What are complex networks?

What is their architecture?

What is their nature?

What is their function?

How do we study, optimize, and use them?

S. N. Dorogovtsev

Departamento de Física da Universidade de Aveiro & I3N,
Ioffe Institute, St. Petersburg

Literature:

[1] M. E. J. Newman, *Networks: An Introduction* (Oxford University Press, 2010).

[2] A.-L. Barabási, *Network Science* (Cambridge University Press, 2016).

[3] S. N. Dorogovtsev and J. F. F. Mendes, *Evolution of Networks: From the Biological Nets to the Internet and WWW* (Oxford University Press, 2003).

[4] S. N. Dorogovtsev, *Lectures on Complex Networks* (Oxford University Press, 2010).

[5] G. Bianconi, *Multilayer Networks* (Oxford University Press, 2018).

R&D spending:

2016	
US	511B US$
China	452B US$, PPP
Japan	166B US$, PPP

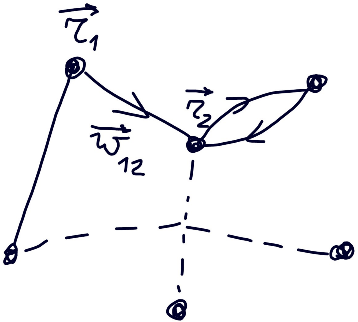
AI, 2017	
US	$\sim$ 10B US$
China	$\sim$ 7B US$, PPP ?
Japan	$\sim$ 0.7B US$, PPP ?

Total energy consumption of data centers:
between 1% and 2% of the world energy consumption

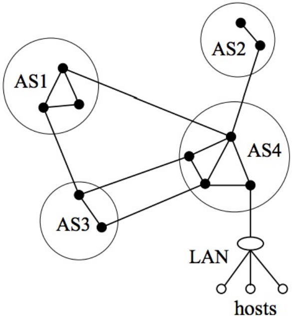
Where the money is:

Big data — control and manipulation

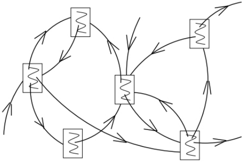
The “NSA” graph

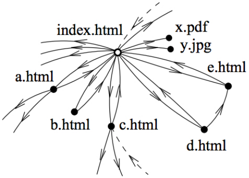


The Internet: the router level, the AS level



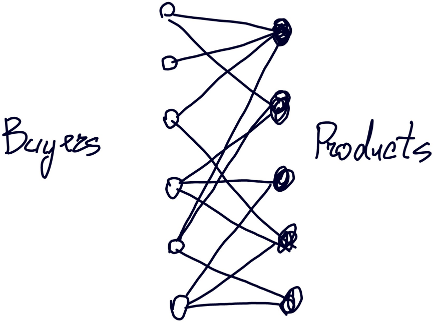
The WWW



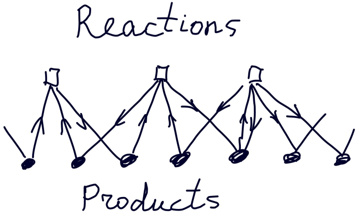


Large social networks — everything, from where you can get data; sociology

Recommendation networks

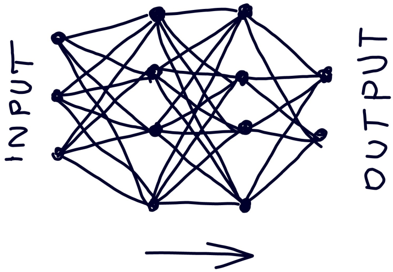


Biology:
Chemical reaction networks, market ∼1B US$



Genetic regulatory networks
Power grids
Transportation networks

Deep learning



Brain networks

Simple vs Complex

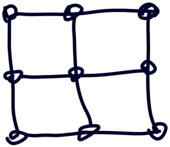
- Progress: simple $\longrightarrow$ complex
- Regress: complex $\longrightarrow$ simple

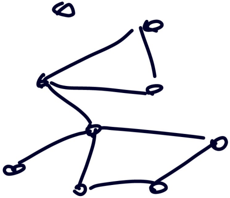
Complexity: mathematical meaning — the time of exact solution $t(N)$

