Maximum Entropy Network Ensembles

LTCC Course
Lesson 4

Ginestra Bianconi

School of Mathematical Sciences Queen Mary University of London

Queen Mary
University of London

Forth lesson

- Correlated and uncorrelated networks
- Exponential random graphs in the uncorrelated limit
- Microcanonical ensemble (Configuration model)
- 2-star model and Strauss model

Correlated and Uncorrelated networks

References

Books

- Mark Newman ***Networks: An introduction*** (Oxford University Press,2010)
- Ginestra Bianconi ***Multilayer networks:Structure and Function*** (Oxford University Press,2018)

Articles

- Park, J. and Newman, M.E., 2004. Statistical mechanics of networks. *Physical Review E*, *70*(6), p.066117.
- Bianconi, G., 2007. The entropy of randomized network ensembles. *EPL (Europhysics Letters)*, *81*(2), p.28005.
- Bianconi, G., 2009. Entropy of network ensembles. Physical Review E, 79(3), p.036114.
- Anand, K. and Bianconi, G., 2009. Entropy measures for networks: Toward an information theory of complex topologies. Physical Review E, 80(4), p.045102.
- Anand, K. and Bianconi, G., 2010. Gibbs entropy of network ensembles by cavity methods. Physical Review E, 82(1), p.011116.

Description of

correlated and uncorrelated networks

in terms of degree classes

A network has

degree correlations

if the probability that a random link is connected to a node of degree k $\pi_{k|k'}$

depends on the degree k'

of the node at

the other end of the link

Assortative and disassortative networks

In assortative networks

"hubs connect preferentially to hubs"

In disassortative networks

"hubs connect preferentially to low degree nodes"

Assortative and disassortative networks

Social networks

are generally **assortative**

Protein-interaction networks

are **disassortative.**

Technological networks

are generally **disassortative**

(ex. Internet).

Measure of degree correlations

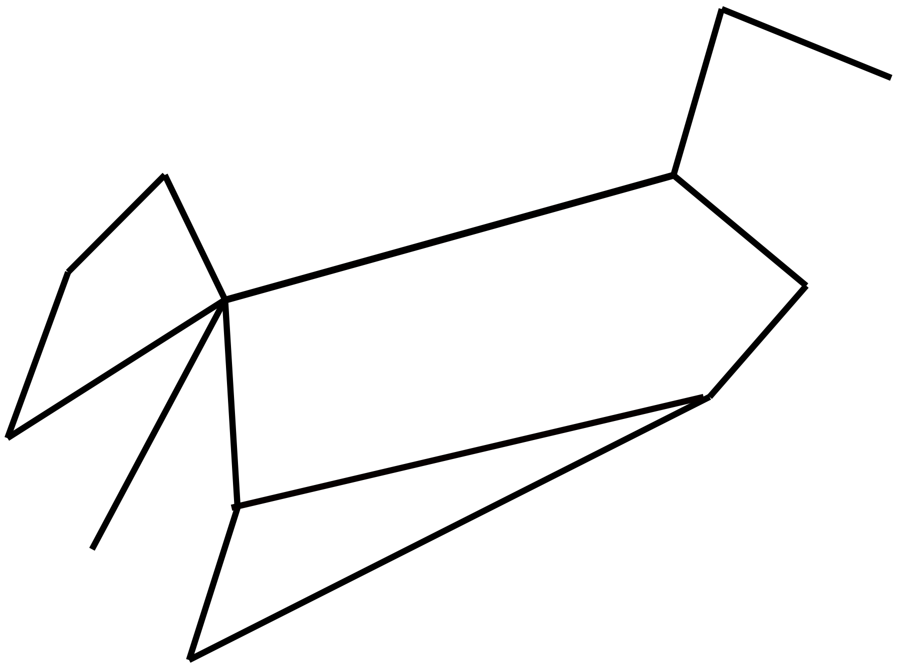
The most direct measure of the matrix $\pi_{k,k'}$ is the direct measure of the probability

This method has some limitations

A. The network might be too sparse to have enough statistics to reconstruct the full matrix

B. In presence of large degree the model cannot be compared directly with the uncorrelated network limit. In order to have a null model usually the random swapping of connection is considered.

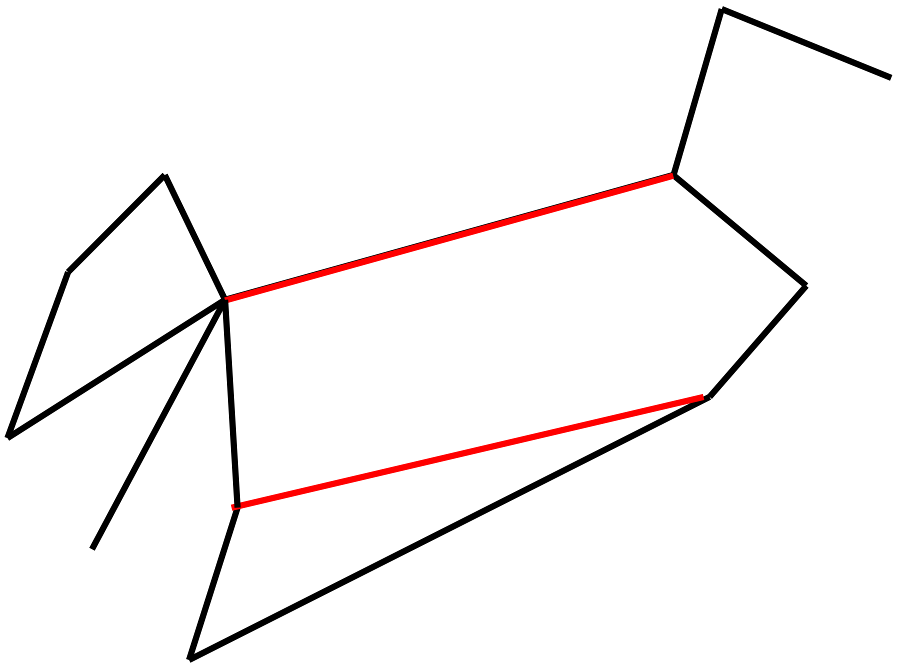
Randomization of a network swap of connections



- *Choose two random links linking four distinct nodes*

Maslov & Sneppen 2002

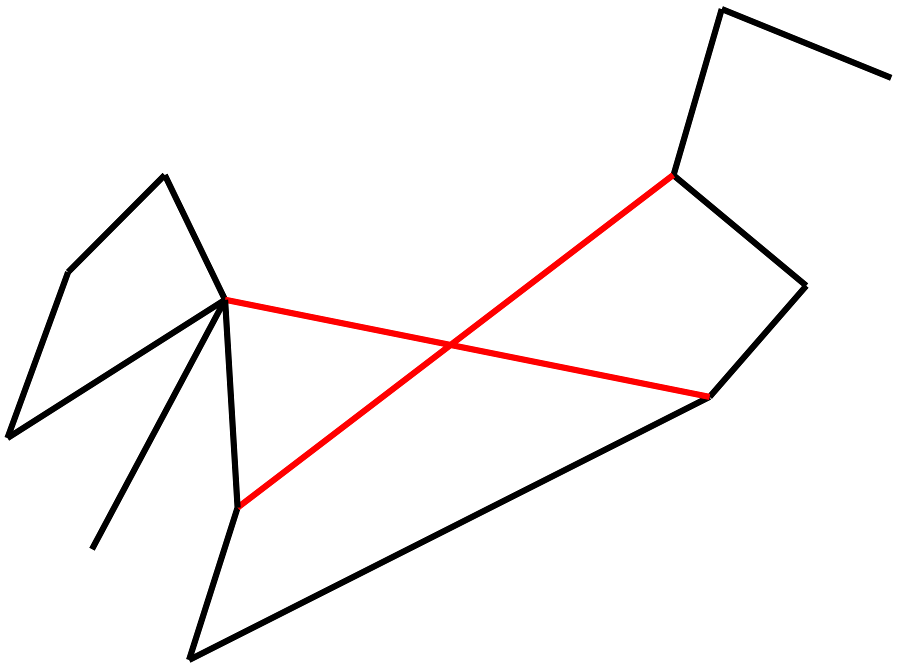
Randomization of a network swap of connections



- *Choose two random links linking four distinct nodes*
- *If possible (not already existing links) swap the ends of the links*

