istituto
dei sistemi
complessi

Consiglio Nazionale delle Ricerche

Statistical Physics for Heterogeneous Random Networks

Andrea Gabrielli

Dip. di Ingegneria, Università Roma Tre, Italy

email: andrea.gabrielli@uniroma3.it

Central questions

- How to build maximally random graph ensembles satisfying some global or local topological/metric constraints (e.g. degree distribution)?
- Can statistical physics tools useful for this task?

Two fundamental applications

Reconstruction or modelling of a real network from partial topological information (e.g. degree sequence)

Examples: reconstruction of financial or ecological networks from incomplete information (e.g. whole topology by knowing only node degrees)

Null models for the statistical validation of the high order properties of a real network

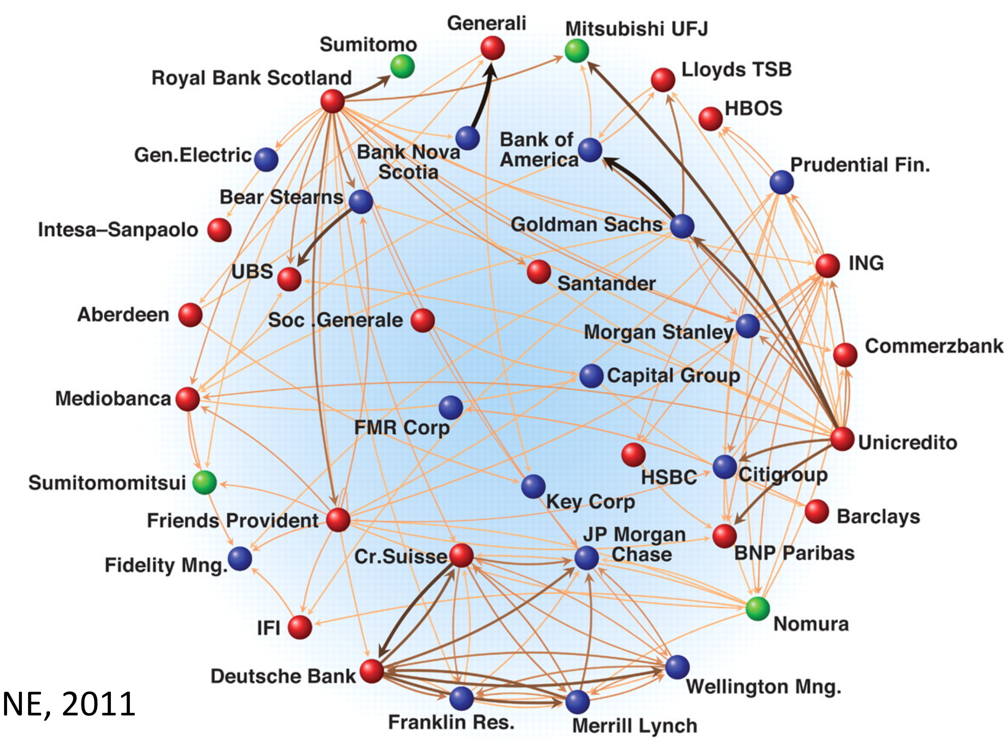
Examples: statistical validation of high order motifs of a real network vs a randomized model sharing only low order properties (e.g. validation of clustering properties by imposing only degrees)

Many social and biological networks, characterized by epidemic diffusion, have interactions/contacts that are only partially known

Reconstructing statistical properties of the network from partial information is one of the outstanding problems in the statistical physics of networks

Example: nodes = financial institutions
edges = various types of financial ties, e.g. loans or derivative contracts

These ties, whose information is usually limited, result in dependencies among institutions and constitute the ground for the propagation of financial distress across the network

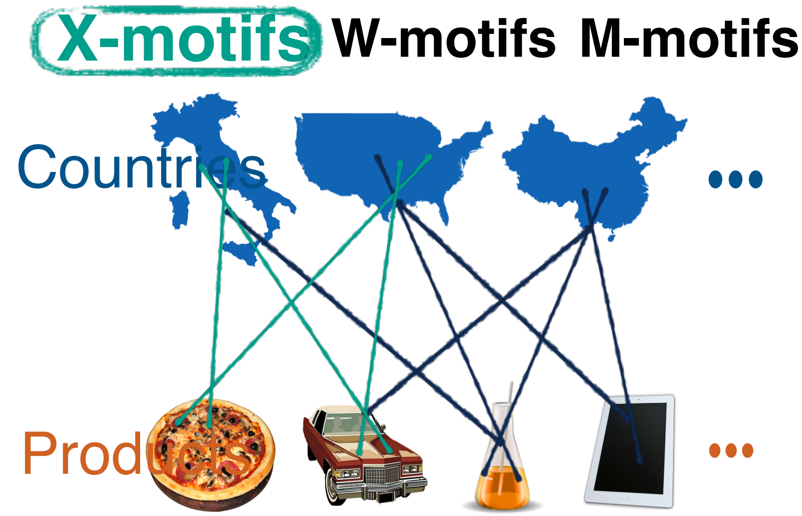


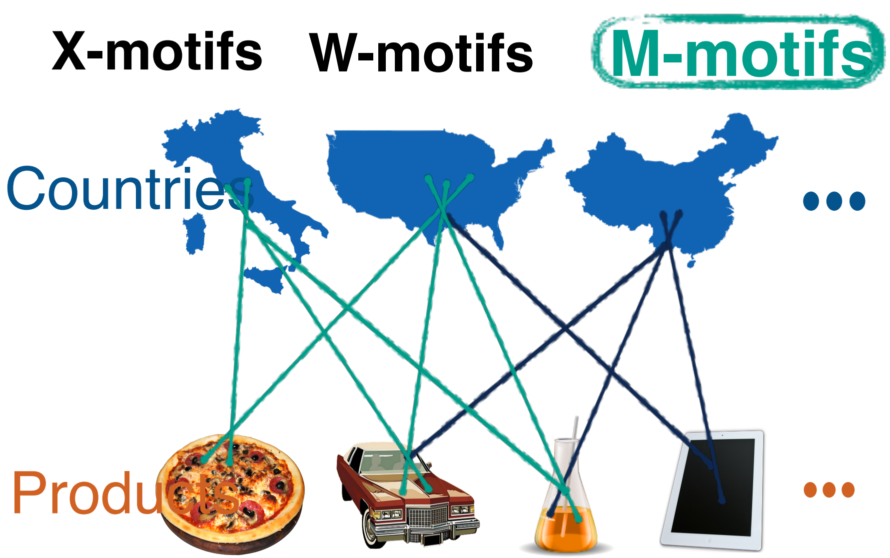
S. Vitali, J. B. Glattfelder, S. Battiston, PLOS ONE, 2011