Complex networks: more complex than classical random graphs
The term coined by Laszló Barabási

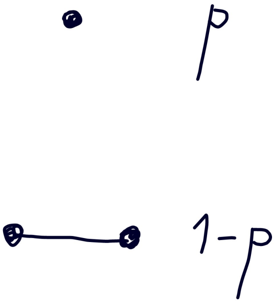
Searching for simple solutions of complex systems
— it's not easy

Random or deterministic



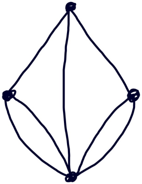


Random graph



History:

Leonhard Euler, 1735, Königsberg bridge problem

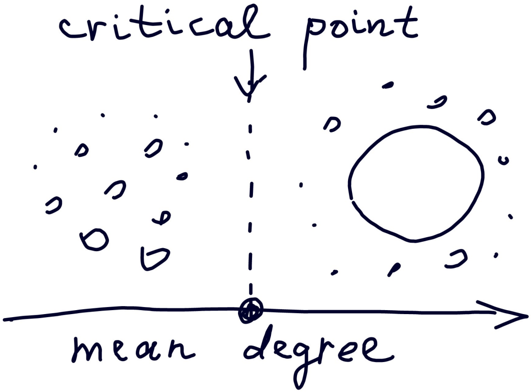



GEORGE UDNY YULE

George Udny Yule, $\sim$1910
Explaining heavy tails of distributions in nature

Ray Solomonoff

Anatol Rapoport
1950, discovery of the giant connected component transition


critical point
mean degree

Paul Erdős

Alfréd Rényi
1950s, Erdős–Rényi random graphs

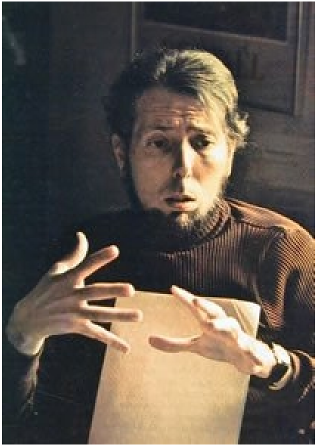
Herbert Alexander Simon
1950s, theory of fat-tailed distributions in nature

Paul Baran

Donald Watts Davies

∼1960, architecture of the Internet

Derek John de Solla Price
~1960+, generated growing scale-free networks



Stanley Milgram
1967, small worlds, six degrees of separation

Mark Granovetter
$\sim$ 1970, the strength of weak ties, sociology

Per Bak
1980s, self-organized criticality

Trends

- Sampling
- Inference
- Reconstruction

Deep learning

- Batch size:

if — small, then fast training, high accuracy, low generalization

- Overfitting

Exploding/vanishing gradients

- Dependence on initial conditions

Treating random systems in mathematics and statistical mechanics:

- Statistical ensemble:

the full set of possible configurations of a system
together with
their statistical weights ($\sim$ probabilities of realization)

Classical random graphs:

$G_{n,p}$ model vs ER model

Canonical ensemble vs microcanonical ensemble

Equivalence in the infinite size limit

Violation of the equivalence in other models (the configuration model vs Chung-Lu)

Locally tree-like graphs when sparse.

The birth of the giant connected component at $\langle q \rangle = 1$.

Equilibrium trees vs recursive trees

Equilibrium trees: $d_H = \infty, \quad \langle \ell \rangle \sim \ln N$

Recursive trees: $d_H = 2, \quad \langle \ell \rangle \sim N^{1/2}$

Both have Poisson degree distributions

Measuring the “Hausdorff” dimension:

$$\langle \mathcal{N}(\ell) \propto \ell^{d_H}$$

Example: measuring d_H and d_S of evolving networks of triangulations — strongly constrained planar graphs.