Maslov & Sneppen 2002

Randomization of a network swap of connections



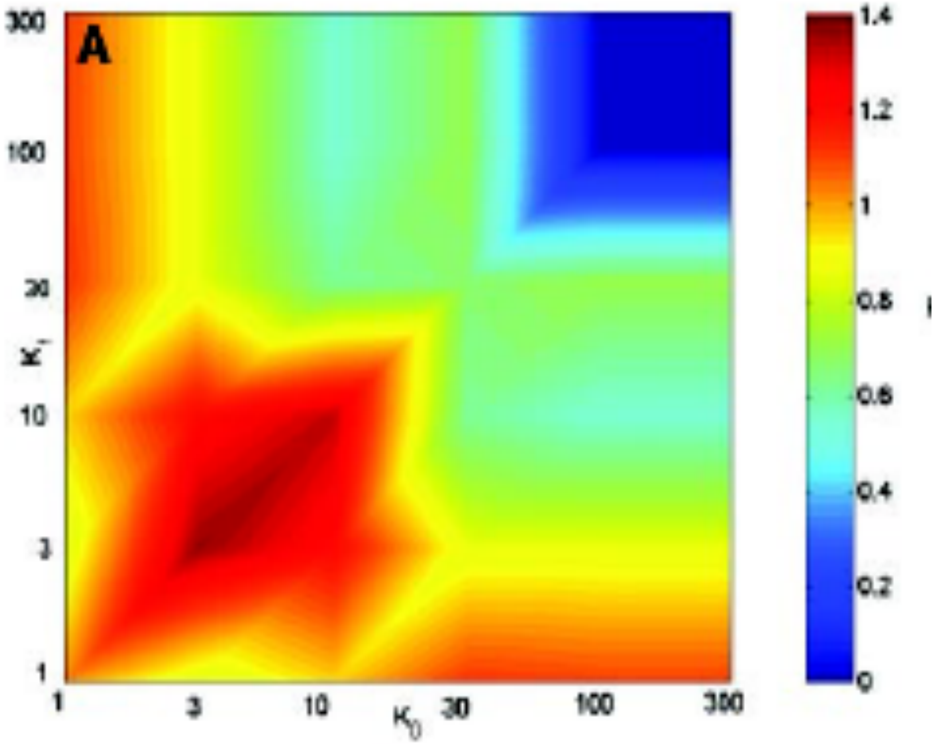
- *Choose two random links linking four distinct nodes*
- *If possible (not already existing links) swap the ends of the links*

Maslov & Sneppen 2002

Direct measurement of degree correlations

$\pi_{k,k'}$ Probability that nodes of degree k and k' are connected by a link

$\tilde{\pi}_{k,k'}$ Same probability in randomised networks



S. Maslov and K. Sneppen Science 2002

The map of

$$\frac{\pi_{k,k'}}{\tilde{\pi}_{k,k'}}$$

reveals the correlations in the protein interaction map

The average degree of neighbour nodes

The average degree of the neighbours of a node is given by

$$k_{nn}(i) = \frac{1}{k_i} \sum_{j=1}^{N} a_{ij} k_j$$

The average degree of the neighbours of nodes of degree k is given by

$$k_{nn}(k) = \left\langle \frac{1}{k_i} \sum_{j=1}^{N} a_{ij} k_j \right\rangle_{k_i = k} = \frac{1}{N(k)} \sum_{i | k_i = k} k_{nn}(i)$$

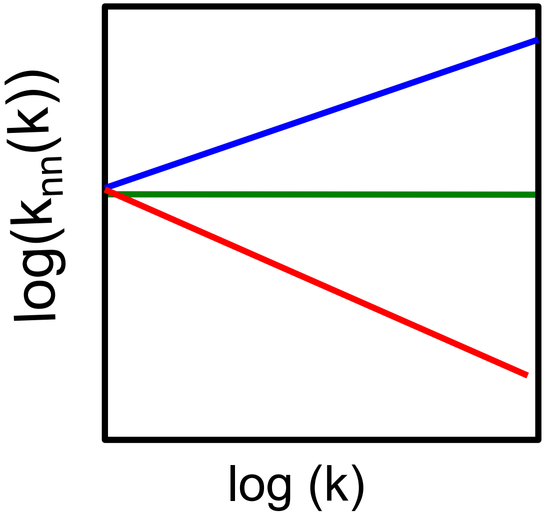
The average degree of neighbour nodes

The average degree of the neighbours of nodes of degree k

Comments

- This is a more coarse grained measure for which there is better statistics
- A monotonically increasing indicates assortative correlations
- A monotonically decreasing indicates disassortative correlations
- A drawback is that in the case in which is not monotonic we cannot classify the correlations.

Average degree of the neighbour of a node of degree k



$$k_{nn}(k) \sim k^{\alpha}$$

Assortative networks $\alpha>0$

Uncorrelated networks $\alpha=0$

Disassortative networks $\alpha<0$

Average degree of a neighbour of a node of degree k

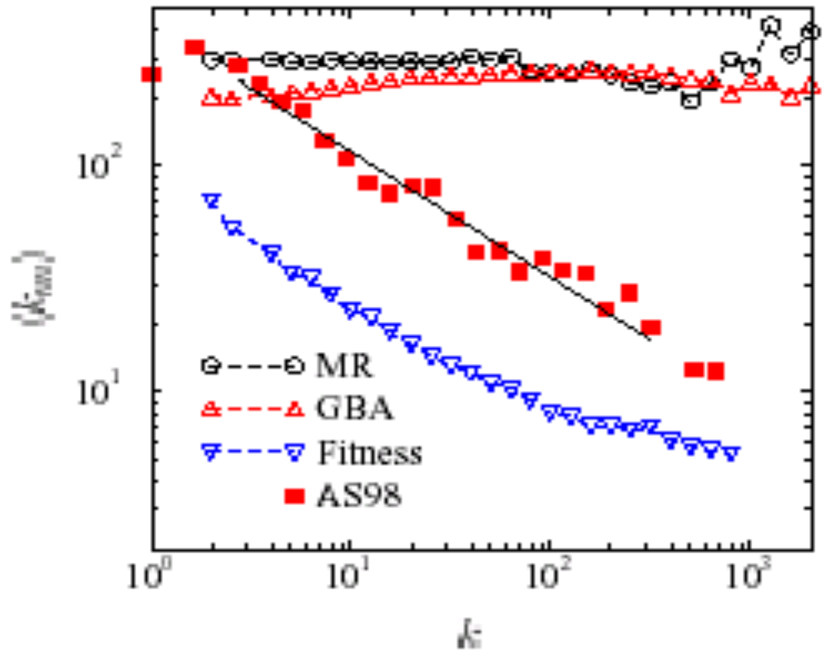
$$k_{nn}(k) = \left\langle \frac{1}{k_i} \sum_{j=1}^{N} a_{ij} k_j \right\rangle_{k_i=k}$$

Disassortative correlations in the Internet at the AS level

The average degree of the neighbours of nodes of degree k

$$k_{nn}(k) = \left\langle \frac{1}{k_i} \sum_{j=1}^{N} a_{ij} k_j \right\rangle_{k_i=k}$$

reveals that the
the Internet at the AS level is
disassortative



Vazquez et al. PRL (2001)

Newman correlation coefficient

The Newman correlation coefficient is a global parameter that provides a unique number $r \in [-1,1]$

given by

$$r = \frac{\sum_{k,k'} kk'(\pi_{k,k'} - q_k q_{k'})}{\sum_k k^2 q_k - \left(\sum_k k q_k\right)^2}$$

We have a classification of the networks depending on the sign of r

$r > 0$ **assortative network**
$r < 0$ **disassortative network**

Description of

correlated and uncorrelated networks

in terms of node labels

Uncorrelated networks

Definition

In ***uncorrelated networks***

in which each node *i* has expected degree $\bar{k}_i$

the probability that a random link

connects a node *i* at one end to a node *j* at the other end

is given by

$$\pi_{ij} = \frac{\bar{k}_i \bar{k}_j}{(\langle k \rangle N)^2}$$

Uncorrelated networks

Proposition

In an uncorrelated network in which each node i has expected degree $\bar{k}_i$ the probability that a random link is connected to node i given that is connected to node j a the other end is given by

$$q_i = \pi_{i|j} = \frac{\bar{k}_i}{\langle \bar{k} \rangle N}$$

Comments

- The probability q_i only depends on the degree of node i and is independent of node j

- The probability q_i can be interpreted as the probability that in an uncorrelated network we reach node i by following the link of any random node

Proof

Given the the expression $\pi_{ij} = \frac{\bar{k}_i \bar{k}_j}{(\langle k \rangle N)^2}$

we want to show that in uncorrelated networks we have

$$\pi_{i|j} = \frac{\bar{k}_i}{\langle \bar{k} \rangle N} = q_i$$

According to the Bayes rule we have $\pi_{i|j} = \frac{\pi_{ij}}{\sum_{j'=1}^{N} \pi_{jj'}}$