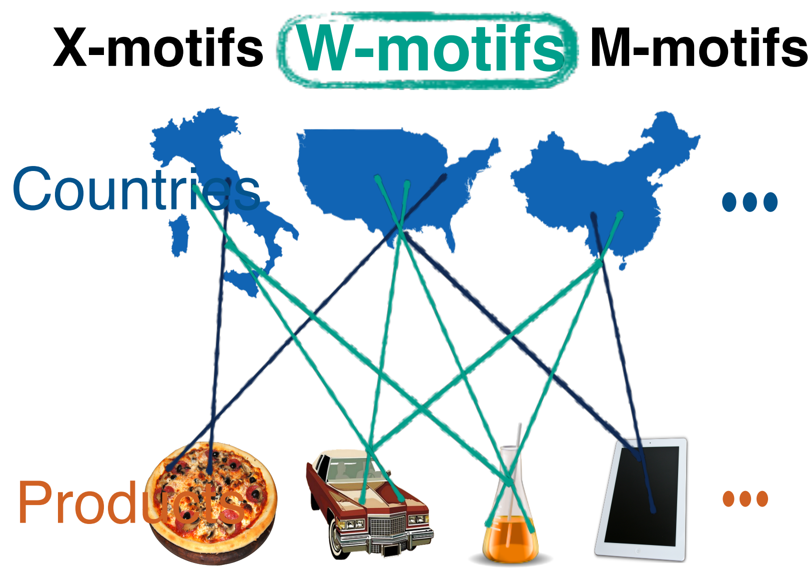
Example: validation of motifs in the bipartite country-product network in different years with respect degree conserving randomized bipartite network.

Detecting early signs of 2007-8 financial crisis in real economy (fast increase of randomization in immediately previous years)

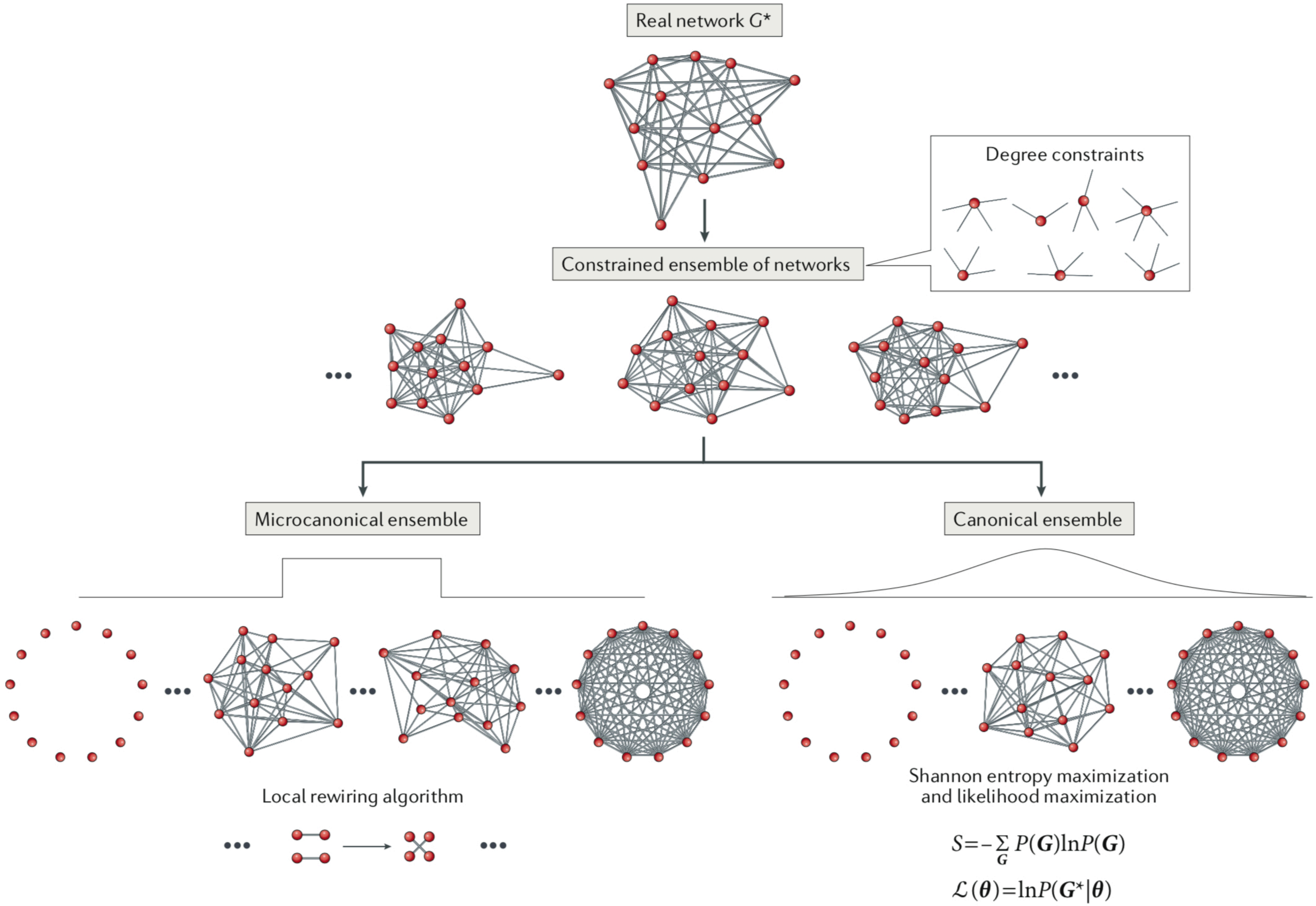
F. Saracco, R. Di Clemente, A. Gabrielli, T. Squartini, Scientific Reports **6**, 30286 (2016)







Reconstruction or constrained randomization methods



Exponential Random Graphs (ERG)

Information theory: Maximum Entropy approach

$\{C_a\}$ = set of graph properties that we want to constrain (e.g. we know their values for a real network) = *sufficient statistics*

ERG defines the maximally random ensemble Ω of graphs with N nodes compatible with the constraints (in average): Canonical Ensemble in Stat. Phys.

$$\langle C_a \rangle \equiv \sum_G C_a(G) P(G) = C_a^* \quad \forall a \quad (1)$$

G = generic graph of the ensemble with N nodes

$P(G)$ = measure on the ensemble $\Rightarrow$ it is found by maximizing the entropy

$$S(G) = -\sum_G P(G) \log P(G) \text{ with the constraints } (1) \Rightarrow$$

$$P(G) = \frac{1}{Z} \exp[-H(G)] \text{ where } H = \sum_a \theta_a C_a(G) \text{ with } \theta_a = \text{ Lagrange multipliers}$$

