Maximum Entropy Network Ensembles

LTCC Course
Lesson 3

Ginestra Bianconi

School of Mathematical Sciences Queen Mary University of London

Queen Mary
University of London

Microcanonical

and

Canonical

Network Ensembles

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.

Network Ensemble

Definition (for simple networks)

A network ensemble $\mathcal{G}$ is a triple $(G, \Omega_G, P(G))$ where G is any possible network $G = (E . V)$ belonging to the set of all simple networks with N nodes Ω_G and $P(G) \geq 0$ with $\sum_{G \in \Omega_G} P(G) = 1$ is the probability associate to each graph G

Generalization

The definition can be extended to non simple networks such as directed, weighted networks and also to generalised network structures by suitably changing the definition of Ω_G

Entropy of network ensembles

Definition

The *entropy of a network ensemble* is given by

$$S = -\sum_{G \in \Omega_G} P(G) \ln P(G)$$

It can be thought as the logarithm of the typical number of networks in the ensemble.

Here we have chosen the natural logarithm for simplicity

Constraints

We distinguish between soft constraints and hard constraints.

The ***soft constraints*** are the constraints satisfied in average over the ensemble of networks.

$$\sum_{G\in\Omega_G} F_\mu(G)P(G) = C_\mu \textbf{ for } \mu = 1,2\ldots,P$$

The ***hard constraints*** are the constraints satisfied by each network in the ensemble.

$$F_\mu(G) = C_\mu \textbf{ for } \mu = 1,2\ldots,P$$

Anand Bianconi 2009

Examples of hard constraints

$$\boxed{F_\mu(G) = C_\mu \text{ for } \mu = 1,2\ldots,P}$$

- Example 1:We can fix the total number of links L

$$\sum_{i<j} a_{ij} = L \Rightarrow \begin{cases} P = 1 \\ F_1(G) = \sum_{i<j} a_{ij} \\ C_1 = L \end{cases}$$

- Example 2: We can fix the entire degree sequence

$$\sum_{j=1}^{N} a_{ij} = k_i \text{ for } i = 1,2,\ldots,N \Rightarrow \begin{cases} P = N \\ F_i(G) = \sum_{j=1}^{N} a_{ij} \\ C_i = k_i \end{cases}$$

Examples of soft constraints

$$\boxed{\sum_{G\in\Omega_G} F_\mu(G)P(G) = C_\mu \textbf{ for } \mu = 1,2\ldots,P}$$

- Example 1:We can fix the expected total number of links $\bar{L}$

$$\sum_{G\in\Omega_G}\left(\sum_{i<j} a_{ij}\right)P(G) = \bar{L} \quad \Rightarrow \quad \begin{cases} P = 1 \\ F_1(G) = \sum_{i<j} a_{ij} \\ C_1 = \bar{L} \end{cases}$$

- Example 2: We can fix the expected degree sequence

$$\sum_{G\in\Omega_G}\left(\sum_{j=1}^{N} a_{ij}\right)P(G) = \bar{k}_i \textbf{ for } i = 1,2,\ldots,N \quad \Rightarrow \quad \begin{cases} P = N \\ F_i(G) = \sum_{j=1}^{N} a_{ij} \\ C_i = \bar{k}_i \end{cases}$$

Canonical and microcanical ensembles

- The **microcanonical ensemble** is the maximum entropy ensemble satisfying a given set of hard constraints of the type

$$F_\mu(G) = C_\mu \textbf{ for } \mu = 1,2\ldots,P$$

- The **canonical ensemble** is the maximum entropy ensemble satisfying a given set of soft constraints of the type

$$\sum_{G\in\Omega_G} F_\mu(G)P(G) = C_\mu \textbf{ for } \mu = 1,2\ldots,P$$

Anand Bianconi 2009

Conjugated ensembles

A microcanonical ensemble and a canonical ensemble

are **conjugated**

when they satisfy corresponding constraints,

i.e. when they satisfy

$$F_\mu(G) = C_\mu \textbf{ for } \mu = 1,2\ldots,P$$

$$\sum_{G\in\Omega_G} F_\mu(G)P(G) = C_\mu \textbf{ for } \mu = 1,2\ldots,P$$

with the same choice of $F_\mu(G)$ and C_μ respectively.

