and Rényi model
  - Watts-Strogatz model
  - Configuration model

- **Non-equilibrium network models**

  The number of nodes N grows

  - Classical random growing grahs
  - Preferential attachment, Barabasi-Albert model

Complex Systems

Complex Networks

STRUCTURAL PROP
Global
Distances
SW
Degree distribution
ANND
Clustering
K-cores
Centrality
Communities

DESCRIPTIONS
Weighted
Directed
Bipartite
Multilayer

NETWORK MODELS
ER
SW
CM + Rewiring
BA
S1/H2
Null models

EQUILIBRIUM MODELS: ER

Classical random graphs, Erdös and Rényi

- **(1) Each pair of nodes in the network are connected with probability p**

  Grand canonical statistical ensamble of graphs with equal probability (standard statistical definition: temperature and chemical potential constant)

- **(2) A given number of links E connect randomly chosen pairs of vertices.** Self-connections and multiple connections may happen unless a specific restriction to avoid them is added

  Microcanonical statistical ensamble of graphs with equal probability (standard statistical definition: energy and number of particles constant)

In the large graph limit (thermodynamic limit) both constructions are equivalent

Complex Systems

Complex Networks

STRUCTURAL PROP
Global
Distances
SW
Degree distribution
ANND
Clustering
K-cores
Centrality
Communities

DESCRIPTIONS
Weighted
Directed
Bipartite
Multilayer

NETWORK MODELS
ER
SW
CM + Rewiring
BA
S1/H2
Null models

EQUILIBRIUM MODELS: ER

Classical random graphs, Erdös and Rényi

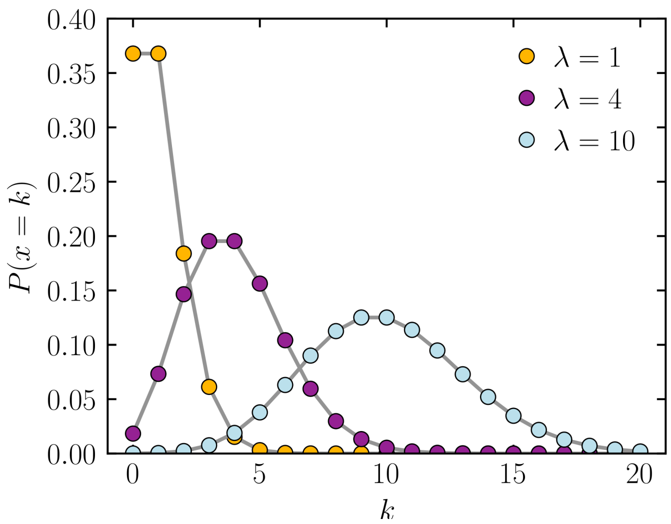
Degree distribution: Binomial,
for large N (p small, sparse limit) it takes the Poisson form

$$P(k) = \binom{N-1}{k} p^k (1-p)^{N-1-k}$$

$$\overline{k} = p(N-1)$$

$$P(k) = e^{-\overline{k}}\, \overline{k}^{\,k} / \, k!$$

$$E = pN(N-1)/2$$



Homogeneous
No hubs

EQUILIBRIUM MODELS: ER

Classical random graphs, Erdös and Rényi

Degree distribution: Binomial,
for large N (p small, sparse limit) it takes the Poisson form

$$P(k) = \binom{N-1}{k} p^k (1-p)^{N-1-k}$$ Type equation here.

$$\bar{k} = p(N-1)$$

$$P(k) = e^{-\bar{k}}\,\bar{k}^{k} / \,k!$$

$$E = pN(N-1)/2$$

Uncorrelated by definition

$$P(k'|k) = \frac{k' P(k')}{\langle k \rangle} \Rightarrow \bar{k}_{nn}(k) = \sum_{k'} k' P(k'|k)$$

$$P(k'|k) = \frac{k' N_{k'}}{2E}$$

$$\bar{k}^{unc}_{nn}(k) = \frac{\langle k^2 \rangle}{\langle k \rangle} = \frac{<k>^2 + <k>}{<k>} = <k> + 1$$