The denominator reads $\sum_{j'=1}^{N} \pi_{jj'} = \sum_{j'=1}^{N} \frac{\bar{k}_j \bar{k}_{j'}}{(\langle \bar{k} \rangle N)^2} = \frac{k_j}{(\langle \bar{k} \rangle N)}$

Therefore we have

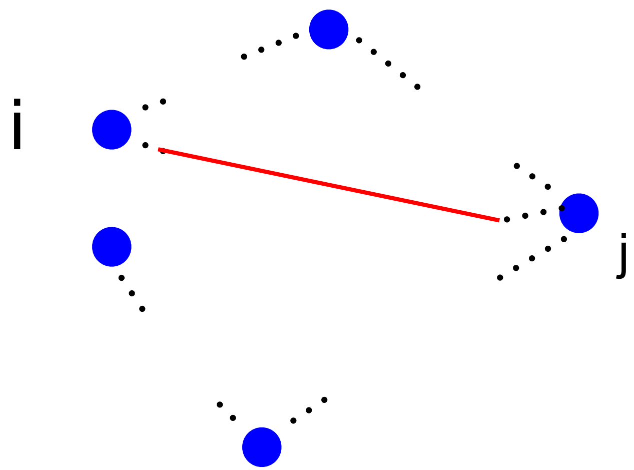
$$\pi_{i|j} = \frac{\pi_{ij}}{\sum_{j'=1}^{N} \pi_{jj'}} = \left(\frac{\bar{k}_i \bar{k}_j}{(\langle \bar{k} \rangle N)^2} \right) \left(\frac{\langle k \rangle N}{\bar{k}_j} \right) = \frac{\bar{k}_i}{\langle \bar{k} \rangle N} = q_i$$

Example

The probability that a random link connects node i to node j is given by

$$\pi_{ij} = \frac{\bar{k}_i \bar{k}_j}{(\langle \bar{k} \rangle N)^2}$$

Example



$\bar{k}_i$

$$\pi_{ij} = \frac{2}{\langle \bar{k} \rangle N} \frac{3}{\langle \bar{k} \rangle N}$$

The probability that the link connects one end to node i is $\frac{2}{\langle \bar{k} \rangle N}$

The probability that the link connects the other end to node j is $\frac{3}{\langle \bar{k} \rangle N}$

Marginal probability in uncorrelated simple networks

Proposition

In uncorrelated simple networks the probability that a node *i* is linked to a node *j* is given by

$$p_{ij} = \frac{\bar{k}_i \bar{k}_j}{\langle \bar{k} \rangle N}$$

Proof

In an uncorrelated network the expected number of links between node I and node j is given by

$$n_{ij} = 2\bar{L}\pi_{ij} = (\langle \bar{k} \rangle N)\frac{\bar{k}_i \bar{k}_j}{(\langle \bar{k} \rangle N)^2} = \frac{\bar{k}_i \bar{k}_j}{(\langle \bar{k} \rangle N)}$$

Since the network is by hypothesis simple

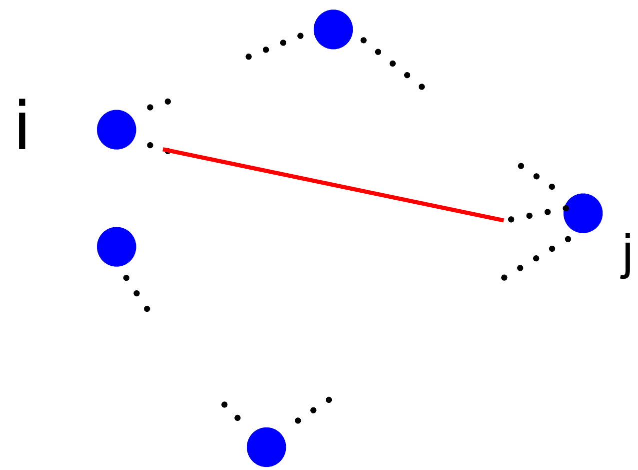
$$p_{ij} = \langle a_{ij} \rangle = n_{ij} = \frac{\bar{k}_i \bar{k}_j}{(\langle \bar{k} \rangle N)}$$

Example

The probability that a node connects node i to node j is given by

$$p_{ij} = \frac{\bar{k}_i \bar{k}_j}{(\langle \bar{k} \rangle N)}$$

Example



$\bar{k}_i$

$$p_{ij} = 2\frac{3}{\langle \bar{k} \rangle N}$$

The probability that one link of node i connects node i to node j is

$$\frac{3}{\langle \bar{k} \rangle N}$$

Since node i has an expected degree there is a factor 2

$$\bar{k}_i = 2$$

Structural cutoff

Simple uncorrelated networks must necessarily have the structural cutoff

$$K_S = \sqrt{\langle \bar{k} \rangle N}$$

i.e. the expected degrees of the nodes should be smaller than the structural cutoff

$$\max_i \bar{k}_i = K \leq K_S = \sqrt{\langle \bar{k} \rangle N}$$

Proof

In uncorrelated network the probability that two nodes are connected is

$$p_{ij} = \frac{\bar{k}_i \bar{k}_j}{\langle \bar{k} \rangle N} \leq 1 \, \forall i, j \in \{1,2,\ldots,N\}$$

Therefore taking $\bar{k}_i = \bar{k}_j = K = \max_n \bar{k}_n$ we must necessarily have

$$p_{ij} = \frac{K^2}{\langle \bar{k} \rangle N} \leq 1$$

It follows that

$$K \leq K_S = \sqrt{\langle \bar{k} \rangle N}$$

The natural cutoff of scale-free networks

For scale-free networks with degree distribution

$$P(k) \simeq Ck^{-\gamma}$$

the

natural cutoff

(maximum degree of a network of N nodes

if no constraint on the maximum degree is imposed

scales like

$$K = K_N \sim N^{\frac{1}{\gamma - 1}}$$

Natural and structural cutoff of scale-free networks

For scale-free networks with degree distribution

$$P(k) \simeq Ck^{-\gamma} \textbf{ for } k \gg 1$$

the

natural cutoff is larger than the structural cutoff

$$K_N \gg K_s = \sqrt{\langle k \rangle N}$$

for

$$\gamma \leq 3$$

Uncorrelated scale-free networks

Sparse uncorrelated networks with power-law exponent γ must have a maximum degree K (cutoff) that scales like

$$K \sim \min\left[N^{\frac{1}{\gamma-1}}, N^{\frac{1}{2}}\right]$$

Maximum entropy ensembles

Degree sequence as constraint

Expected degree sequence as constraint

Canonical ensemble or exponential random graph with given expected degree sequence

We consider the

canonical network ensemble

in which we impose the N soft constraints

$$\bar{k}_i = \sum_{G \in \Omega_G} \left[P(G) \left(\sum_{j=1}^{N} a_{ij} \right) \right] \quad i = 1,2,\ldots,N$$

Canonical ensemble

Proposition

The canonical ensemble in which we fix the expected degree sequence has Gibbs measure

$$P(\mathbf{a}) = \frac{1}{Z} e^{-\sum_{i=1}^{N} \lambda_i \sum_{j=1}^{N} a_{ij}}$$

Proof

This follow directly from the general Gibbs measure of canonical network ensemble

$$P(\mathbf{a}) = P(G) = \frac{e^{-\sum_{i=1}^{N} \lambda_i F_i(G)}}{Z}$$

where we take as constraints

$$P = N, \qquad F_i(G) = \sum_{j=1}^{N} a_{ij}, \qquad C_i = \bar{k}_i \qquad \text{for } i = 1,2\ldots,N$$

Marginal and equation for the Lagrangian multipliers

In the canonical ensemble with given expected degree sequence the marginal probability of a link (i,j)

$$p_{ij} = \sum_{\mathbf{a}} a_{ij} P(\mathbf{a})$$

is given by

$$p_{ij} = \frac{e^{-\lambda_i - \lambda_j}}{1 + e^{-\lambda_i - \lambda_j}}$$

where λ_i are the Lagrangian multipliers fixing the expected degrees, i.e. satisfying

$$\bar{k}_i = \sum_{j \neq i} p_{ij} = \sum_{j \neq i} \frac{e^{-\lambda_i - \lambda_j}}{1 + e^{-\lambda_i - \lambda_j}}$$