Canonical network ensemble

Proposition

The canonical ensemble satisfying the set of soft constraints

$$\boxed{\sum_{G \in \Omega_G} F_\mu(G) P(G) = C_\mu \ \textbf{for} \ \ \mu = 1,2\ldots, P}$$

is determined by a probability given by

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

where Z is a normalisation constant $H(G) = \sum_{\mu=1}^{P} \lambda_\mu F_\mu(G)$ is called the Hamiltonian

and the Lagrangian multipliers λ_μ are fixed by the constraints.

For this reason the canonical network ensembles are also called exponential random graphs

Log-likelihood

Consider a network G coming from an unknown network ensemble $P(G)$

We assume that the unknown distribution of the ensemble is coming from an ensemble with distribution $P_{\vec{\lambda}}(G)$ dependent on the parameters $\vec{\lambda}$

Definition

The *log-likelihood* of a parameters $\vec{\lambda}$ is defined as

$$\boxed{\mathscr{L}(\vec{\lambda}|G) = -\ln P_{\vec{\lambda}}(G)}$$

Maximum likelihood estimation

The maximum likelihood estimation of the parameters $\overrightarrow{\lambda^\star}$

corresponding to the distribution $P_{\overrightarrow{\lambda^\star}}(G)$

that best approximate the observed network

(according to maximum likelihood estimation) takes the form

$$\boxed{\overrightarrow{\lambda}^\star = \mathbf{argmax}_{\overrightarrow{\lambda}} \mathscr{L}(\overrightarrow{\lambda}\,|\,G) = \mathbf{argmin}_{\overrightarrow{\lambda}} \left[-\ln P_{\overrightarrow{\lambda}}(G)\right]}$$

Relation between maximum entropy and maximum likelihood

Assuming that $P_{\vec{\lambda}}(G)$ is the Gibbs measures of the type

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

Maximum likelihood estimation of the parameters $\vec{\lambda}^\star$

$$\vec{\lambda}^\star = \textbf{argmax}_{\vec{\lambda}} \mathscr{L}(\vec{\lambda} \,|\, G)$$

Implies that $P_{\vec{\lambda}}(G)$ is the maximum entropy ensemble with constraints fixed by the data

$$\boxed{F_\mu(G) = \langle F_\mu(G) \rangle_{ENSEMBLE} = \sum_{G' \in \Omega_G} P_{\vec{\lambda}}(G') F_\mu(G')}$$

Proof

Minimising the negative log-likelihood

$$-\mathscr{L}(\overrightarrow{\lambda}\,|\,G) = -\ln P_{\overrightarrow{\lambda}}(G) = \sum_{\mu} \lambda_\mu F_\mu(G) + \ln Z$$

We get

$$0 = \frac{\partial \mathscr{L}(\overrightarrow{\lambda}\,|\,G)}{\partial \lambda_\mu} = F_\mu(G) + \frac{\partial \ln Z}{\partial \lambda_\mu} \textbf{ for } \mu = 1,2\ldots,P$$

Therefore

$$F_\mu(G) = -\frac{\partial \ln Z}{\partial \lambda_\mu} = \sum_{G' \in \Omega_G} P_{\overrightarrow{\lambda}}(G')F_\mu(G') \textbf{ for } \mu = 1,2,\ldots,P$$

Therefore we have

$$\boxed{F_\mu(G) = \langle F_\mu(G) \rangle_{ENSEMBLE} = \sum_{G' \in \Omega_G} P_{\overrightarrow{\lambda}}(G')F_\mu(G')}$$