J. Park and M.E.J. Newman, Phys. Rev. E, **70**, 066117 (2004)

Example: binary directed network

If $\{C_a\} = \{k_i^{in}, k_i^{out}\}$ i=1,…,N ➔ $H(G)=\Sigma_i (\theta_i^{in} k_i^{in} + \theta_i^{out} k_i^{out})$

$\{\theta_i^{in}, \theta_i^{out}\}$ are fixed by the knowledge of $\langle k_i^{(in,out)} \rangle$

$\langle k_i^{(in,out)} \rangle = k_i^{*(in,out)}$ (Maximum likelihood if single configuration values)

$$p_{i\to j} = \frac{x_i^{out} x_j^{in}}{1 + x_i^{out} x_j^{in}}$$

Where $x_i^{(in,out)} = \exp[-\theta_i^{(in,out)}]$ = in - out fitness of node i

- N. Musmeci, S. Battiston, G. Caldarelli, M. Puliga, A. Gabrielli, J. of Stat. Phys., **151**, 220 (2013)
- G. Caldarelli, A. Chessa, A. Gabrielli, F. Pammolli, M. Puliga, Nature Phys., **9**, 125 (2013)
- G. Cimini, T. Squartini, D. Garlaschelli, A. Gabrielli, Scientific Reports **5**, 15758 (2015)
- G. Cimini, T. Squartini, A. Gabrielli, D. Garlaschelli, Phys. Rev. E **92**, 040802(R) (2015)
- F. Saracco, R. Di Clemente, A. Gabrielli, T. Squartini, Scientific Reports **5**,10595 (2015)

+

Reviews (see next slide)

ELSEVIER

Physics Reports
Volume 757, 8 October 2018, Pages 1-47

Reconstruction methods for networks: The case of economic and financial systems

Tiziano Squartini [a], Guido Caldarelli [a, b, c], Giulio Cimini [a, b], Andrea Gabrielli [b, a], Diego Garlaschelli [a, d]

⊞ Show more

https://doi.org/10.1016/j.physrep.2018.06.008

Technical Review | Published: 08 January 2019

The statistical physics of real-world networks

Giulio Cimini, Tiziano Squartini, Fabio Saracco, Diego Garlaschelli, Andrea Gabrielli & Guido Caldarelli

Nature Reviews Physics **1**, 58–71 (2019) | Download Citation

ELSEVIER

Journal of Financial Stability

Available online 27 May 2017

In Press, Corrected Proof

The missing links: A global study on uncovering financial network structures from partial data ☆

Kartik Anand [a], Iman van Lelyveld [b], Ádám Banai [c], Soeren Friedrich [a], Rodney Garratt [d], Grzegorz Hałaj [e], Jose Fique [f], Ib Hansen [g], Serafín Martínez Jaramillo [h], Hwayun Lee [i], José Luis Molina-Borboa [h], Stefano Nobili [j], Sriram Rajan [k], Dilyara Salakhova [l], Thiago Christiano Silva [m], Laura Silvestri [n], Sergio Rubens Stancato de Souza [m]

[a] Deutsche Bundesbank, Germany
[b] De Nederlandsche Bank, The Netherlands
[c] Magyar Nemzeti Bank, Hungary
[d] Federal Reserve Bank of New York, United States
[e] European Central Bank, Germany
[f] Bank of Canada, Canada
[g] Danmarks Nationalbank, Denmark
[h] Banco de Mexico, Mexico
[i] Bank of Korea, South Korea
[j] Banca d'Italia, Italy
[k] Office of Financial Research, United States
[l] Banque de France, France
[m] Banco Central do Brasil, Brazil
[n] Bank of England, United Kingdom

"In presenting our results we face the challenge that some algorithms produce an ensemble of networks while others produce a single matrix. This makes a straightforward comparison difficult. Fortunately the *Cimi* method is the clear winner between the ensemble methods"

May we do the analogy with statistical physics stricter?

Let us go deeper in the Jayne's formulation of statistical mechanics: heterogeneous weighted undirected networks

$A =$ symmetric adjacency matrix

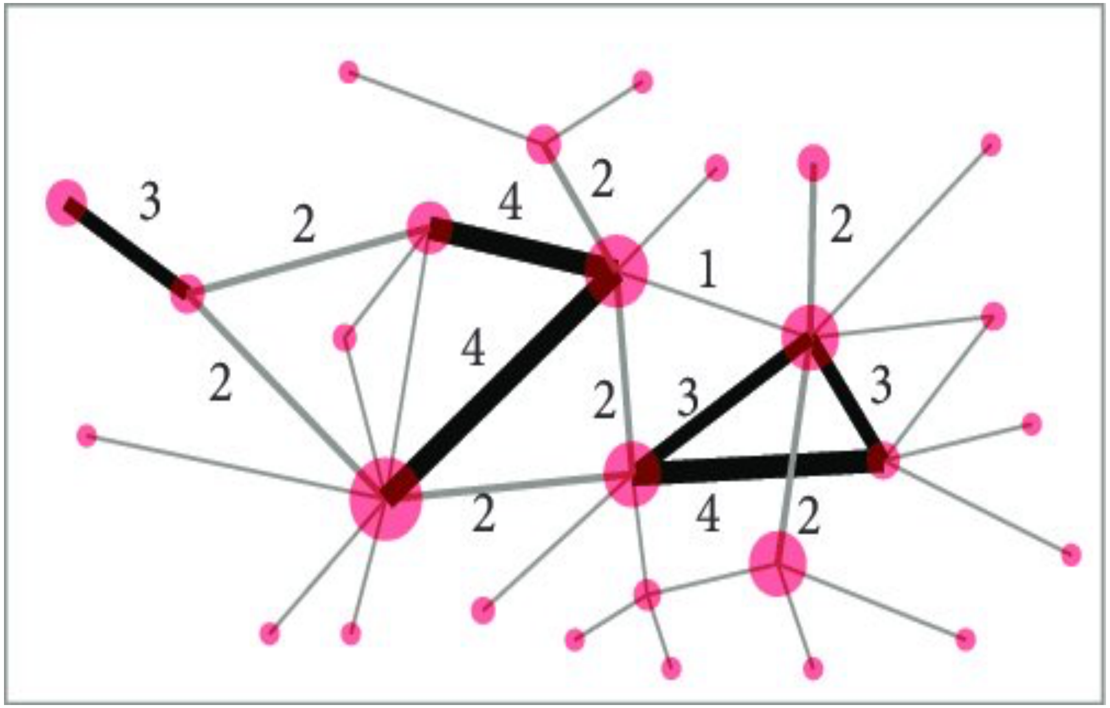
$a_{ij} = a_{ji} = 0,1$ with $i, j = 1,\dots,N$