Natural correlations

Since the marginal probabilities

$$p_{ij} = \frac{e^{-\lambda_i - \lambda_j}}{1 + e^{-\lambda_i - \lambda_j}}$$

do not factorise in terms depending exclusively on single nodes,

the configuration model leads to

natural correlations

which are

disassortative

Evidence of disassortative correlations

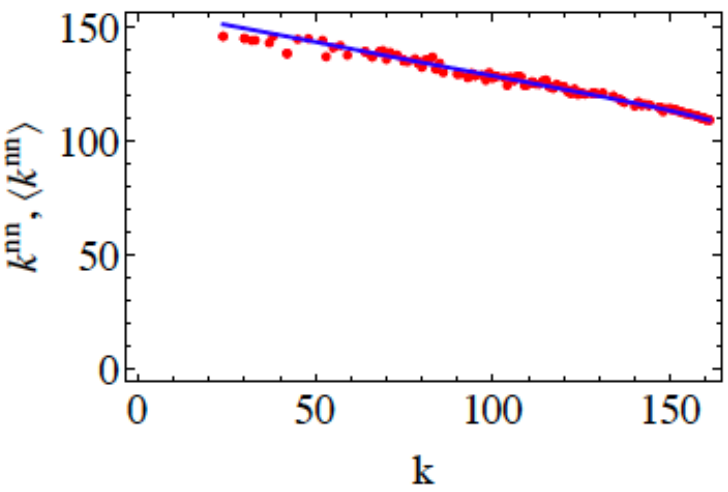
Average degree of the neighbour of a node in the data

$$k_{nn}(k) = \left\langle \frac{1}{k_i} \sum_{j=1}^{N} k_j a_{ij} \right\rangle_{k_i=k}$$

Expected average degree of the neighbour of a node in the canonical network ensemble

$$\langle k_{nn}(k) \rangle = \left\langle \frac{1}{k_i} \sum_{j=1}^{N} k_j p_{ij} \right\rangle_{k_i=k}$$

World-Trade network



Squartini, et al. Randomizing world trade I. (2011)

Uncorrelated limit

Only in presence of the structural cutoff

$$K_S = \sqrt{\langle \bar{k} \rangle N}$$

where the expected degree are bounded

$$\bar{k}_i \ll K_S = \sqrt{\langle \bar{k} \rangle N} \quad \forall i \in \{1,2,\ldots,N\}$$

The configuration model is an uncorrelated network and the marginal probabilities read

$$p_{ij} = \frac{\bar{k}_i \bar{k}_j}{\langle k \rangle N}$$

Proof

If we assume $e^{-\lambda_i} \ll 1$

We can express the marginals as $p_{ij} = \dfrac{e^{-\lambda_i - \lambda_j}}{1 + e^{-\lambda_i - \lambda_j}} \simeq e^{-\lambda_i - \lambda_j}$

Enforcing the expected degree we get

$$\bar{k}_i = \sum_{j=1}^{N} e^{-\lambda_i - \lambda_j} = e^{-\lambda_i} Q$$

Therefore

$$e^{-\lambda_i} = \frac{\bar{k}_i}{Q}$$

with Q defined as

$$Q = \sum_{j=1}^{N} e^{-\lambda_j} = \sum_{j=1}^{N} \frac{\bar{k}_j}{Q}$$

Proof (continuation)

The equation

$$Q = \sum_{j=1}^{N} e^{-\lambda_j} = \sum_{j=1}^{N} \frac{\bar{k}_j}{Q}$$

implies that

$$Q^2 = \sum_{j=1}^{N} \bar{k}_j = \langle \bar{k} \rangle N$$

Therefore

$$Q = \sqrt{\langle \bar{k} \rangle N}$$

By inserting this equation in the expression for the Lagrangian multiplier

$$e^{-\lambda_i} = \frac{\bar{k}_i}{Q} = \frac{\bar{k}_i}{\sqrt{\langle k \rangle N}}$$ **and** $$p_{ij} = \frac{\bar{k}_i \bar{k}_j}{\langle k \rangle N}$$

We get that the initial hypothesis is only satisfied for

$$e^{-\lambda_i} \ll 1 \textbf{ iff } k_i \ll \sqrt{\langle \bar{k} \rangle N}$$

Entropy of the ensemble

Given that the Gibbs entropy for the canonical ensemble with given expected degrees factories in single links contributions

$$P(\mathbf{a}) = \prod_{i<j} p_{ij}^{a_{ij}}(1-p_{ij})^{1-a_{ij}}$$

The entropy of the canonical ensemble

$$S = -\sum_{\mathbf{a}} P(\mathbf{a}) \ln P(\mathbf{a})$$

can be written as

$$S = -\sum_{i<j}^{N} \left[p_{ij} \ln p_{ij} + (1-p_{ij}) \ln(1-p_{ij}) \right]$$

Entropy of the canonical ensemble

In the uncorrelated limit, when the marginal probabilities are given by

$$p_{ij} = \frac{\bar{k}_i \bar{k}_j}{\langle \bar{k} \rangle N}$$

The entropy of the canonical ensemble

$$S = -\sum_{i<j}^{N} \left[p_{ij} \ln p_{ij} + (1 - p_{ij}) \ln(1 - p_{ij}) \right]$$

can be written as

$$S = -\sum_{i<j}^{N} \left[\frac{\bar{k}_i \bar{k}_j}{\langle \bar{k} \rangle N} \ln \frac{\bar{k}_i \bar{k}_j}{\langle \bar{k} \rangle N} + \left(1 - \frac{\bar{k}_i \bar{k}_j}{\langle \bar{k} \rangle N}\right) \ln \left(1 - \frac{\bar{k}_i \bar{k}_j}{\langle \bar{k} \rangle N}\right) \right]$$

Entropy of the canonical ensemble in the uncorrelated network limit

In the uncorrelated limit, the entropy of the canonical ensemble scales like

$$S \simeq \underbrace{\frac{1}{2}(\langle\bar{k}\rangle N)\ln(\langle\bar{k}\rangle N)}_{\mathcal{O}(N\ln N)} - \underbrace{\sum_{i=1}^{N}\bar{k}_i\ln\bar{k}_i + \frac{1}{2}\langle\bar{k}\rangle N}_{\mathcal{O}(N)} - \frac{1}{4}\underbrace{\left(\frac{\langle\bar{k}^2\rangle}{\langle\bar{k}\rangle}\right)^2}_{o(N)}$$

$\mathcal{O}(N\ln N)$: **Only dependent on the average degree**

$\mathcal{O}(N)$: **Dependent on the degree distribution**

$o(N)$: **Sublinear but diverging with N for power-law networks**

Proof

In the uncorrelated limit, the entropy of the canonical ensemble is given by

$$S = -\frac{1}{2}\sum_{i,j}^{N}\left[\frac{\bar{k}_i\bar{k}_j}{\langle\bar{k}\rangle N}\ln\frac{\bar{k}_i\bar{k}_j}{\langle\bar{k}\rangle N} + \left(1-\frac{\bar{k}_i\bar{k}_j}{\langle\bar{k}\rangle N}\right)\ln\left(1-\frac{\bar{k}_i\bar{k}_j}{\langle\bar{k}\rangle N}\right)\right]$$

Using the expansions

$$\ln(1-x) \simeq -x - \frac{1}{2}x^2 \textbf{ for } x \ll 1$$

$$(1-x)\ln(1-x) \simeq -x + \frac{1}{2}x^2 \textbf{ for } x \ll 1$$

with $x = \frac{\bar{k}_i\bar{k}_j}{\langle\bar{k}\rangle N}$

$$S = \frac{1}{2}(\langle\bar{k}\rangle N)\ln(\langle\bar{k}\rangle N) - \sum_{i=1}^{N}\bar{k}_i\ln\bar{k}_i + \frac{1}{2}\langle\bar{k}\rangle N - \frac{1}{4}\left(\frac{\langle\bar{k}^2\rangle}{\langle\bar{k}\rangle}\right)^2$$

Proof

In the uncorrelated limit, the entropy of the canonical ensemble scales like

$$S \simeq \frac{1}{2}(\langle \bar{k} \rangle N)\ln(\langle \bar{k} \rangle N) - \sum_{i=1}^{N} \bar{k}_i \ln \bar{k}_i + \frac{1}{2}\langle \bar{k} \rangle N - \frac{1}{4}\left(\frac{\langle \bar{k}^2 \rangle}{\langle \bar{k} \rangle}\right)^2$$

Using the entropy of the random graph G(N,p) we get

$$S_{G(N,p=\langle k \rangle/N} \simeq \frac{1}{2}(\langle \bar{k} \rangle N)\ln(\langle \bar{k} \rangle N) - N\langle \bar{k} \rangle \ln\langle \bar{k} \rangle + \frac{1}{2}\langle \bar{k} \rangle N$$

can be written as

$$S \simeq S_{G(N,p=\langle \bar{k} \rangle/N)} - \sum_{i=1}^{N} \bar{k}_i \ln \bar{k}_i + N\langle \bar{k} \rangle \ln(\langle \bar{k} \rangle) - \frac{1}{4}\left(\frac{\langle \bar{k}^2 \rangle}{\langle \bar{k} \rangle}\right)^2$$