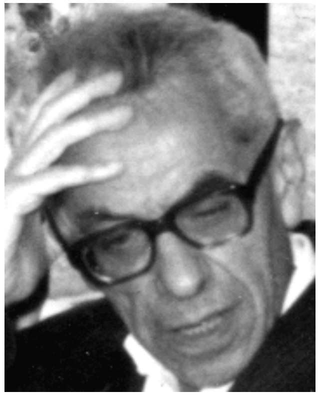
Random graphs

Random graphs

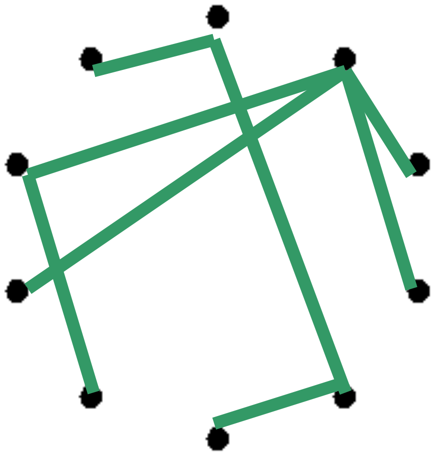
G(N,L) ensemble

Graphs with exactly
N nodes and
L links

G(N,p) ensemble
Graphs with N nodes
Each pair of nodes linked
with probability p



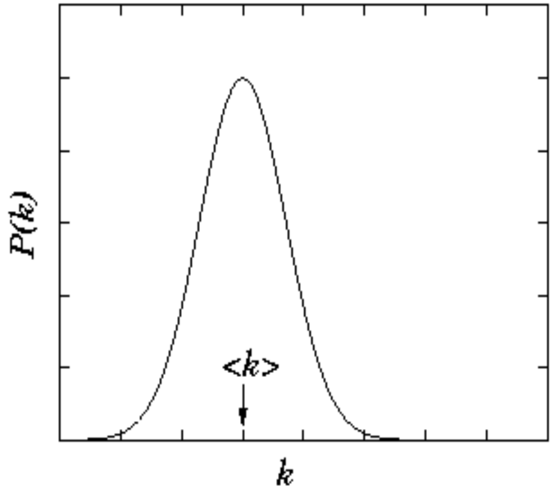
Poisson distribution



Sparse regime

$$\frac{2L}{N} \rightarrow \langle k \rangle$$

$$p = \frac{2\bar{L}}{N(N-1)} \rightarrow \frac{\langle k \rangle}{N-1}$$



Constraints of random graphs

Microcanonical ensemble

We can fix the total number of links L

$$\boxed{\sum_{i<j} a_{ij} = L}$$

Canonical ensemble

We can fix the expected total number of links $\bar{L}$

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

Canonical ensemble
The G(N,p) ensemble

According to the general theory of exponential random graph if we constraint the expected total number of links the ensemble is specified by the probability

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

where

$$P = 1, \qquad F_1(G) = \sum_{i<j} a_{ij}, \qquad C_1 = \bar{L}.$$

Each graph G is specified by its adjacency matrix so we have

$$P(\mathbf{a}) = P(G)$$

can alternatively write

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

The G(N,p) ensemble

The probability of a network in the G(N,p) ensemble can be written as

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

where

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

are the marginal probability of having a link.

Since these marginal are equal for every link, i.e. $p_{ij} = p = \dfrac{e^{-\lambda}}{1+e^{-\lambda}} = \dfrac{2\bar{L}}{N(N-1)} \quad \forall i,j$

we have

$$\boxed{P(\mathbf{a}) = p^{L}(1-p)^{N(N-1)/2-L}}$$

Proof

Let us start by calculating the partition function

$$Z = \sum_{\mathbf{a}} e^{-\lambda \sum_{i<j} a_{ij}} = \sum_{a_{12}=0,1} \sum_{a_{1,3}=0,1} \cdots \sum_{a_{N-1,N}=0,1} \prod_{i<j} e^{-\lambda a_{ij}}$$

$$= \sum_{a_{12}=0,1} e^{-\lambda a_{12}} \sum_{a_{1,3}=0,1} e^{-\lambda a_{13}} \ldots \sum_{a_{N-1,N}=0,1} e^{-\lambda a_{N-1,N}} = \prod_{i<j} \sum_{a_{ij}=0,1} e^{-\lambda a_{ij}}$$

$$= \prod_{i<j} \left(1 + e^{-\lambda}\right) = \left(1 + e^{-\lambda}\right)^{N(N-1)/2}$$

i.e.