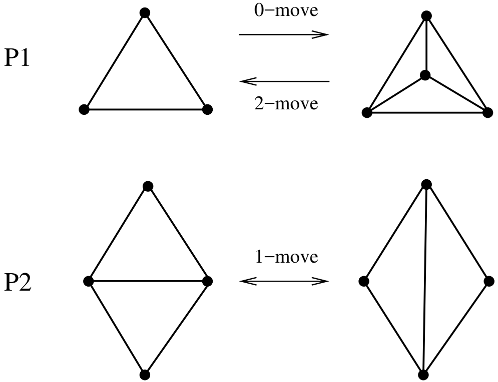
Evolution of triangulation networks:

• local structure — ???
(degree distribution, degree–degreecorrelations)

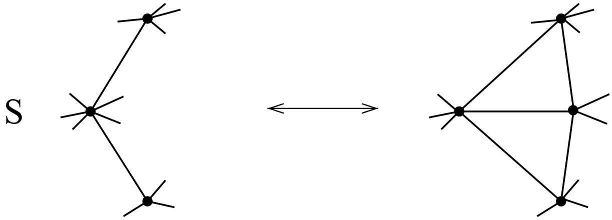
• space dimension — ???
(Hausdorff and spectral dimensions)

Triangular mesh operations—1:

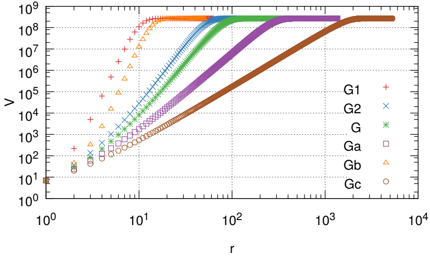
Pachner moves:

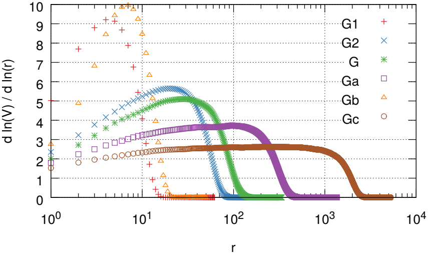


Triangular mesh operations—2:



Hausdorff dimension:

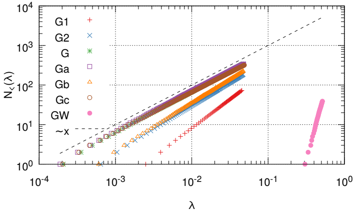


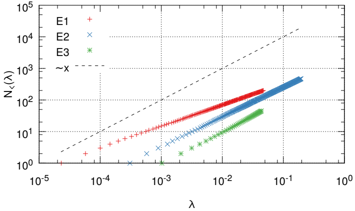


Spectral dimension:

$$\varphi_0(t) \sim t^{-d_S/2}, \qquad \rho(\lambda) \sim \lambda^{d_S/2-1}$$

$$d_S \leq d_H$$





E1 vs. equilibrium random trees

Model E1:
random addition and addition vertices of degree 3 with equal rates.

$$d_H \sim 2(?), \qquad d_S = 1.4(2).$$

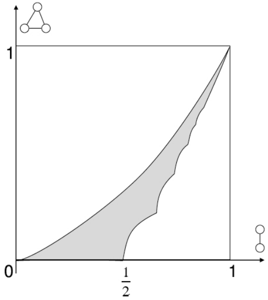
Equilibrium random trees:

$$d_H = 2, \qquad d_S = 4/3.$$

vs typical random planers who's $d_H = 4$.

Triangles

Density of triangles vs density of edges:



$$d_2 \equiv \frac{N_2}{N(N-1)/2}, \qquad d_3 \equiv \frac{N_3}{N(N-1)(N-2)/6}$$

$$d_3 \leq d_2^{3/2}$$

Clustering of classical random graphs:

$$C = \overline{C} = \frac{\langle q \rangle}{N} = p$$

$$N_3 = \langle q \rangle^3 / 6$$

For classical random graphs:

$$d_3 = d_2^3$$

Clustering coefficient C vs average clustering $\overline{C}$:

$$C(k) \equiv \frac{\langle l_{\text{nn}}(k) \rangle}{k(k-1)/2}$$

$$\overline{C} \equiv \sum_k P(k) C(k)$$

$$C \equiv \frac{\sum_k P(k) \langle l_{\text{nn}}(k) \rangle}{\sum_k P(k) k(k-1)/2} = \frac{\sum_k k(k-1) P(k) C(k)}{\langle k^2 \rangle - \langle k \rangle}$$