EQUILIBRIUM MODELS: ER

Classical random graphs, Erdös and Rényi

Degree distribution: Binomial,
for large N (p small, sparse limit) it takes the Poisson form

$$P(k) = \binom{N-1}{k} p^k (1-p)^{N-1-k}$$

$$P(k) = e^{-\bar{k}}\, \bar{k}^{k} / \, k!$$

$$\bar{k} = p(N-1)$$

$$E = pN(N-1)/2$$

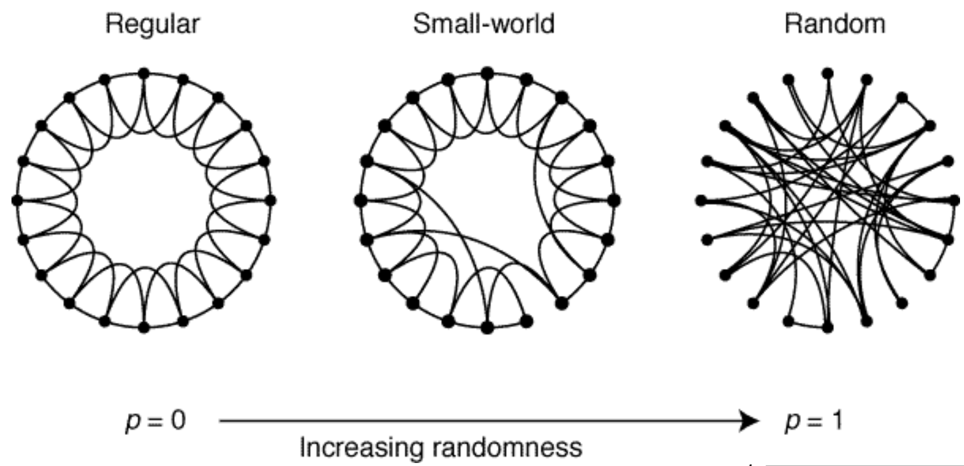
Uncorrelated by definition

$$P(k'|k) = \frac{k' P(k')}{\langle k \rangle} \Rightarrow \bar{k}_{nn}^{unc}(k) = \frac{\langle k^2 \rangle}{\langle k \rangle} = \frac{<k>^2 + <k>}{<k>} = <k> + 1$$

$$C = \frac{(\text{number of triangles})}{(\text{number of connected triples})} = \frac{\binom{n}{3} p^3}{\binom{n}{3} p^2} = p = \frac{\langle k \rangle}{n-1}$$

$\Rightarrow$ Locally tree-like

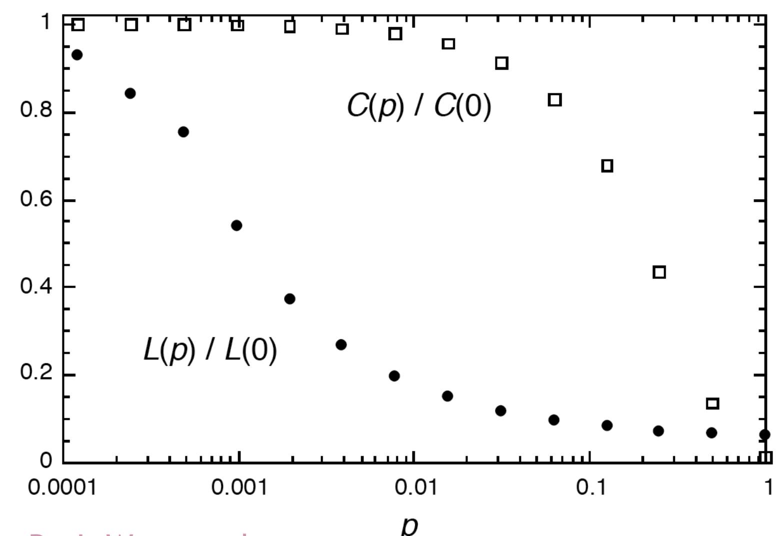
EQUILIBRIUM MODELS: SW



random network
Low clustering
Small-world property

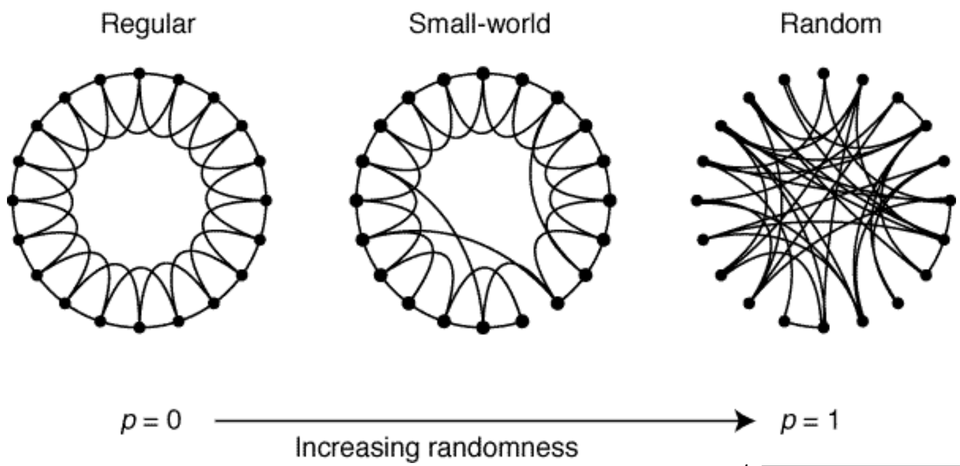
regular ring lattice
High clustering
High average path length

small worlds even with small concentrations of shortcuts



Collective dynamics of 'small-world' networks, D. J. Watts and S. H. Strogatz, Nature 393, 440-442 (1998)