Entropy of the canonical ensemble in the uncorrelated network limit

In the uncorrelated network limit, the entropy of the canonical ensemble scales like

$$S \simeq S_{G(N,p=\langle \bar{k} \rangle/N)} - \sum_{i=1}^{N} \bar{k}_i \ln \bar{k}_i + N\langle \bar{k} \rangle \ln(\langle \bar{k} \rangle) - \frac{1}{4}\left(\frac{\langle \bar{k}^2 \rangle}{\langle \bar{k} \rangle}\right)^2$$

$\mathcal{O}(N \ln N)$

Only dependent on the average degree

$\mathcal{O}(N)$

Dependent on the degree distribution

$o(N)$

Sublinear but diverging with N for power-law networks

True degree distribution of node I in the uncorrelated limit

In the uncorrelated network limit

the probability that node i has degree k_i

is given by a Poisson distribution

with average given by the expected degree $\bar{k}_i$ of node i

$$\mathbb{P}(k_i = k) = \frac{\bar{k}_i^k}{k!} e^{-\bar{k}_i}$$

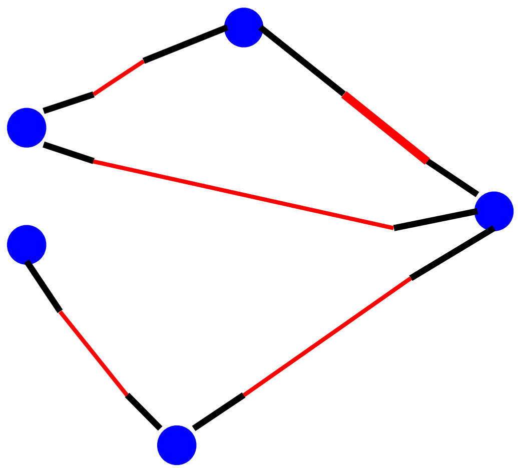
Microcanonical network ensemble

Configuration model

Microcanonical network ensemble
The configuration model

Probability of a network

$$P(G) = \frac{1}{Z_M} \prod_{i=1}^{N} \delta\left(k_i, \sum_{j=1}^{N} a_{ij}\right)$$



Ensemble of networks with exact
degree sequence

Graphicality

A degree sequence is graphical

if it is the degree sequence of

at least a simple network

Not all degree sequence are graphical!

Erdö-Gallai Theorem

A non-decreasing degree sequence $\{k_1, k_2, \ldots, k_N\}$

is graphical if and only if the following two conditions are satisfied:

1. the sum of the degree is even;

2. for all $1 \leq m < N$ we have

$$\sum_{i=1}^{m} k_i \leq m(m-1) + \sum_{i=m+1}^{N} \min(m, k_i)$$

Solution to graphicality problem

- Check directly for graphicality of the degree sequence
- If the degree sequence is not graphical search for minimal modifications that can make the degree sequence graphical
- Start from the degree sequence of a real network (null model)

Generation of networks in the configuration model

A. Consider a graphical degree sequence $\{k_1, k_2, \ldots, k_N\}$

B. Assign k_i half-stubs to each node i

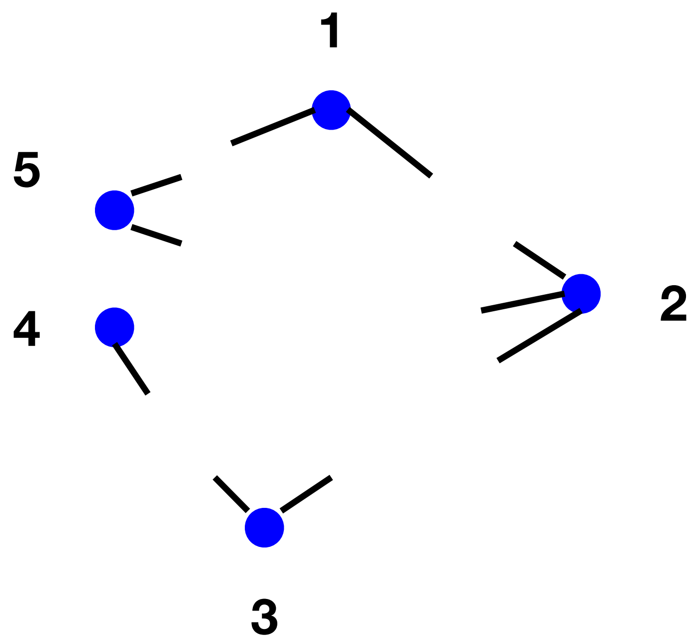
C. Randomly match the half-stubs

D. If in the process tadpoles or multiple edges are generated start from point B.

Example

Assign k_i half-stubs on each node i

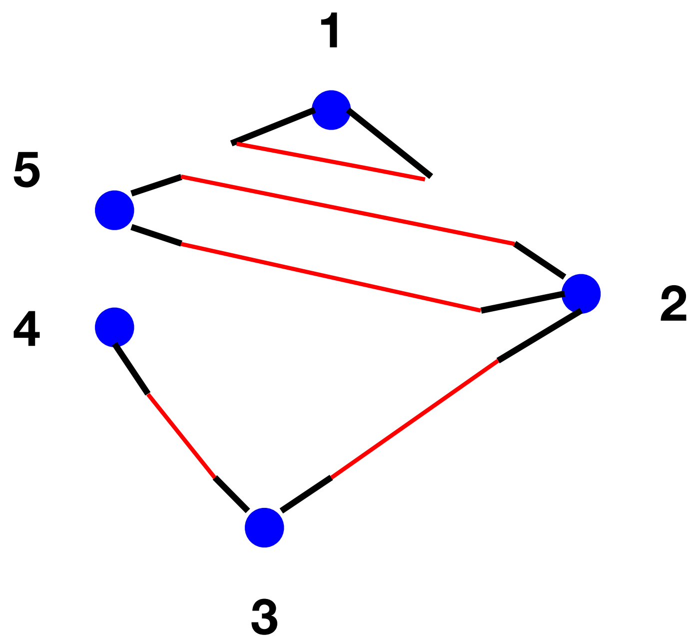
$$\{k_1, k_2, k_3, k_4, k_5\} = \{2,3,4,1,2\}$$



Example

Randomly match the half-stubs

$$\{k_1, k_2, k_3, k_4, k_5\} = \{2,3,4,1,2\}$$



(1,1): Tadpole

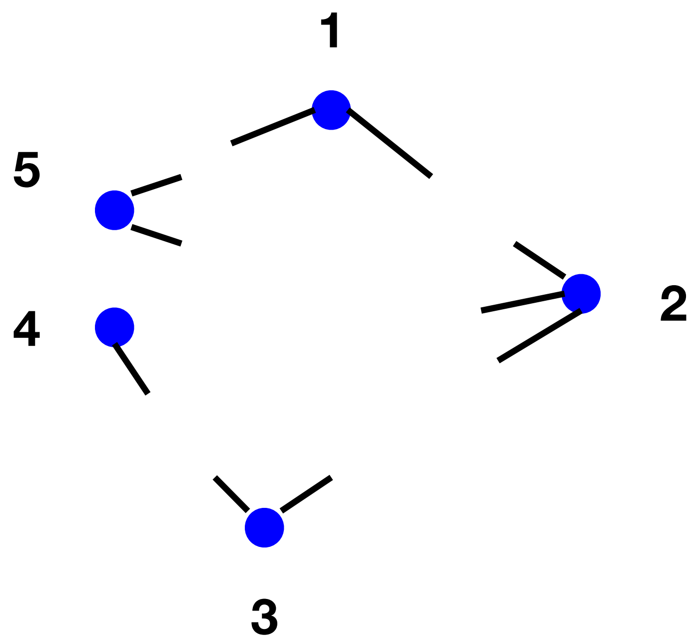
(2,5): Multiple edge

This network realisation should be discarded

Example

Restart from the beginning

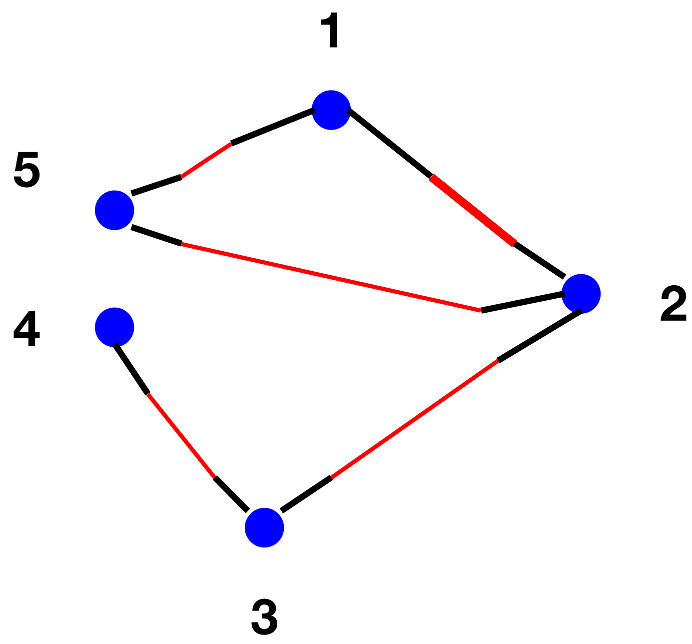
$$\{k_1, k_2, k_3, k_4, k_5\} = \{2,3,4,1,2\}$$