Generalizing classical random graphs

The configuration model (random graph with a given degree sequence) — a micro-canonical ensemble

Models with hidden variables, Chung-Lu model (random graph with a desired degree sequence) — a canonical ensemble

Uncorrelated networks

Local tree-likeness of these models if the networks are sparse

The problem of non-equivalence

Degree distribution of the nearest neighbour

Let the degree distribution of an uncorrelated network be $P(q)$

Then the degree distribution of a neighbour of a randomly chosen vertex is

$$\frac{qP(q)}{\langle q \rangle}.$$

This is also the distribution of an end vertex of a randomly chosen edge.

The joint distribution of the degrees of the ends of an edge is

$$P(q, q') = \frac{qP(q)}{\langle q \rangle} \frac{q'P(q')}{\langle q \rangle}.$$

Average branching $\langle b \rangle = \dfrac{\langle q^2 \rangle}{\langle q \rangle} - 1.$

Ultra-small worlds if $\langle q^2 \rangle \to \infty$.

The Molloy–Reed criterion for the existence of the giant connected component, $\langle b \rangle > 1$, i.e.,

$$\langle q^2 \rangle > 2\langle q \rangle.$$

Size S of the giant connected component in uncorrelated tree-like nets:

$$X = \sum_q P(q) X^q \equiv G(X),$$

where $G(X)$ is the generating function of $P(q)$.

$$1 - S = \sum_q \frac{qP(q)}{\langle q \rangle} X^{q-1} = \frac{G'(X)}{\langle q \rangle},$$

Super-resilience if $\langle q^2 \rangle \to \infty$
Degree distribution cutoffs

Laplace operator

$$D: \quad d_{ij} = q_i \delta_{ij}.$$

$$L = D - A, \quad L_{ij} = q_i \delta_{ij} - a_{ij}.$$

$$\frac{d\varphi_i(t)}{dt} = -\sum_j L_{ij} \varphi_j(t)$$

$$\frac{d\varphi_i}{dt} = -\sum_{j \in i} (\varphi_i - \varphi_j)$$

Spectral dimension d_S:

$$\varphi_0(t) \propto t^{-d_S/2}$$

$$\rho_L(\lambda) \propto \lambda^{d_S/2-1}$$

$$d_S \leq d_H$$

Modularity:

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

$k_i = \sum_j A_{ij}, \quad m = \frac{1}{2} \sum_{ij} A_{ij}$.
— comparison with the null model $\equiv$ the configuration model.
Q for a given set of communities
$\equiv$
(the fraction of edges that fall within the communities)
$-$
(the expected fraction of edges within these communities for the configuration model with the same node degrees).

$$-1 < Q < 1$$

Bad points:

• Q can be nonzero for graphs even without communities (trees);

• resolution limit;

• overlapping communities;

• a difficult optimization problem.

Stochastic block model:

The main test tool for community detection algorithms

k types of vertices — k blocks in a network, $\alpha = 1, ..., k$

q_α, $\alpha = 1, .., k$, is the probability that a vertex $\in$ block α.

If vertex $i \in$ block α and vertex $j \in$ block β,
then they are connected with the probability $p_{\alpha\beta}$.

c_{in} is the total fraction of edges within blocks.
c_{out} is the total fraction of edges between blocks.

According to the communities detection algorithm times,

$k \leq 4$:

detectable — undetectable

detectable if $\frac{c_{\text{out}}}{c_{\text{in}}} < \epsilon_c$

(ϵ_c is the point of the second order phase transition)

undetectable if $\frac{c_{\text{out}}}{c_{\text{in}}} > \epsilon_c$

$k > 4$:

detectable — hard detectable — undetectable

(the first order phase transition).