EQUILIBRIUM MODELS: SW

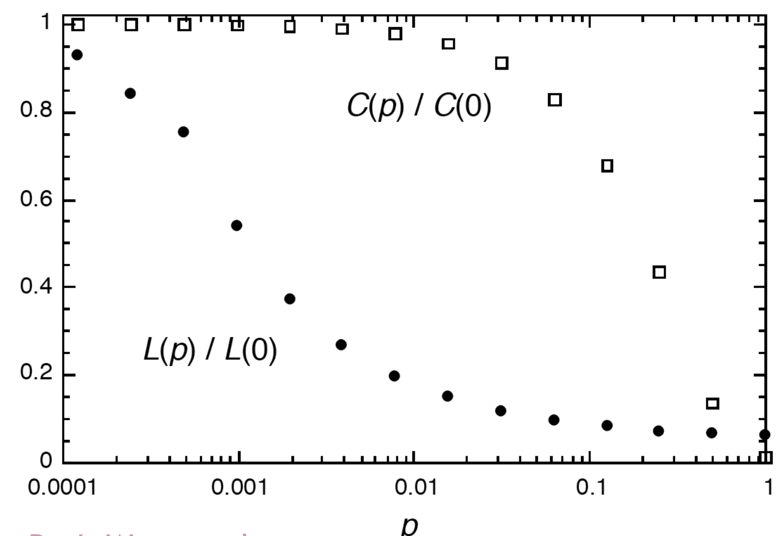


random network
Low clustering
Small-world property

ASSIGNMENT 4

Reproduce the SW figure

DELIVERABLE: commented code + report

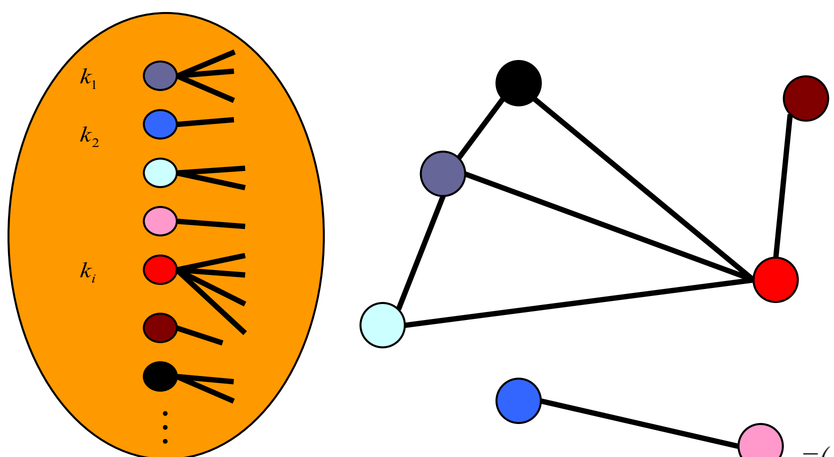


Collective dynamics of 'small-world' networks, D. J. Watts and S. H. Strogatz, Nature 393, 440-442 (1998)

EQUILIBRIUM MODELS: CM

The **configuration model** is usually used as a null model for real complex networks. It produces maximally random networks with a preassigned degree sequence (uncorrelated except for structural unavoidable correlations necessary to close the networks)

$$P(k) \qquad k_1 \quad k_2 \quad \cdots \quad k_i \quad \cdots \quad k_N$$



Avoiding multiple and self-connections

ANND and clustering (except for structural correlations)

$$\frac{\langle k^2 \rangle}{\langle k \rangle}$$

$$\bar{c}(k) = C = \frac{1}{N}\frac{(\langle k^2\rangle - \langle k\rangle)^2}{\langle k\rangle^3}\ , \ k > 1$$

For weighted networks

M. A. Serrano and M. Boguñá. Weighted Configuration Model, AIP Conference Proceedings 776 (1), 101-107 (2005).

M. Molloy and B. Reed, Random Struct. Algorithms 6, 161 (1995)
M. Molloy and B. Reed, Combinatorics, Probab. Comput. 7, 295 (1998)