Example

Randomly match the half-stubs

$$\{k_1, k_2, k_3, k_4, k_5\} = \{2,3,4,1,2\}$$



This network realisation is viable

Entropy of the micro canonical ensemble

Proposition

The entropy of the micro canonical ensemble is given by

$$\Sigma = -\sum_{G \in \Omega_G | \{F_\mu(G) = C_\mu\}_{\mu=1,2\ldots,P}} P(G) \ln P(G) = \ln Z_M$$

Proof

In fact we have

$$P(G) = \frac{1}{Z_M} \prod_{\mu=1}^{P} \delta\left(F_\mu(G), C_\mu\right) \quad \textbf{with} \quad Z_M = \sum_{G \in \Omega_G} \prod_{\mu=1}^{P} \delta\left(F_\mu(G), C_\mu\right)$$

Therefore

$$S = -\sum_{G \in \Omega_G | \{F_\mu(G) = C_\mu\}_{\mu=1,2\ldots,P}} \frac{1}{Z_M} \ln\left(\frac{1}{Z_M}\right) = \ln Z_M$$

Entropy of conjugated ensembles

Proposition

The entropy of a micro canonical ensemble Σ and the entropy S of the conjugated canonical ensemble are related by

$$\Sigma = S - \Omega$$

where

$$\Omega = -\ln \sum_{G \in \Omega_G} P_C(G) \prod_{\mu=1}^{P} \delta\left(F_\mu(G), C_\mu\right) \qquad P_C(G) = \frac{1}{Z} e^{-\sum_{\mu=1}^{P} \lambda_\mu F_\mu(G)}$$

and where $\delta(x, y)$ indicates the Kronecker delta.

Proof

Our aim is to calculate

$$\Omega = -\ln \sum_{G \in \Omega_G} P_C(G) \prod_{\mu=1}^{P} \delta\left(F_\mu(G), C_\mu\right)$$

where

$$P_C(G) = \frac{1}{Z} e^{-\sum_{\mu=1}^{P} \lambda_\mu F_\mu(G)}$$

By inserting this explicit expression we obtain

$$\Omega = -\ln\left[\sum_{G \in \Omega_G} \frac{1}{Z} e^{-\sum_{\mu=1}^{P} \lambda_\mu F_\mu(G)} \prod_{\mu=1}^{P} \delta\left(F_\mu(G), C_\mu\right)\right] = -\ln\left[\sum_{G \in \Omega_G} \frac{1}{Z} e^{-\sum_{\mu=1}^{P} \lambda_\mu C_\mu} \prod_{\mu=1}^{P} \delta\left(F_\mu(G), C_\mu\right)\right]$$

$$\Omega = -\ln\left[\frac{1}{Z} e^{-\sum_{\mu=1}^{P} \lambda_\mu C_\mu} \sum_{G \in \Omega_G} \prod_{\mu=1}^{P} \delta\left(F_\mu(G), C_\mu\right)\right] = -\ln\left[e^{-S} Z_M\right] = -\ln e^{-S+\Sigma} = S - \Sigma$$

Entropy of micro canonical network ensemble

The entropy of the microcanonical ensemble Σ is given by

$$\Sigma = S - \Omega$$

where

$$\Omega = -\ln\left[\sum_{\mathbf{a}} P_C(\mathbf{a}) \prod_{i=1}^{N} \delta\left(k_i, \sum_{j=1}^{N} a_{ij}\right)\right]$$

and

$$P_C(\mathbf{a}) = \frac{e^{-\sum_{i=1}^{N} \lambda_i \sum_{j=1}^{N} a_{ij}}}{Z} \qquad \bar{k}_i = \sum_{\mathbf{a}} P_C(\mathbf{a}) \left[\sum_{j=1}^{N} a_{ij}\right] = k_i \ \forall i \in \{1,2,\ldots,N\}$$

K. Anand & G. Bianconi (2010)

Entropy of the microcanonical ensemble

In the uncorrelated limit we have

$$\Sigma = \ln Z_M = S - \Omega$$

$$\Omega = -\sum_{i=1}^{N} \ln\left(\frac{k_i^{k_i}}{k_i!} e^{-k_i}\right)$$

Ω is extensive (order N)

There is no equivalence of the canonical and microcanonical ensembles

Bianconi et. al (2008) K. Anand & G. Bianconi (2010)

There is no equivalence of the ensembles as long as the number of constraints is extensive

Example

Microcanonical esemble

Regular networks

$$p_{ij} = \frac{\langle k \rangle}{N}$$

Canonical ensemble

Poisson networks

$$p_{ij} = \frac{\langle k \rangle}{N}$$

but

$$\Sigma < S$$

K. Anand, G. Bianconi PRE 2009

Canfield-Bender formula

The asymptotic formula
for the number of networks in
the uncorrelated network limit
of the configuration model
is given by

$$\mathcal{N} = Z_M \simeq \frac{(\langle k \rangle N)!!}{\prod_{i=1}^{N} k_i!} e^{-\frac{1}{4}\left(\frac{\langle k^2 \rangle}{\langle k \rangle}\right)^2}$$

Combinatorial meaning of the Canfield-Bender formula

$$\mathcal{N} = Z_M \simeq \frac{(\langle k \rangle N)!!}{\prod_{i=1}^{N} k_i!} e^{-\frac{1}{4}\left(\frac{\langle k^2 \rangle}{\langle k \rangle}\right)^2}$$

$$(\langle k \rangle N)!!$$

Number of possible matchings between the half-stubs

$$\prod_{i=1}^{N} k_i!$$

Degeneracy of the counting due to the permutation of the half-stubs of each node

$$e^{-\frac{1}{4}\left(\frac{\langle k^2 \rangle}{\langle k \rangle}\right)^2}$$

Asymptotic correction for obtaining simple networks

Proof

Starting from

$$\Sigma = S - \Omega$$

with

$$S \simeq \frac{1}{2}(\langle \bar{k} \rangle N)\ln(\langle \bar{k} \rangle N) - \sum_{i=1}^{N} \bar{k}_i \ln \bar{k}_i + \frac{1}{2}\langle \bar{k} \rangle N - \frac{1}{4}\left(\frac{\langle \bar{k}^2 \rangle}{\langle \bar{k} \rangle}\right)^2$$

and

$$\Omega = -\sum_{i=1}^{N} \ln\left(\frac{1}{k_i!} k_i^{k_i} e^{-k_i}\right) = -\sum_{i=1}^{N} k_i \ln k_i + \sum_{i=1}^{N} \ln k_i! + \sum_{i=1}^{N} k_i$$

We can express Σ as

$$\Sigma \simeq \frac{1}{2}(\langle \bar{k} \rangle N)\ln(\langle \bar{k} \rangle N) - \sum_{i=1}^{N} \ln \bar{k}_i! - \frac{1}{2}\langle \bar{k} \rangle N - \frac{1}{4}\left(\frac{\langle \bar{k}^2 \rangle}{\langle \bar{k} \rangle}\right)^2$$

Therefore we have

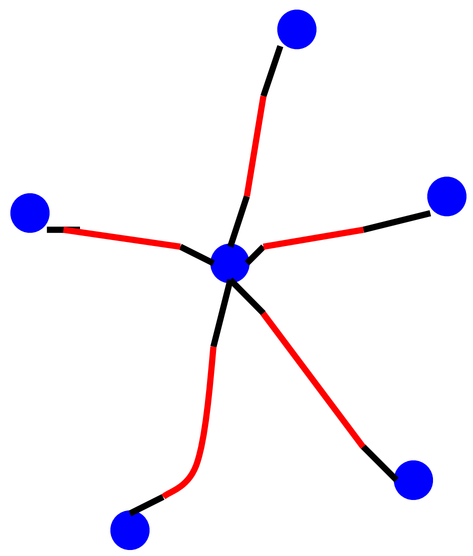
$$Z_M = e^{\Sigma} \sim \frac{(\langle k \rangle N)!!}{\prod_{i=1}^{N} k_i!} e^{-\frac{1}{4}\left(\frac{\langle \bar{k}^2 \rangle}{\langle \bar{k} \rangle}\right)^2}$$