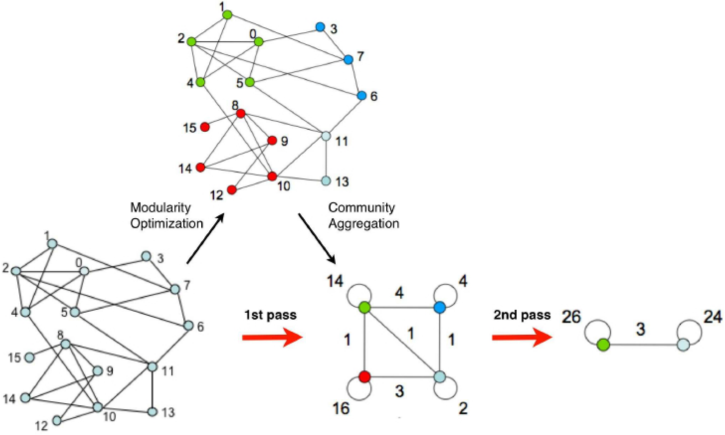
Spectral algorithms for community detection

Studying 1st and 2nd eigenvectors of A or L, etc.
Works well for dense graphs where the localization of the components of eigenvectors on hubs is absent.
This localization is due to returns of a random walk to hubs in sparse networks.
How to avoid it?
Study the eigenvectors of the matrix B which describes a non-backtracking walk:

$$B_{i\to j,k\to l} = \delta_{jk}(1 - \delta_{il}).$$

It reduces the problem of localization.

Greedy detection of communities in large nets:



V. D. Blondel, J.-L. Guillaume, R. Lambiotte, E. Lefebvre (2008)

• modularity is optimized by allowing only local changes of communities;

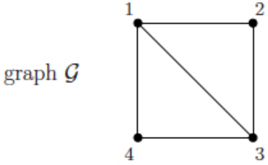
• the found communities are aggregated forming a new network of communities;

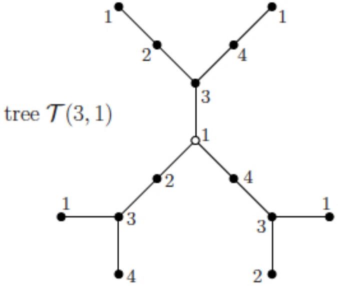
• repeat ... ...

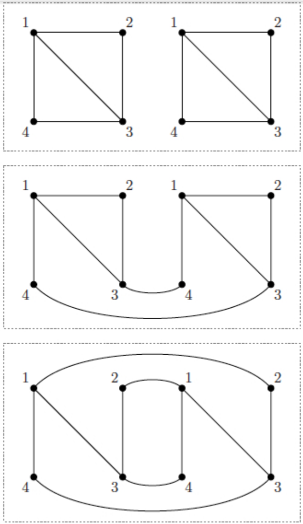
The simplicity of the algorithm is due to the fact that at each step all its communities are single-node.

• For each node i consider all $j \in i$ and evaluate the gain of Q that would take place by removing i from its community to the community of j.
• The node i is then placed in the community for which this gain is maximum, but only if this gain is positive.
• If no positive gain is possible, i stays in its original community.
• This process is applied repeatedly and sequentially for all nodes until no further improvement can be achieved.

Treating networks with short cycles:







The average branching $\langle b \rangle = \lambda_1$ of the matrix B

Centrality index for optimal percolation:

$$\sum_{j \in \partial i} v_{i \leftarrow j}^{(\lambda_1)} v_{j \leftarrow i}^{(\lambda_1)}$$

Strong heterogeneity produces non-mean-field critical exponents:

Upper critical dimension

Uncorrelated scale-free networks

Super-resilience if $\gamma \leq 3$ ($\langle q \rangle \to \infty$)

Growing networks with a BKT phase transition

The Ising model on complex networks

Non-equilibrium statistical mechanics:

$$\partial_t P(g,t) = \sum_{g' \in G} [W(g,g')P(g',t) - W(g',g)P(g,t)]$$

Random recursive graphs

At each time step, attach the new vertex to uniformly randomly selected vertex/vertices.