$k_i = \sum_{j(\neq i)} a_{ij} =$ degree

$W =$ symmetric weight matrix

$w_{ij} = w_{ji} =$ continuous or discrete weights

$s_i = \sum_{j(\neq i)} w_{ij} =$ strength



We want to build the maximally random ensemble of weighted graphs with $\langle k_i \rangle = k_i^*$ and $\langle s_i \rangle = s_i^*$ for all $i=1,\dots,N$

Standard approach

The knowledge of the matrix W implies the knowledge of A if we interpret $w_{ij} = 0$ as a non existing link, i.e. $a_{ij} = 0$

1. One can think to define a probability measure $P(W)$ over all possible matrices W

2. Maximise $S[P(W)] = -\sum_W P(W) \log P(W)$ with the constraints $\langle k_i \rangle = k_i^*$ and $\langle s_i \rangle = s_i^*$ for all $i = 1, \dots, N$ to find $P(W)$

However, since $\langle k_i \rangle = k_i^*$ is a singular constraint in $S[P(W)]$ when w_{ij} are continuous variables, it is possible only in the discrete case $w_{ij} = n_{ij} w_0$ with $n_{ij} = 0,1,2, \dots$

Enhanced Configuration Model

D. Garlaschelli and M. I. Loffredo, Phys. Rev. Lett. **102**, 038701 (2009).
R. Mastrandrea, T. Squartini, G. Fagiolo, and D. Garlaschelli, New J. Phys. **16**, 043022 (2014).

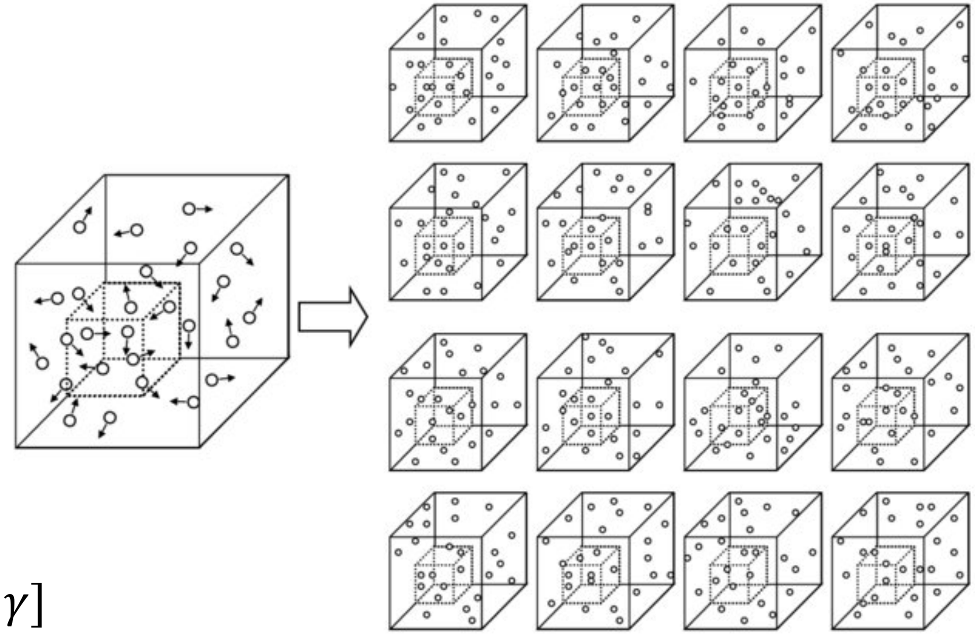
How to treat the case of continuous and defined in an arbitrary (left bounded) interval?

Jayne's formulation of the Grand Canonical Ensemble for Hamiltonian systems

$C = (N, \boldsymbol{x})$ with $x \in \mathbb{R}^{6N}$

$H_N(\boldsymbol{x})$=Hamiltonian of N particles

$$\sum_C = \sum_N \int d^{6N}x$$



$P(C) = P(N, \boldsymbol{x})$ maximises

$$S_C = S - \sum_C P(N,\boldsymbol{x})[\alpha N + \beta H_N(\boldsymbol{x}) + \gamma]$$

where

$$S[P] = - \sum_C P(N,\boldsymbol{x}) \log P(N,\boldsymbol{x})$$

with constraints $\langle N \rangle = N^*$ and $\langle H_N \rangle = E$

$$P(N,\boldsymbol{x}) = Z_G^{-1}(\alpha,\beta)\exp[-\alpha N - \beta H_N(\boldsymbol{x})]$$

$$Z_G(\alpha,\beta) = \sum_{N=1}^{\infty} e^{-\alpha N} Z_C(\beta, N)$$

$$Z_C(\beta, N) = \int dx^{6N} e^{-\beta H_N(\boldsymbol{x})}$$

$$\langle N \rangle \equiv \sum_C P(N,\mathbf{x})N \equiv -\frac{\partial \log Z_G(\alpha,\beta)}{\partial \alpha} = N^*$$

$$\langle H_N \rangle \equiv \sum_C P(N,\mathbf{x})H_N(\mathbf{x}) \equiv -\frac{\partial \log Z_G(\alpha,\beta)}{\partial \beta} = E$$

$$\Phi_G = -kT \log Z_G(\alpha,\beta)$$

$$\alpha = -\frac{\mu}{kT} = \frac{\partial \Phi_G}{\partial N^*}$$

$$\beta = \frac{1}{kT} = \frac{\partial \Phi_G}{\partial E}.$$

The same approach can be adopted for weighted random graphs *G* with *N* nodes

We can build a similar *Fock space* of configurations with variable number of links

Roughly speaking A plays the role of the number of particles *N* and the weights W define the energy function H_N : existing links