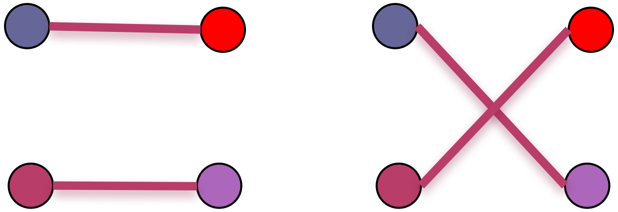
EQUILIBRIUM MODELS: RW

Alternatively, one could apply a random rewiring process to links in order to destroy correlations (again, structural correlations necessary to close the network are not avoidable). If one wants to preserve the degree distribution:

i) select two different links in the network

ii) exchange one end of one of the two links for one of the ends of the second link (avoiding multiple and self-connections)

iii) repeat at least E times



Different possibilities for weighted networks

M. A. Serrano. Rich-club vs rich-multipolarization phenomena in weighted networks. Physical Review E 78, 026101 (2008).

Structural correlations

Scale-free degree distributions - finite size

Natural cut-off : The expected maximum value of the degree (random variable) given the degree distrution and that the network is of finite size N

Definition: It is calculated taking into account that the number of nodes with degrees greater than the cut-off has to be of the order of 1

$$N \int_{k_{cut}}^{\infty} dk\, P(k) \sim 1 \qquad k_{\max} = \min\left(N-1, k_0 N^{1/(\gamma-1)}\right)$$

Power law degree distributions are observable only in large networks!!!

The same scaling is found from extreme value theory: if we draw N observations (random variables) from the degree distribution (many times), the maximum value of the sample will also be a random variable, the natural cut-off is the average value of the maximum (increasing with N)

Boguna, M; Pastor-Satorras, R; Vespignani, A;
EUROPEAN PHYSICAL JOURNAL B 38, 205-209 (2004)

Structural correlations

Structural cut-off for the degree distribution (for uncorrelated networks):

In real networks, the degrees are not drawn independetly from P(k), but must satisfy some topological constraints due to network structure and closeness

$$k_s(N) \sim (\langle k \rangle N)^{1/2}$$

The structural cut-off is calculated by imposing that the probabiliy that a link extist between two nodes of degrees k and k' is less or equal to 1 (physical constraint)

$$r_{kk'} = \frac{kk'}{\langle k \rangle N}$$

Boguna, M; Pastor-Satorras, R; Vespignani, A; EUROPEAN PHYSICAL JOURNAL B 38, 205-209 (2004)

Structural correlations

$$k_c(N) \sim N^{1/(\gamma-1)} \qquad k_s(N) \sim (\langle k \rangle N)^{1/2}$$

The structural cut-off vs the natural cut-off

- The structural cut-off coincides with the natural cut-off when γ=3 (Barabasi-Albert networks)

- If γ>3, the structural cut-off diverges faster than the natural one and the later dominates

- If γ<3, the natural cut-off diverges faster than the structural one and the later dominates. If the actual cut-off is imposed to be larger than the structural cut-off, then some negative two-point correlations (or multiple connections) are necessary to fulfill network closeness contraints

Structural correlations!!!

Boguna, M; Pastor-Satorras, R; Vespignani, A;
EUROPEAN PHYSICAL JOURNAL B 38, 205-209 (2004)

COMPLEX NETWORKS

ASSIGNMENT 5

Given a list of degrees, make a computer program that generates CM random networks with the given degrees

Use the list of degrees calculated in ASSIGNMENT 1

- Calculate 1 CM random network instance
- Compute topological properties with program in ASSIGNMENT 2
- Calculate 100 CM random network realizations
- Compute the average values over realizations of the topological properties with program in ASSIGNMENT 2

DELIVERABLE: commented code + report with a discussion of the results in comparison with results of ASSIGNMENT 2

GROWING MODELS: BA

- **At each time step, add one vertex to the network and attach it to a randomly chosen vertex** Growing exponential network with exponential degree distribution

- **At each time step, add one vertex to the network and attach it to m randomly chosen vertices with probability proportional to their degree** Preferential attachment, BA model, power-law degree distributions

Growth alone or preferential attachment alone is not sufficient to produce scale-free degree distributions

A little bit of history about preferential attachment

First use to explain power-law distributions by Yule in 1925.

The modern master equation method was applied by H. A. Simon in 1955 in the course of studies of the sizes of cities and other phenomena.

It was first applied to the growth of networks by D. de Solla Price in 1976, who was interested in the networks of citation between scientific papers.