$$P(q,t) = \frac{1}{t}\langle N(q,t)\rangle.$$

$$\langle N(q,t{+}1)\rangle = \langle N(q,t)\rangle + \frac{1}{t}[\langle N(q{-}1,t)\rangle - \langle N(q,t)\rangle] + \delta_{q,1}.$$

$$(t+1)P(q,t+1) - tP(q,t) = P(q-1,t) - P(q,t) + \delta_{q,1}.$$

$$P(q,t) + t\partial_t P(q,t) \xrightarrow{t\to\infty} P(q) = P(q-1) - P(q) + \delta_{q,1}.$$

$$P(q) = 2^{-1-q},$$

i.e, an exponential degree distribution.

Preferential attachment

Proportional preference:
at each time step, attach the new vertex to a vertex selected with the probability $\frac{q}{t\langle q\rangle} = \frac{q}{2t}$.

$$N(q,t+1)\rangle=\langle N(q,t)\rangle+\frac{1}{2t}[(q-1)\langle N(q-1,t)\rangle-q\langle N(q,t)\rangle]+\delta_{q,1}$$

$$(t+1)P(q,t+1)-tP(q,t)=(q-1)P(q-1,t)-qP(q,t)+\delta_{q,1}.$$

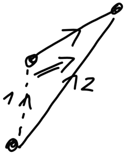
$$P(q) = (q-1)P(q-1) - qP(q) + \delta_{q,1}.$$

$$P(q) \propto q^{-3},$$

$$\text{Prob}(i) \propto q_i + A \Longrightarrow P(q) \propto q^{-\gamma},\ 2 < \gamma \leq \infty$$

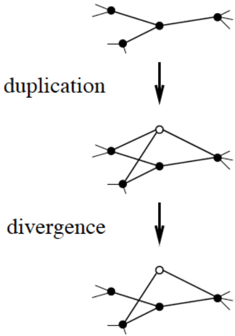
Link copying:

A directed network:



Choose uniformly at random a vertex (1) and make a link to its descending neighbour (2).
This produces (approximately) the proportional preferential attachment (Prob $\propto q_{\text{in}}$).

Duplication mechanism:



The white vertex is a new protein.
The duplication mechanism effectively produces the proportional preferential attachment.

k-cores:

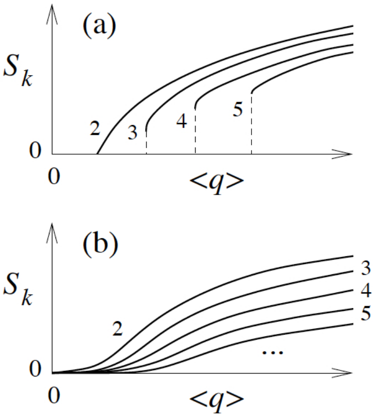
- The k-core of a network is its largest subgraph whose nodes have at least k connections (within this subgraph).

- The local pruning algorithm:

Remove from a graph all nodes of degree less than k.

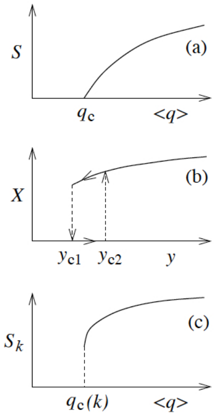
Some of the remaining nodes may occur with less than k links. Prune these nodes, and so on.

The final result, if it exists, is the k-core.



k-core sets in uncorrelated networks.
(a) $\langle q \rangle < \infty$.
(b) $\langle q \rangle \to \infty$.

Hybrid transition: $S \cong S_c + C\sqrt{\langle q\rangle - q_c}$

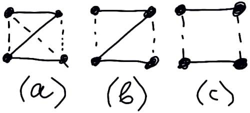


(a) Continuous, (b) first order, (c) hybrid phase transitions. HPT is the limit of stability of the FOPT.

Multilayer networks

Multiplex networks, interdependent networks, nets of nets.