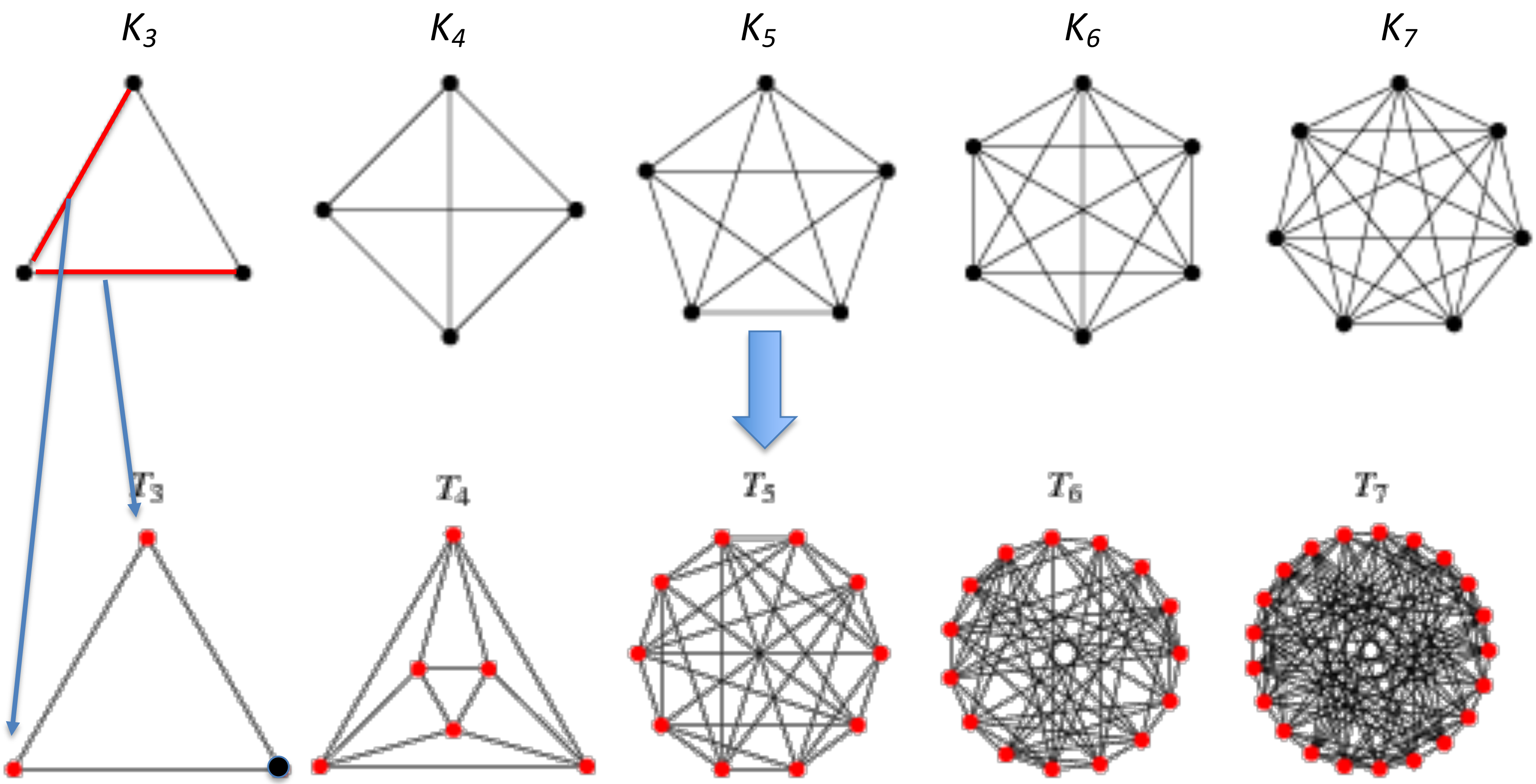
More precisely,

The graph *G* is mapped into a configuration of a lattice gas embedded in an *appropriate space*:

- Each link of *G* corresponds to an *occupied* edge of K_N , the complete graph of *N* nodes.
- We can define a lattice using the *line graph* of K_N, also called the *triangular graph* T_N of order *N*: each edge of K_N corresponds to a vertex (site) on the *lattice* T_N
- Therefore each link of *G* is mapped into a particle of a gas on the lattice T_N

A. Gabrielli, R. Mastrandrea, G. Caldarelli, G. Cimini, *Grand canonical ensemble of weighted networks*, Phys. Rev. E **99**, 030301(R) (2019)