The entropy of the canonical ensemble depends on the degree distribution

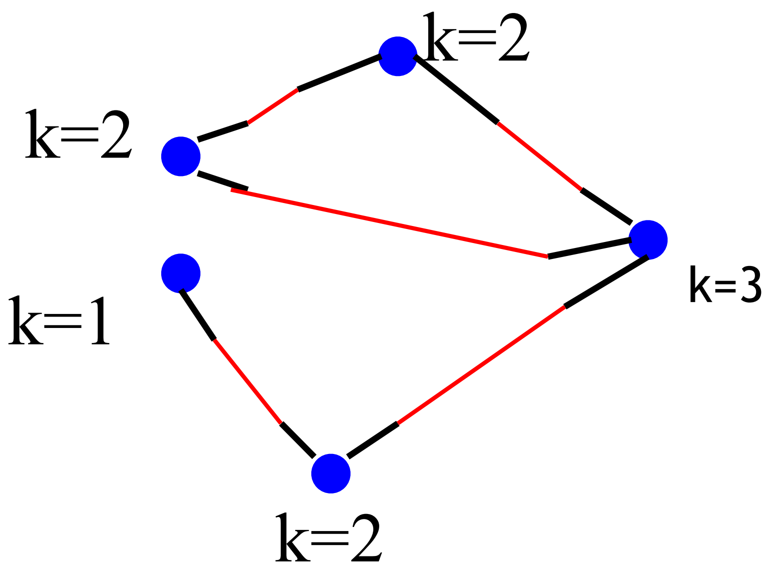
Exponential random graphs with the same average degree but different degree distribution have different entropy

Two examples of given degree sequence

Zero entropy

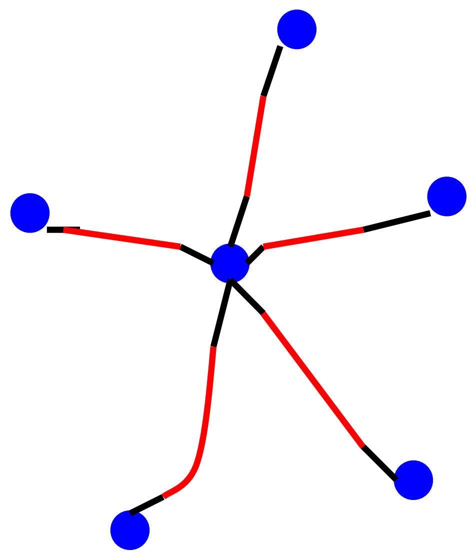


Non-zero entropy

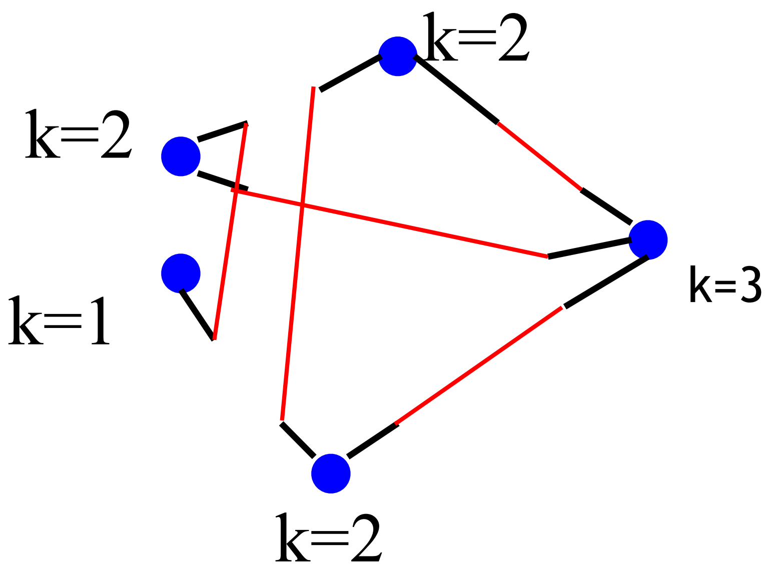


Two examples of given degree sequence

Zero entropy

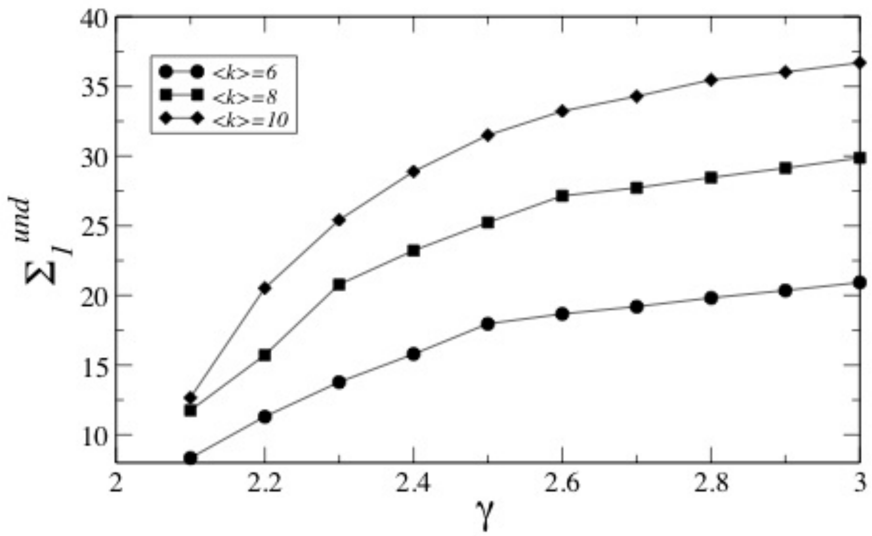


Non-zero entropy



The entropy of random scale-free networks

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



The entropy decreases as $\gamma \to 2$
quantifying a higher order in networks with fatter tails

Randomization

Edge swaps randomisation

can be shown

to provide a

biased sampling

of the networks in the configuration model

Metropolis-Hastings algorithm

- Start from a given network of N nodes and given target degree sequence.
- Iterate the following procedure until convergence of observables

1. Pick randomly a pair of links (i, j) (r, s)
2. If allowed perform a edge swap transition $\mathbf{a} \rightarrow \mathbf{a}'$ with probability

$$\Pi_{\mathbf{a}\rightarrow\mathbf{a}'} = \textbf{min}\left[1, \frac{P(\mathbf{a}')\,|\,\Phi_{\mathbf{a}}\,|}{P(\mathbf{a})\,|\,\Phi_{\mathbf{a}'}\,|}\right]$$

Where $|\Phi_{\mathbf{a}}|$ indicates the number of viable edge swaps allowed starting from adjacency matrix $\mathbf{a}$

Coolen, De Martino,Annibale, J Stat Phys (2009) 136: 1035–1067

Average number of loops of finite size L

In the uncorrelated network limit the number of loops of finite size L are given by

$$\langle \mathcal{N}_L \rangle = \frac{1}{2L}\left(\frac{\langle k(k-1)\rangle}{\langle k \rangle}\right)^L$$

- For Poisson networks we have $\langle k(k-1)\rangle = \langle k \rangle^2 = \mathcal{O}(1)$

therefore the number of small loops is finite even in an infinite network

- Scale free networks the second moment diverge $\langle k(k-1)\rangle = \mathcal{O}(N^{(3-\gamma)/2})$

therefore we have an infinite number of small loops

G. Bianconi, M. Marsili JSTAT (2005)

Expected clustering coefficient

In the uncorrelated network limit the expected average clustering coefficient of a node is independent of the degree of the node and given is given by

$$\langle C_i | k_i \rangle = \frac{1}{3\langle k \rangle N} \left(\frac{\langle k(k-1) \rangle}{\langle k \rangle} \right)^2$$

- For Poisson networks we have $\langle k(k-1) \rangle = \langle k \rangle^2 = \mathcal{O}(1)$

therefore $\langle C_i | k_i \rangle = \mathcal{O}(N^{-1})$

- For scale free networks we have $\langle k(k-1) \rangle = \mathcal{O}(N^{(3-\gamma)/2})$

and we still observe vanishing average clustering coefficient

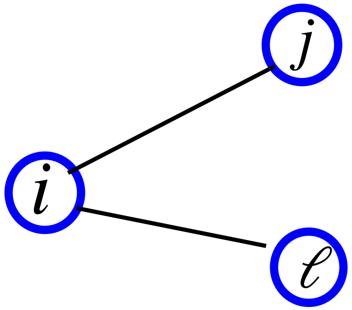
$$\langle C_i | k_i \rangle = \mathcal{O}(N^{2-\gamma})$$

Phase transitions in

Maximum Entropy Ensembles

2-star model

A wedge is a triple of nodes connected by two links



$$a_{ij}a_{i\ell} = 1$$

The 2 star model is

the maximum entropy canonical network model

in which we fix

- the expected total number of links
- the expected number of wedges

The soft constraints of the 2 star model

2 star model

In this case we impose the expected total number of links as a soft constraint

$$\sum_{G \subset \Omega_G} \left[\sum_{i<j} a_{ij} \right] P(G) = \bar{L}$$

and the expected total number of wedges as a soft constraint

$$\sum_{G \in \Omega_G} \left(\sum_{i}^{N} \sum_{j \neq \ell | j, \ell \neq i} a_{ij} a_{i\ell} \right) P(G) = \bar{C}$$

Phase transition in the 2-star model

By solving the 2 star model in the mean-field approximation

a first order phase transition is found

between a low density phase and

a high density phase

including a region of the phase-space

with coexistence of the two phases.