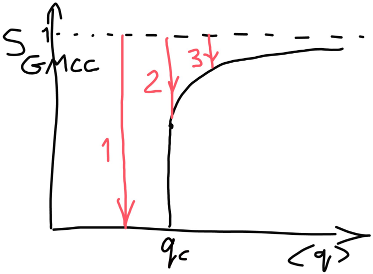
Three percolation problems for multiplex networks:



The giant analogies of (a), (b) are the results of global pruning; of (c) is the result of local pruning.

A giant mutually connected component in interdependent networks corresponds to (a).

Cascade of failures in interdependent networks:



1 — the pruning takes a finite time,
2 — power-law (infinite) relaxation,
3 — exponential (infinite) relaxation.

Google PageRank:

$$r_i = \frac{p}{N} + (1-p) \sum_{j \in \overrightarrow{\partial} i} \frac{r_j}{q_{\text{out},j}},$$

$$p = 0.15$$

Uncorrelated networks:

$$\langle r \rangle (q_{\text{in}}, q_{\text{out}}) = \frac{p}{N} + \frac{1-p}{N} \frac{q_{\text{in}}}{\langle q_{\text{in}} \rangle}.$$

Prisoner's dilemma:

Each of two players independently decides for himself which is better:

to cooperate, C, or to defect, D?

The decision is based on the set of payoffs for the players:

Players:	Their payoffs:
(C, C)	$(1, 1)$
(C, D)	$(0, b)$
(D, C)	$(b, 0)$
(D, D)	$(0, 0)$

$$b > 1, \quad \text{so defect}$$

Evolutionary games on networks:

Many players, can adopt/imitate strategies of their neighbours.
$\sigma_i = C, D$ is the state/strategy of player i
An adaptive player: adopts the most successful strategy in his close environment (the player himself and his nearest neighbours).

spatial prisoners dilemma

Stochastic version:

• The initial concentration of cooperators $c = 1/2$.

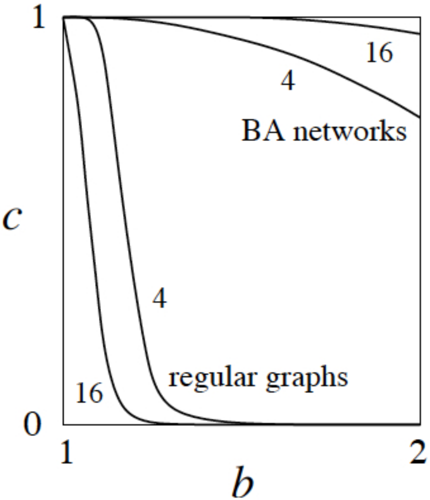
• Then all of the pairs of nearest neighbours play the game independently, and after this round, each player i accumulates his payoffs as P_i.

• After that, for each player i, choose at random one of its neighbours, j, and compare the scores P_i and P_j.

(a) If $P_i > P_j$, then leave σ_i unchanged.

(b) Otherwise, let player i accept the strategy of j with some probability.

• Then, pass to the next round, recalculate all the scores and so on.


1
16
4
BA networks
c
4
regular graphs
16
0
1
2
b

Biased imitation in coupled evolutionary games:

Layer 1: PD, layer 2: Snowdrift Game

Players imitate neighbours in their layer with prob p, and neighbours from the other layer with prob $1-p$.