Mapping between complete graphs K_n and triangular graphs T_n

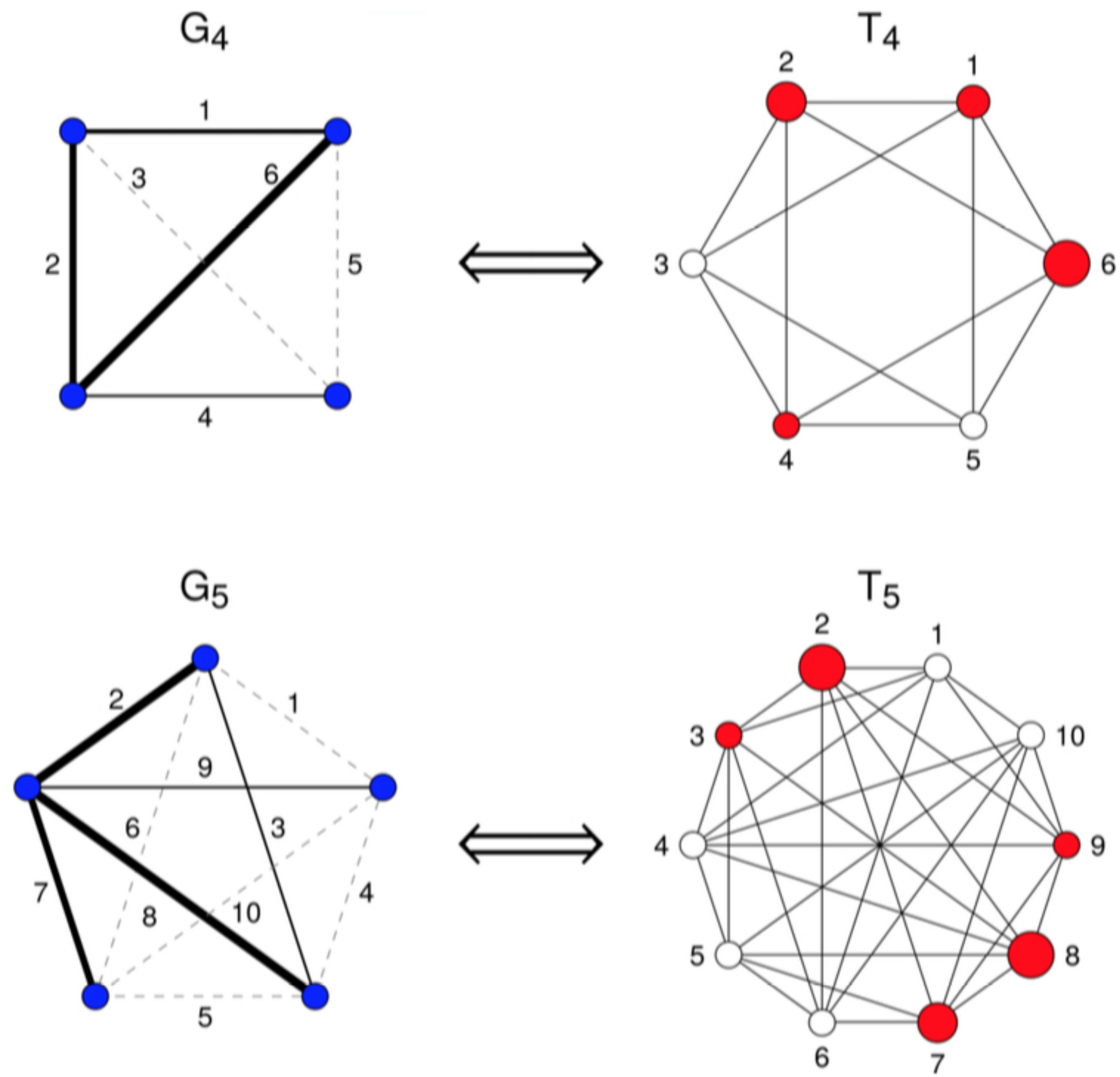


In T_N number of nodes $M=N(N-1)/2$ and $2(N-2)$ first n.n. for each node.
Only first and second neighbours: diameter = 2

We map each weighted graph configuration into a configuration of a lattice gas with an internal continuous variable (energy/magnetization)

N=4

N=5



- The binary adjacency matrix A of the network fixes the number L and the positions of the gas particles $[1 \leq L \leq \frac{N(N-1)}{2} = |V|]$ on $T_{N,}$
- The weights matrix W defines the internal coordinates of the existing particles.

Therefore we can exactly map the Maximum Constrained Entropy method for the definition of the weighted graphs ensemble into the definition of the Grand Canonical Ensemble for a lattice gas on T_N

$$P(\mathcal{C}) = P(\mathsf{A}, \mathsf{W}) \equiv P(\mathcal{L}_{\mathsf{A}}, \{w_{ij}\}_{(i,j)\in\mathcal{L}_{\mathsf{A}}})$$

$\mathcal{L}_{\mathsf{A}} \subseteq \mathcal{V}$ with $|\mathcal{L}_{\mathsf{A}}| = L$ (i.e., the set of node pairs with $a_{ij} = 1$)

$$\sum_{\mathcal{C}} P(\mathcal{C}) \cdot \Longleftrightarrow \sum_{\mathsf{A}} \prod_{i<j}^{\mathcal{L}_{\mathsf{A}}} \int_0^\infty dw_{ij} P(\mathcal{L}_{\mathsf{A}}, \{w_{ij}\}_{(i,j)\in\mathcal{L}_{\mathsf{A}}})$$

The GC ensemble is defined by the maximization of

$$S_c = -\sum_C P(C) \log P(C) - \sum_i \lambda_i \langle O_i(C) \rangle$$

Simplest case: weighted Erdos-Renyi graphs

Global (homogeneous) constrained averages

$$\langle L \rangle \equiv \langle \sum_{i<j}^{\mathcal{V}} a_{ij} \rangle = L^* \qquad \langle W \rangle \equiv \langle \sum_{i<j}^{\mathcal{L}_A} w_{ij} \rangle = W^*$$

$$P(\mathrm{A}, \mathrm{W}, \alpha, \beta) = Z_G^{-1}(\alpha, \beta) e^{-H(\mathrm{A},\mathrm{W},\alpha,\beta)}$$

$$H(\mathrm{A}, \mathrm{W}, \alpha, \beta) = \alpha \sum_{i<j}^{\mathcal{V}} a_{ij} + \beta \sum_{i<j}^{\mathcal{L}_A} w_{ij}$$

$$Z_G(\alpha, \beta) = \sum_{\mathcal{C}} e^{-H(\mathrm{A},\mathrm{W},\alpha,\beta)} = \sum_{\mathrm{A}} e^{-\alpha \sum_{i<j}^{\mathcal{V}} a_{ij}} Z_C(\beta) = \sum_{L=0}^{V} \binom{V}{L} \frac{e^{-\alpha L}}{\beta^L} = \left[1 + \frac{e^{-\alpha}}{\beta}\right]^V$$

$$Z_C(\beta) = \prod_{i<j}^{\mathcal{L}_A} \int_0^\infty dw_{ij} e^{-\beta w_{ij}} = \beta^{-\sum_{i<j}^{\mathcal{V}} a_{ij}} = \beta^{-L}$$

$$\langle L \rangle \equiv -\partial_\alpha \log Z_G(\alpha, \beta) \equiv \frac{V}{\beta e^\alpha + 1} = L^*,$$

$$\langle W \rangle \equiv -\partial_\beta \log Z_G(\alpha, \beta) \equiv \frac{V \beta^{-1}}{\beta e^\alpha + 1} = W^*$$

Homogeneity implies decoupling

By assuming $\alpha' = \alpha + \log \beta$

Conditional weights distr.

$$P(\mathrm{A}, \mathrm{W}) = \left[\prod_{i<j}^{\mathcal{V}} \frac{e^{-\alpha' a_{ij}}}{1 + e^{-\alpha'}}\right] \left[\prod_{i<j}^{\mathcal{L}_A} \beta e^{-\beta w_{ij}}\right]$$

Continuous Enhanced Configuration Model (CECM)

Local (heterogeneous) constrained averages (*i=1,...,N*)

$$\langle k_i \rangle \equiv \langle \sum_{j(\neq i)}^{\mathcal{V}} a_{ij} \rangle = k_i^* \qquad \langle s_i \rangle = \langle \sum_{j(\neq i)}^{\mathcal{L}_\mathsf{A}} w_{ij} \rangle = s_i^*$$

$$P(\mathsf{A}, \mathsf{W}, \{\alpha_i, \beta_i\}_{i=1}^N) = Z_G^{-1}(\{\alpha_i, \beta_i\}_{i=1}^N) e^{-H(\mathsf{A}, \mathsf{W}, \{\alpha_i, \beta_i\}_{i=1}^N)}$$

$$H(\mathsf{A}, \mathsf{W}, \{\alpha_i, \beta_i\}_{i=1}^N) = \sum_{i<j}^{\mathcal{V}} (\alpha_i + \alpha_j) a_{ij} + \sum_{i<j}^{\mathcal{L}_\mathsf{A}} (\beta_i + \beta_j) w_{ij}$$

$$Z_G(\{\alpha_i, \beta_i\}_{i=1}^N) = \prod_{i<j}^{\mathcal{V}} \left(1 + \frac{e^{-(\alpha_i + \alpha_j)}}{\beta_i + \beta_j}\right)$$

$$P(\mathsf{A}, \mathsf{W}) = \left[\prod_{i<j}^{\mathcal{V}} \frac{e^{-[\alpha_i + \alpha_j + \log(\beta_i + \beta_j)] a_{ij}}}{1 + e^{-[\alpha_i + \alpha_j + \log(\beta_i + \beta_j)]}}\right] \left[\prod_{i<j}^{\mathcal{L}_\mathsf{A}} (\beta_i + \beta_j) e^{-(\beta_i + \beta_j) w_{ij}}\right] = \pi(\mathsf{A}) q(W_{\mathcal{L}_\mathsf{A}})$$

$$p_{ij} = \frac{e^{-[\alpha_i + \alpha_j + \log(\beta_i + \beta_j)]}}{1 + e^{-[\alpha_i + \alpha_j + \log(\beta_i + \beta_j)]}}$$