Complex Systems

Complex Networks

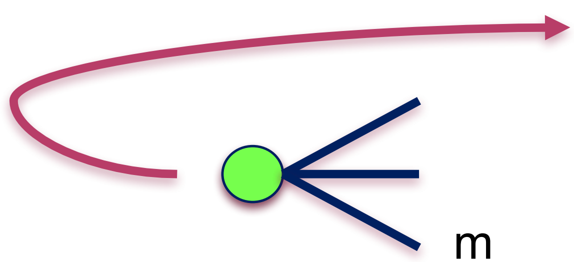
STRUCTURAL PROP
Global
Distances
SW
Degree distribution
ANND
Clustering
K-cores
Centrality
Communities

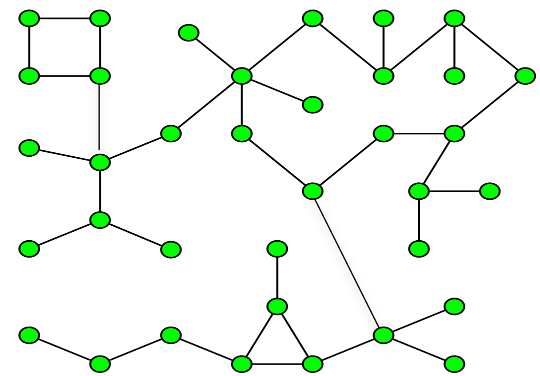
DESCRIPTIONS
Weighted
Directed
Bipartite
Multilayer

NETWORK MODELS
ER
SW
CM + Rewiring
BA
S1/H2
Null models

GROWING MODELS: BA

- At each time step, add one vertex to the network and attach it to m randomly chosen vertices with probability proportional to their degree





Time t = Number of nodes N

Probability that a node i of degree k_i attracts a new edge

Probability that a new edge attaches to a vertex of degree k

Probability that m new edges attach to (m different) vertices of degree k

Complex Systems

Complex Networks

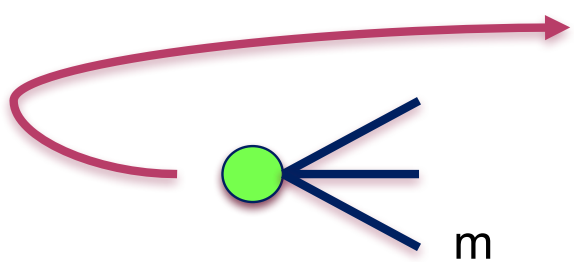
STRUCTURAL PROP
Global
Distances
SW
Degree distribution
ANND
Clustering
K-cores
Centrality
Communities

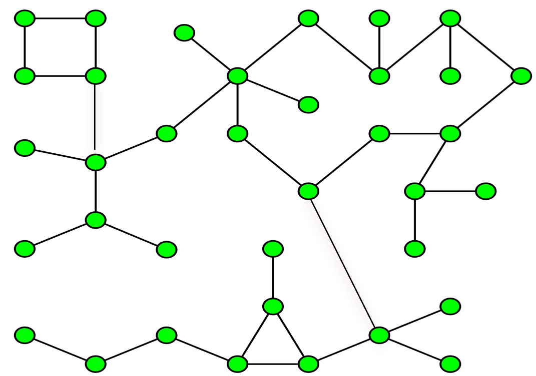
DESCRIPTIONS
Weighted
Directed
Bipartite
Multilayer

NETWORK MODELS
ER
SW
CM + Rewiring
BA
S1/H2
Null models

GROWING MODELS: BA

- At each time step, add one vertex to the network and attach it to m randomly chosen vertices with probability proportional to their degree





Time t = Number of nodes N

Probability that a node i of degree k_i attracts a new edge

$$\frac{k_i}{\sum k_{i,N}} = \frac{k_i}{2E_N} = \frac{k_i}{2mN}$$

Probability that a new edge attaches to a vertex of degree k

$$\frac{kp_k}{\sum_k kp_k} = \frac{kp_k}{2m}$$

Probability that m new edges attach to (m different) vertices of degree k

$$m \times kp_k/2m = \frac{1}{2}kp_k$$

GROWING MODELS: BA

- At each time step, add one vertex to the network and attach it to m randomly chosen vertices with probability proportional to their degree

Probability that a new edge attaches to a vertex of degree k

$$\frac{kp_k}{\sum_k kp_k} = \frac{kp_k}{2m}$$

Master equation for the probability of nodes having degree k at t+1

time evolution of the probability of a system to occupy each of a discrete set of states

time evolution for p_k

GROWING MODELS: BA

- **At each time step, add one vertex to the network and attach it to m randomly chosen vertices with probability proportional to their degree**

Probability that a new edge attaches to a vertex of degree k

$$\frac{kp_k}{\sum_k kp_k} = \frac{kp_k}{2m}$$

Master equation for the probability of nodes having degree k at t+1

$$N_{k,N+1} = N_{k,N} + m\,(k-1)\,p_{k-1,N}/2m - m\,k\,p_{k,N}/2m$$

$$(n+1)p_{k,n+1} - np_{k,n} = \tfrac{1}{2}(k-1)p_{k-1,n} - \tfrac{1}{2}kp_{k,n},$$

for $k > m$, or

added node

$$(n+1)p_{m,n+1} - np_{m,n} = 1 - \tfrac{1}{2}mp_{m,n},$$

for $k = m$, and there are no vertices with $k < m$.