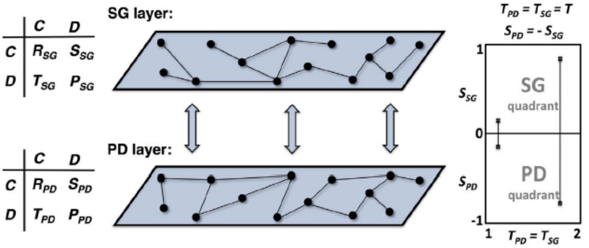
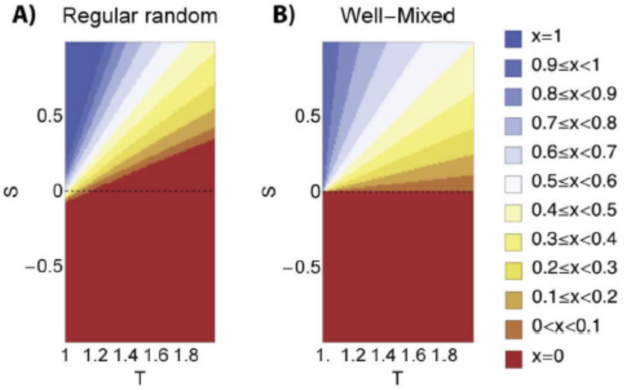
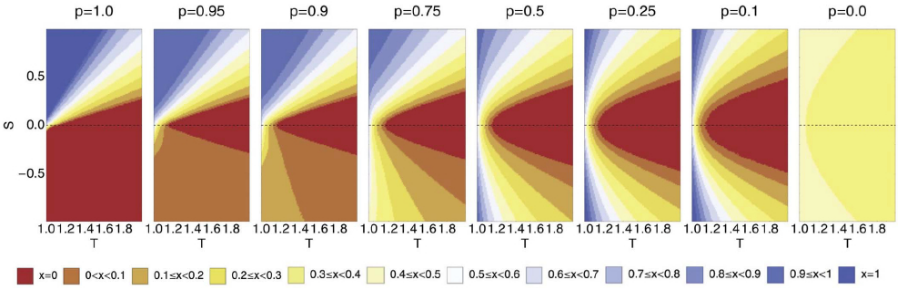


Figure 1 | Scheme of the population structure. The population is organized in two layers. Individuals establish *intralinks* with neighbors of the same layer, and *interlinks* with neighbors of the opposite layer. Depending on the layer in which they are located, individuals compute their payoff taking into account the PD payoff matrix or the SG payoff matrix, indicated on the left of the figure. On the right, we schematize the relation between the parameters of the two games, which we use in this paper. We assume that $T_{PD} = T_{SG} = T$, $S_{SG} = S$ and $S_{PD} = -S$, with $S \in [0, 1]$, $T \in [1, 2]$.

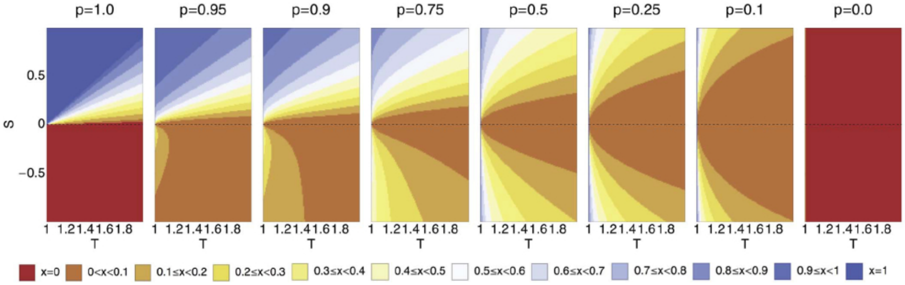
Isolated layers:



Random regular graph



“Exact” solution for well-mixed populations:



This stuff is interesting

P. S. For books [3] and [4], see http://sweet.ua.pt/sdorogov/ and https://sites.google.com/site/sergeydorogovtsev/lectures_on_complex_networks universidade de aveiro

universidade de aveiro

United Nations Educational, Scientific and Cultural Organization

International Year of Light 2015

i3N INSTITUTO DE NANOESTRUTURAS, NANOMODELAÇÃO E NANOFABRICAÇÃO

i3N INSTITUTE FOR NANOSTRUCTURES, NANOMODELLING AND NANOFABRICATION

universidade de aveiro theoria poiesis praxis

POCH

COMPETE 2020

PORTUGAL 2020

UNIÃO EUROPEIA Fundos Europeus Estruturais e de Investimento

POPH

QREN QUADRO DE REFERÊNCIA ESTRATÉGICO NACIONAL

GOVERNO DE PORTUGAL

UNIÃO EUROPEIA Fundo Social Europeu

COMPETE

FCT Fundação para a Ciência e a Tecnologia

FCT Fundação para a Ciência e a Tecnologia
MINISTÉRIO DA EDUCAÇÃO E CIÊNCIA