Entanglement of topology and internal variables

Constraints and the effect of heterogeneity

$$\langle k_i \rangle \equiv -\partial_{\alpha_i} \log Z_G\left(\{\alpha_l, \beta_l\}_{l=1}^{N}\right) \equiv \sum_{j(\neq i)}^{\mathcal{V}} \frac{1}{1 + (\beta_i + \beta_j) e^{\alpha_i + \alpha_j}} = k_i^*$$

$$\langle s_i \rangle \equiv -\partial_{\beta_i} \log Z_G\left(\{\alpha_l, \beta_l\}_{l=1}^{N}\right) \equiv \sum_{j(\neq i)}^{\mathcal{V}} \frac{(\beta_i + \beta_j)^{-1}}{1 + (\beta_i + \beta_j) e^{\alpha_i + \alpha_j}} = s_i^*$$

The existence of different links (particles) are independent events, and the weights of existing ones are also mutually independent,

BUT

heterogeneity (disorder) makes in principle impossible to set topological features of the network independently of the weights distribution.

E.g. a large constrained strength of a node forces the existence of links pointing to it

What happens if we impose independently topological and weight constraints?

Separated Enhanced Configuration Model (SECM)

Two steps entropy maximization

1) First we impose topological constraints on a topological entropy $S[\pi(A)]$

$$\langle k_i \rangle \equiv \langle \sum_{j(\neq i)}^{\mathcal{V}} a_{ij} \rangle = k_i^* \quad \Rightarrow \quad \pi(\mathrm{A}) = \prod_{i<j}^{\mathcal{V}} \frac{e^{-(\alpha_i' + \alpha_j') a_{ij}}}{1 + e^{-(\alpha_i' + \alpha_j')}}$$

2) For each adjacency matrix we maximise the conditional entropy

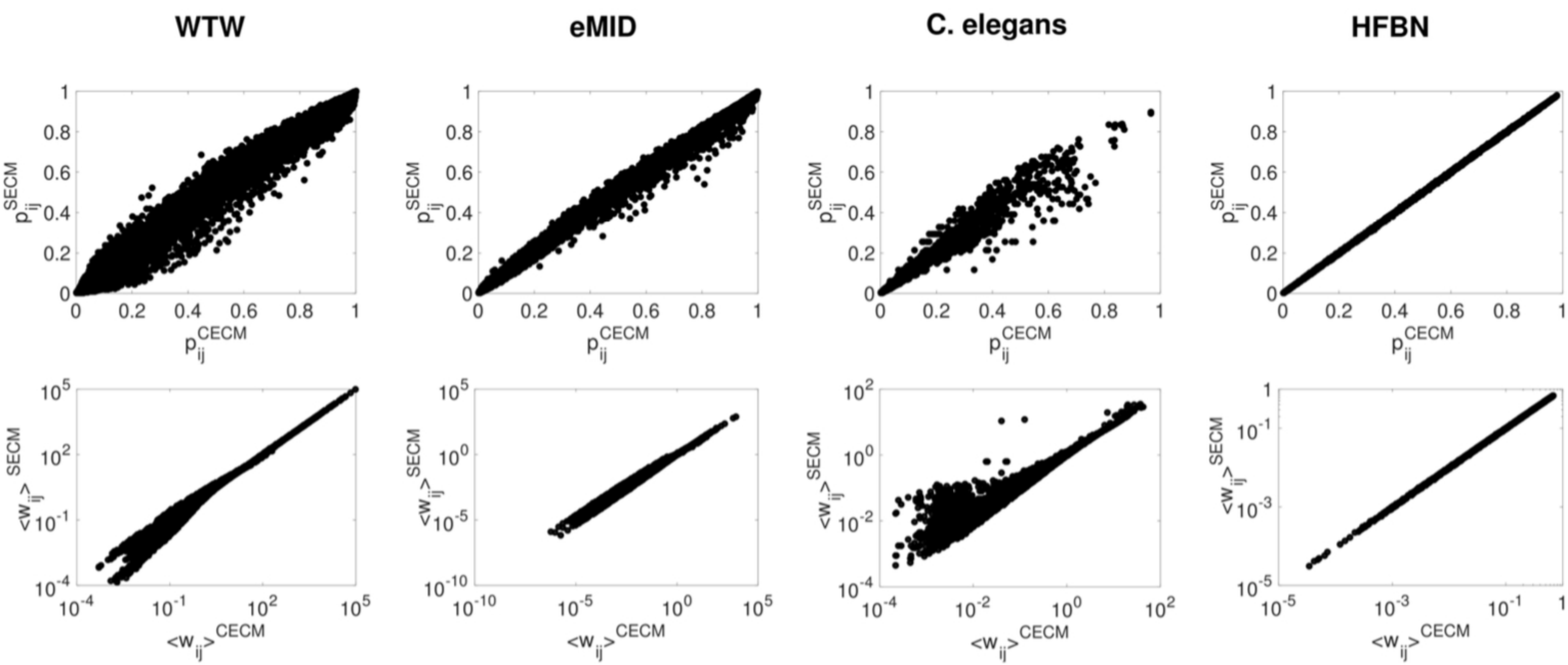
$$\langle s_i \rangle = \langle \sum_{j(\neq i)}^{\mathcal{L}_\mathrm{A}} w_{ij} \rangle = s_i^* \quad \Rightarrow \quad q(W_{\mathcal{L}_\mathrm{A}}) = \prod_{i<j}^{\mathcal{L}_\mathrm{A}} (\beta_i + \beta_j) e^{-(\beta_i + \beta_j) w_{ij}}$$