GROWING MODELS: BA

- At each time step, add one vertex to the network and attach it to m randomly chosen vertices with probability proportional to their degree

$$p_{k,n+1} = p_{k,n} = p_k$$ Stationary solutions

GROWING MODELS: BA

- **At each time step, add one vertex to the network and attach it to m randomly chosen vertices with probability proportional to their degree**

$$p_{k,n+1} = p_{k,n} = p_k$$ Stationary solutions

$$p_k = \frac{(k-1)(k-2)\dots m}{(k+2)(k+1)\dots(m+3)} p_m = \frac{2m(m+1)}{(k+2)(k+1)k}$$

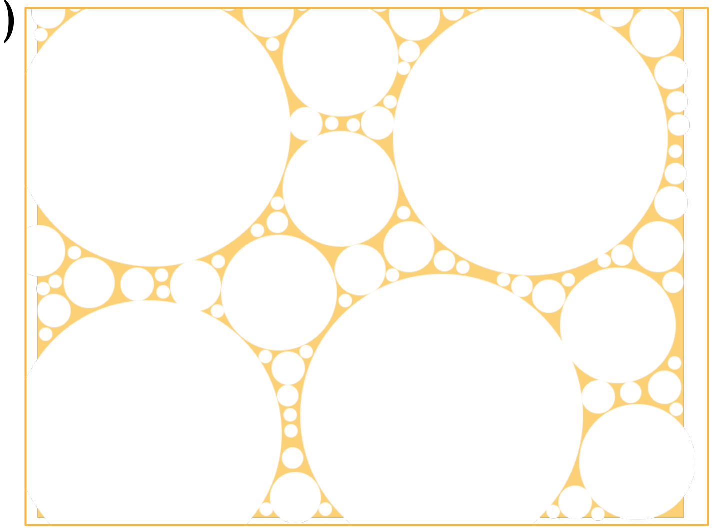
In the limit of large degrees $p_k \sim k^{-3}$

Other features

- Correlation between age and degree
- Very weak two point disassortative degree correlations that vanish in the thermodynamic limit
- Clustering vanishes in the thermodynamic limit

OTHER METHODS TO GENERATE PLS

The Swiss cheese model
(holes vs radius)



combination of exponentials

$$p_Y(y) \sim e^{ay}$$

$$x = e^{by}$$

$$p_X(x) \sim \frac{e^{a \ln x^{1/b}}}{bx} = \frac{1}{bx^{1-a/b}}$$

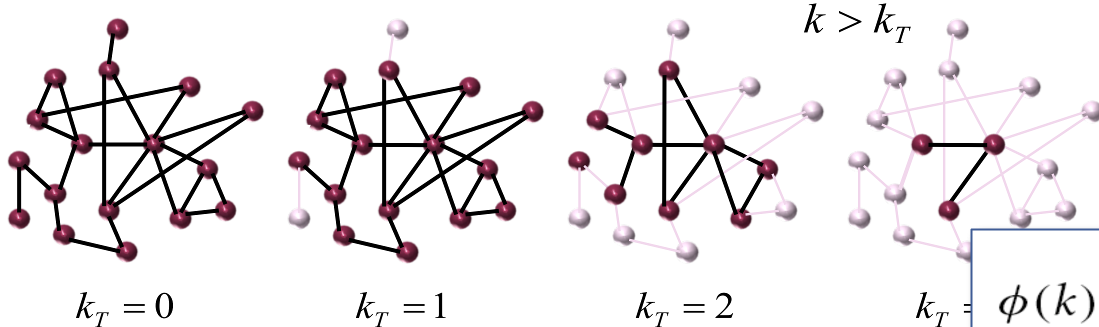
And more...

IMPORTANCE OF NULL MODELS

Rich club phenomenon

Zhou, S. & Mondragon, R. J. The rich-club phenomenon in the Internet topology. *IEEE Commun. Lett.* **8,** 180–182 (2004).