Probability of a network in the 2 star model

According to the general theory of canonical network ensemble the probability of a network can be expressed as

$$P(G) = \frac{1}{Z}\sum_{\mathbf{a}} \exp\left[\lambda \sum_{i<j} a_{ij} + \gamma \sum_{i=1}^{N} \sum_{j\neq\ell,\ell\neq j} a_{ij}a_{i\ell}\right] = \frac{e^{-H(G)}}{Z}$$

with Hamiltonian given by

$$H(G) = -\lambda \sum_{i<j} a_{ij} - \gamma \sum_{i} \sum_{j\neq\ell} a_{ij}a_{i\ell}$$

where λ and γ are Lagrangian multipliers enforcing the constraints

Mean-field approximation

In the mean field approximation we neglect correlations and we put

$$a_{ij}a_{j\ell} \simeq a_{ij}\langle a_{j\ell}\rangle + \langle a_{ij}\rangle a_{j\ell} - \langle a_{ij}\rangle\langle a_{j\ell}\rangle$$

which gives

$$\langle a_{ij}a_{j\ell}\rangle \simeq \langle a_{ij}\rangle\langle a_{j\ell}\rangle$$

Where we assume that the marginal of each link is the same and equal to p, i.e.

$$\langle a_{ij}\rangle = p \ \forall i, j$$

Mean-field approximation

By inserting the mean-field approximation

$$a_{ij}a_{j\ell} \simeq a_{ij}p + a_{j\ell}p - p^2$$

In the expression for the Hamiltonian $H(G) = -\beta \sum_{i<j} a_{ij} - \gamma \sum_i \sum_{j \neq \ell} a_{ij}a_{i\ell}$

We get

$$\begin{aligned}
H_{MF}(G) &= -\beta \sum_{i<j} a_{ij} - \gamma \sum_i \sum_{\ell \neq j, \ell \neq i} [a_{ij}p + a_{j\ell}p - p^2] \\
&= -\beta \sum_{i<j} a_{ij} - \gamma \left[\sum_{i,j} a_{ij} \sum_{\ell \neq j, \ell \neq i} p + \sum_{j\ell} a_{j\ell} \sum_{i \neq j, i \neq \ell} p \right] + C \\
&\simeq -\beta \sum_{i<j} a_{ij} - 4\gamma p N \sum_{i<j} a_{ij} + C = -\sum_{i<j} a_{ij}(\beta + 4N\gamma p) + C
\end{aligned}$$

Self-consistent equation

Assuming that p is known and that the Hamiltonian of the network ensemble is given by its mean-field approximation

$$H_{MF}(G) \simeq -\sum_{i<j} a_{ij}(\beta + 4N\gamma p) - C$$

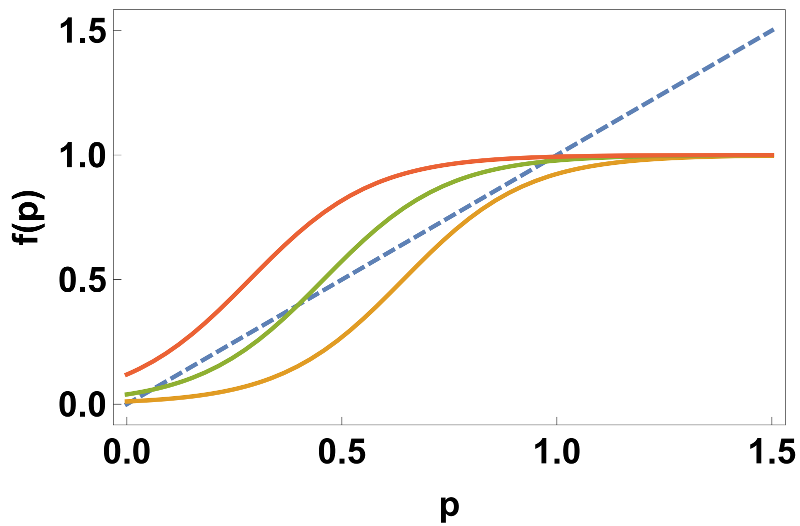
We can calculate the marginal which leads to

the self-consistent equation for p given by

$$p = f(p) = \frac{e^{\beta+4N\gamma p}}{1 + e^{\beta+4N\gamma p}}$$

Phase transition in the 2-star model

$$p = f(p) = \frac{e^{\beta+4N\gamma p}}{1+e^{\beta+4N\gamma p}}$$



For some values of the Lagrangian multipliers

there are two stable solutions at

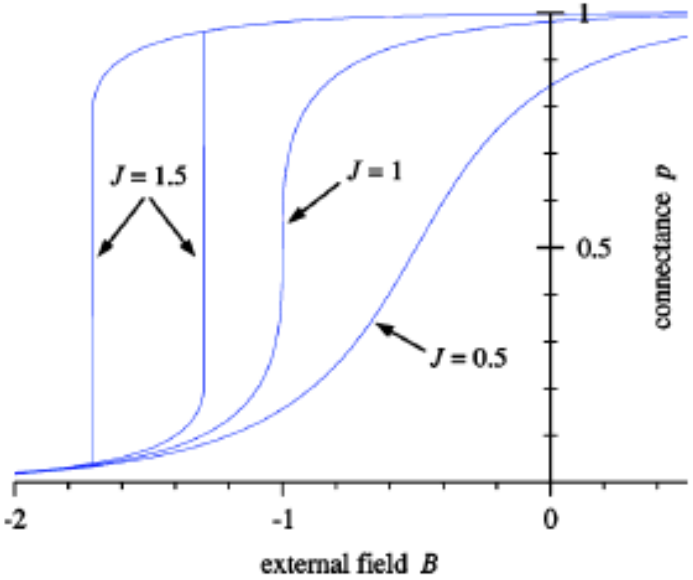
high density (high value of p) and low density (low values of p)

Phase transition in the 2-star model

By putting

$$B = \frac{\beta}{2} \quad J = \gamma N$$

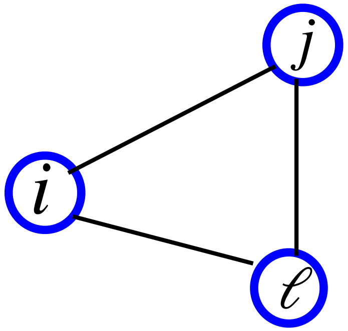
the phase diagram of p as a function of B is given by



Park and Newman (2004)

Strauss model

A triangle is a triple of nodes connected by three links



$$a_{ij}a_{j\ell}a_{\ell i} = 1$$

The Strauss model is

the maximum entropy canonical network model

in which we fix

- the expected total number of links
- the expected number of triangles

The soft constraints of the Strauss model

Strauss model

In this case we impose the expected total number of links as a soft constraint

$$\sum_{G \subset \Omega_G} \left[\sum_{i<j} a_{ij} \right] P(G) = \bar{L}$$

and the expected total number of triangles as a soft constraint

$$\sum_{G \in \Omega_G} \left(\sum_{i<j<\ell} a_{ij} a_{i\ell} a_{j\ell} \right) P(G) = \bar{C}$$

Phase transition in the Strauss model

By solving the Strauss model in the mean-field approximation

a first order phase transition is found

between a low density phase and

a high density phase

including a region of the phase-space with coexistence of the two phases.

In the high density phase one observes a

condensation phenomena

where the network is decomposed in a high density phase including all the triangles and into several disconnected nodes and clusters.