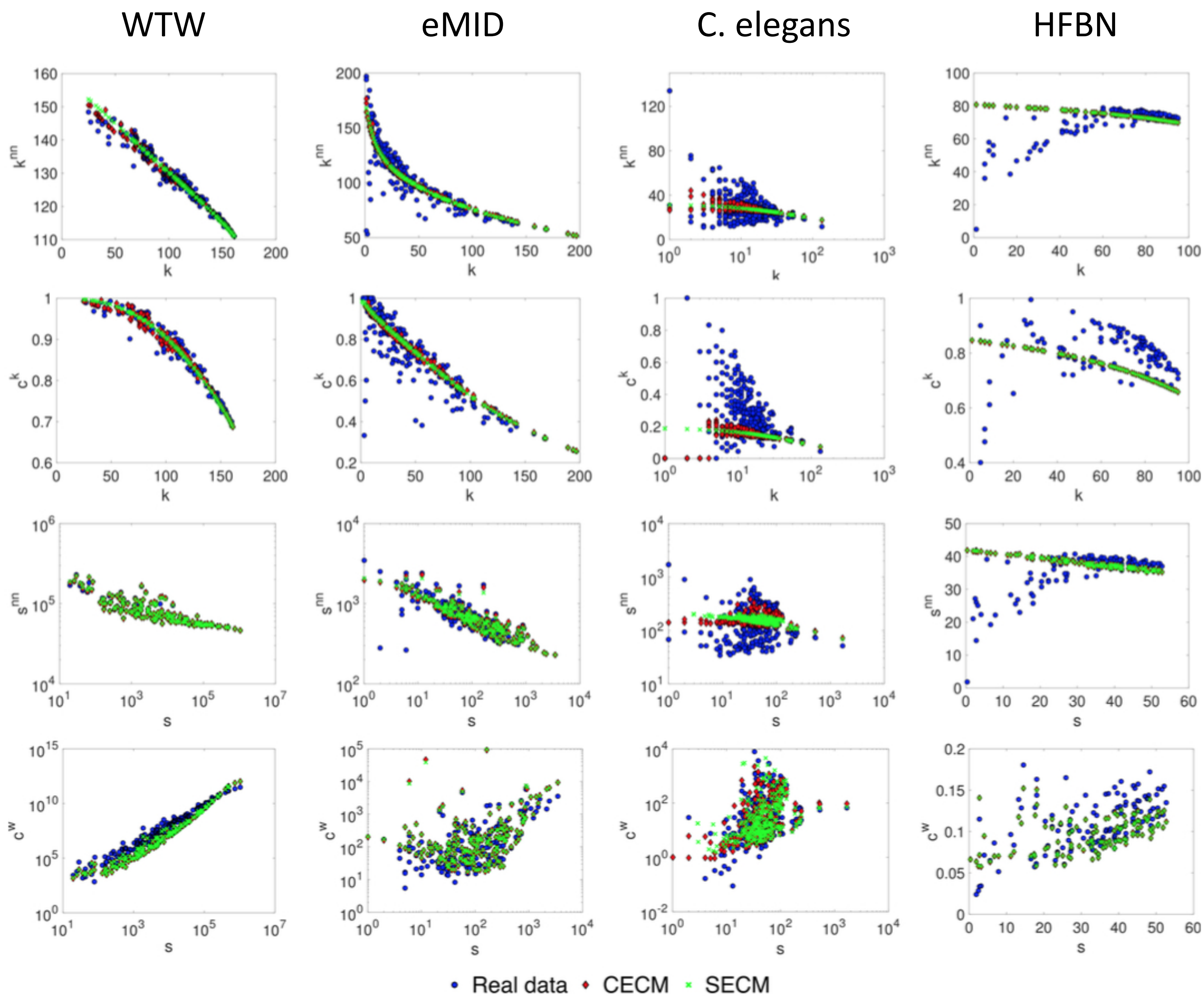
Constraints equations

$$\langle k_i \rangle \equiv \sum_{j(\neq i)}^{\mathcal{V}} \frac{1}{1 + e^{\alpha_i' + \alpha_j'}} = k_i^*,$$

$$\langle s_i \rangle \equiv \sum_{j(\neq i)}^{\mathcal{V}} \frac{(\beta_i + \beta_j)^{-1}}{1 + e^{\alpha_i' + \alpha_j'}} = s_i^*$$

Comparison between exact CECM and approximated SECM




WTW
eMID
C. elegans
HFBN
k^nn
k
c^k
s^nn
s
c^w
Real data
CECM
SECM

Conclusions

- Constrained random models of graphs are fundamental tools for reconstruction
- of networks with partial information and statistical validation of properties of real networks

- They can be formulated in strict analogy with statistical physics of particle systems in particular spaces/lattices (triangular graphs), building typical Fock spaces for the configurations of a lattice gas

- Heterogeneity of topological and weight constraints implies entanglement between these properties

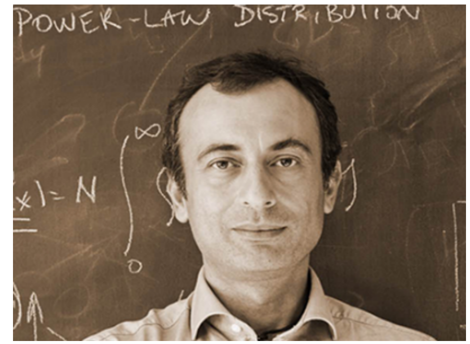
- Disentangled approximation works is much simpler to solve and very often works well

- Finally, this approach open a new perspective on the mathematical approach to this kind of networks and graphs: we can map constrained random network ensembles into statistical physical particle systems

Main Collaborators

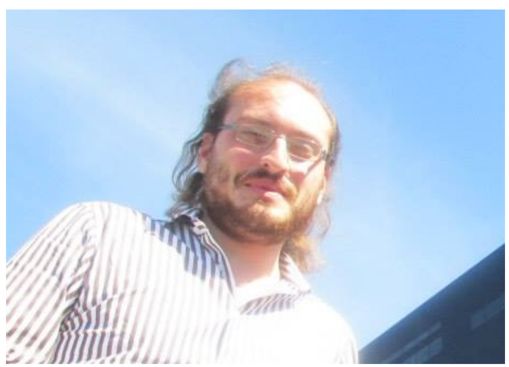
Giulio Cimini
(Univ. “Tor Vergata”, Rome)

Rossana Mastrandrea
(IMT – Lucca)



Guido Caldarelli
(IMT – Lucca)

Diego Garlaschelli
(IMT - Lucca)



Tiziano Squartini
(IMT – Lucca)

Fabio Saracco
(IMT – Lucca)

Riccardo Di Clemente
(UCL, London)