$k > k_T$

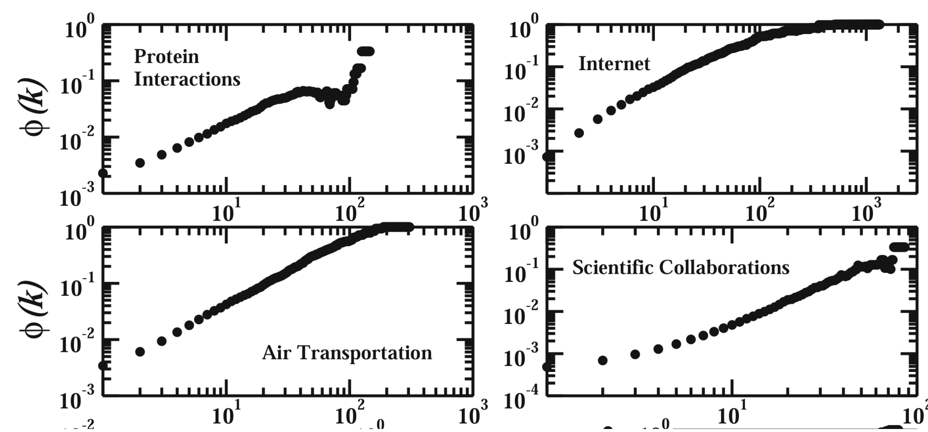


$$\phi(k) = \frac{2E_{>k}}{N_{>k}(N_{>k}-1)}$$

[1]This nested hierarchy of subgraphs turns out to have self-similarity properties for some real scale-free networks such as the Internet at the autonomous system level; see [7].

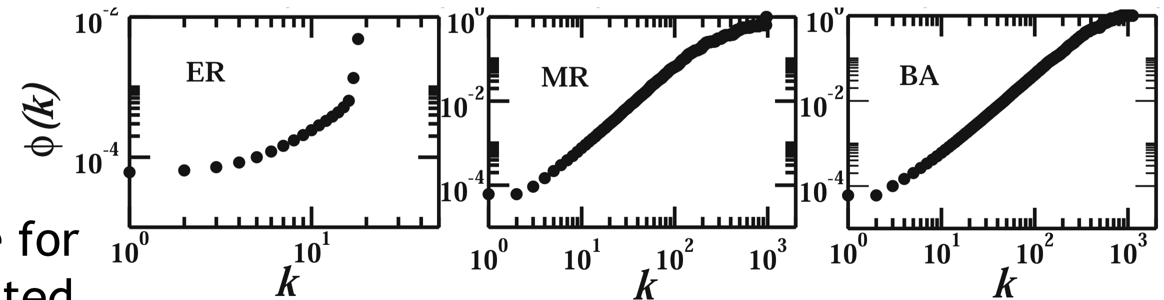
M. A. Serrano, D. Krioukov, and M. Boguñá, Phys. Rev. Lett. **100**, 078701 (2008).

Original interpretation: the growing behavior denotes the presence of rich clubs



IMPORTANCE OF NULL MODELS

Rich club phenomenon



The same for uncorrelated networks!!

$$P_{\mathrm{unc}}(k,k') = kk'P(k)P(k')/\langle k\rangle^2$$

$$\phi(k) = \frac{N\langle k\rangle \int_k^{k_{\max}} \mathrm{d}k' \int_k^{k_{\max}} \mathrm{d}k'' P(k',k'')}{\left[N\int_k^{k_{\max}} \mathrm{d}k' P(k')\right]\left[N\int_k^{k_{\max}} \mathrm{d}k' P(k') - 1\right]}$$

$$\boxed{\phi_{\mathrm{unc}}(k) = \frac{1}{N\langle k\rangle}\left[\frac{\int_k^{k_{\max}} \mathrm{d}k' k' P(k')}{\int_k^{k_{\max}} \mathrm{d}k' P(k')}\right]^2 \underset{k,k_{\max}\to\infty}{\sim} \frac{k^2}{\langle k\rangle N}}$$

V. Colizza, A. Flammini, M. A. Serrano, and A. Vespignani. Detecting rich-club ordering in complex networks, Nature Physics 2, 110-115 (2006).
M. A. Serrano. Rich-club vs rich-multipolarization phenomena in weighted networks. Physical Review E 78, 026101 (2008).

IMPORTANCE OF NULL MODELS

Rich club phenomenon

Compare $\phi(k) = \frac{2E_{>k}}{N_{>k}(N_{>k}-1)}$ with $\phi_{\text{expected}}(k)$

Structural correlations have to be discounted!!!

$$\rho_{\text{unc}}(k) = \phi(k)/\phi_{\text{unc}}(k)$$

$$\phi_{\text{unc}}(k) = \frac{1}{N\langle k\rangle}\left[\frac{\int_k^{k_{\max}} \mathrm{d}k' k' P(k')}{\int_k^{k_{\max}} \mathrm{d}k' P(k')}\right]^2 \underset{k, k_{\max}\to\infty}{\sim} \frac{k^2}{\langle k\rangle N}$$

Not feasible when structural correlations are present $P(k) \approx k^{-\gamma}$

$$\rho_{\text{ran}}(k) = \phi(k)/\phi_{\text{ran}}(k)$$

Maximally random network (MRN) with the same P(k)Configuration model or rewiring

$\rho_{\text{ran}}(k) > 1$ ⇨ RC effect

V. Colizza, A. Flammini, M. A. Serrano, and A. Vespignani. Detecting rich-club ordering in complex networks, Nature Physics 2, 110-115 (2006).
M. A. Serrano. Rich-club vs rich-multipolarization phenomena in weighted networks. Physical Review E 78, 026101 (2008).

Complex Systems

Complex Networks

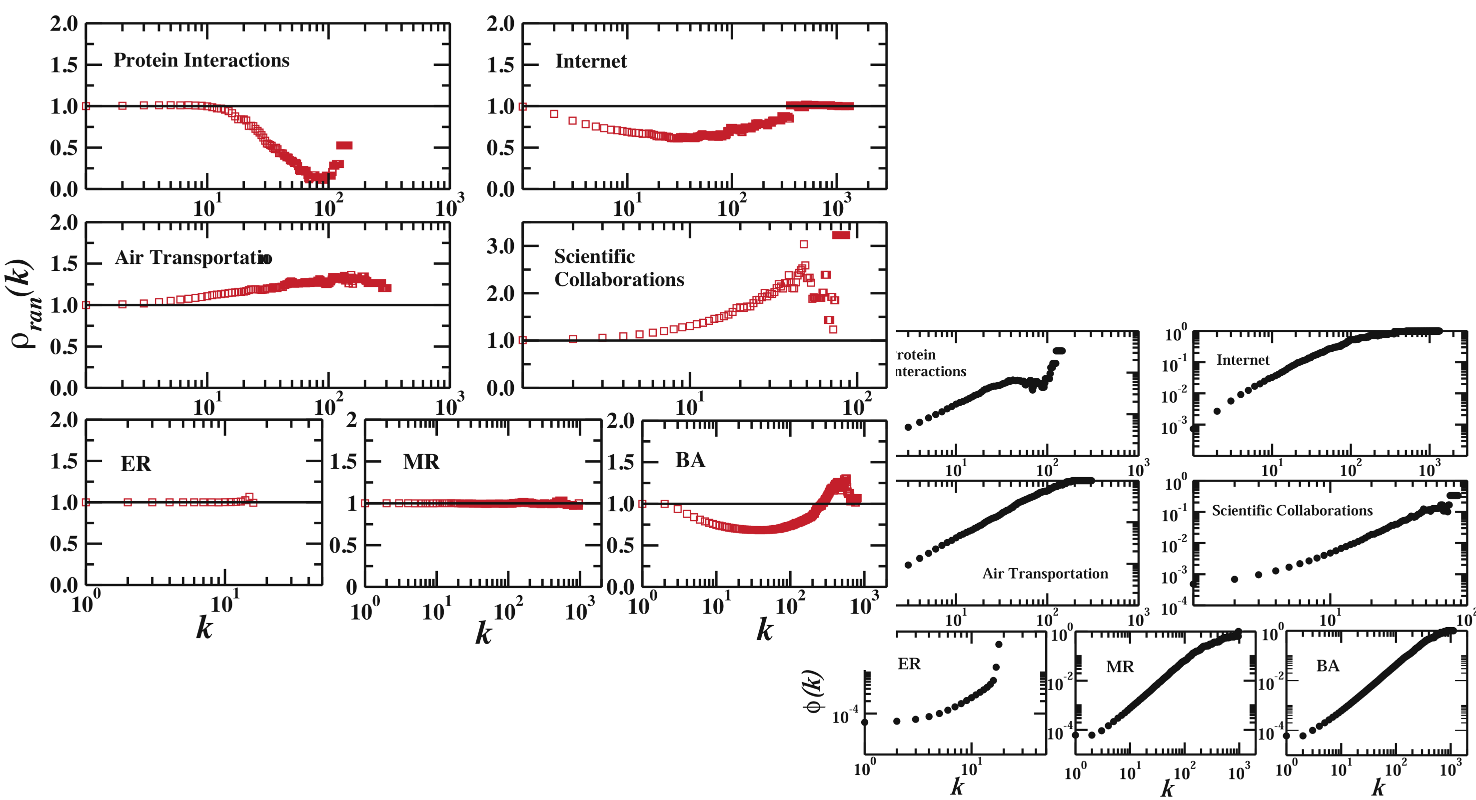
STRUCTURAL PROP
Global
Distances
SW
Degree distribution
ANND
Clustering
K-cores
Centrality
Communities

DESCRIPTIONS
Weighted
Directed
Bipartite
Multilayer

NETWORK MODELS
ER
SW
CM + Rewiring
BA
S1/H2
Null models

IMPORTANCE OF NULL MODELS

Rich club phenomenon



V. Colizza, A. Flammini, M. A. Serrano, and A. Vespignani. Detecting rich-club ordering in complex networks, Nature Physics 2, 110-115 (2006).
M. A. Serrano. Rich-club vs rich-multipolarization phenomena in weighted networks. Physical Review E 78, 026101 (2008).