$$\boxed{Z = \left(1 + e^{-\lambda}\right)^{N(N-1)/2}}$$

Proof

The marginal probability of a link between node (i,j) is given by

$$p_{ij} = \frac{1}{Z}\sum_{\mathbf{a}} a_{ij} e^{-\lambda \sum_{r<s} a_{rs}} = \frac{1}{Z}\sum_{a_{12}=0,1}\sum_{a_{1,3}=0,1}\cdots\sum_{a_{N-1,N}=0,1}\prod_{r<s} a_{ij} e^{-\lambda a_{rs}}$$

$$= \frac{1}{Z}\sum_{a_{12}=0,1} e^{-\lambda a_{12}} \sum_{a_{1,3}=0,1} e^{-\lambda a_{13}} \ldots \sum_{a_{ij}=0,1} a_{ij} e^{-\lambda a_{ij}} \ldots \sum_{a_{N-1,N}=0,1} e^{-\lambda a_{N-1,N}}$$

$$= \frac{1}{Z} e^{-\lambda} \prod_{r<s|(r,s)\neq(i,j))} \sum_{a_{rs}=0,1} e^{-\lambda a_{rs}} = \frac{e^{-\lambda}}{Z}\left(1+e^{-\lambda}\right)^{N(N-1)/2-1} \underset{\text{using}}{=} \frac{e^{-\lambda}}{1+e^{-\lambda}}$$

$$Z = \left(1+e^{-\lambda}\right)^{N(N-1)/2}$$

i.e.

$$\boxed{p_{ij} = p = \frac{e^{-\lambda}}{1+e^{-\lambda}} \quad \forall i,j}$$

Proof (continuation)

Therefore the expected number of links is given by

$$\bar{L} = \sum_{\mathbf{a}} P(\mathbf{a}) \left(\sum_{i<j} a_{ij} \right) = \sum_{i<j} \left(\sum_{\mathbf{a}} P(\mathbf{a}) a_{ij} \right) = \sum_{i<j} p_{ij} = p \frac{N(N-1)}{2}$$

and the marginal probability can be expressed as

$$\boxed{p = \frac{2\bar{L}}{N(N-1)}}$$

Proof (continuation)

`Given the distribution of the G(N,p) ensemble

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

with partition function and marginal given by

$$Z = \left(1 + e^{-\lambda}\right)^{N(N-1)/2} \qquad p = \frac{e^{-\lambda}}{1 + e^{-\lambda}}$$

We can easily show that the distribution factorises over contributions coming from single links, getting

$$\boxed{P(\mathbf{a}) = \prod_{i<j} p^{a_{ij}} (1-p)^{1-a_{ij}} = p^L (1-p)^{N(N-1)/2 - L}}$$

$$\mathscr{L}(p \,|\, \mathbf{a}) = -\sum_{ij} \left[a_{ij} \ln(p) - (1 - a_{ij}) \ln(1-p) \right]$$

Maximum likelihood estimation of p from data

Let us assume that a given network model is described by the G(N,p) ensemble,

$$P_p(\mathbf{a}) = \prod_{i<j} p^{a_{ij}}(1-p)^{1-a_{ij}} = p^L(1-p)^{N(N-1)/2-L}$$

The log-likelihood of p is given by

$$\mathscr{L}(p\,|\,\mathbf{a}) = -\sum_{ij} a_{ij}\ln(p) - (1-a_{ij})\ln(1-p) = -L\ln p - (N(N-1)/2 - L)\ln(1-p)$$

where L is the exact observed number of links in the data.

By maximising the log-likelihood we get

$$\frac{\partial \mathscr{L}(p\,|\,\mathbf{a})}{\partial p} = -\frac{L}{p} + (N(N-1)/2 - L)\frac{1}{1-p} = 0$$

$$(N(N-1)/2 - L)p - L(1-p) = N(N-1)/2p - L = 0$$

Therefore the maximum likelihood estimation of p is $\boxed{p = \frac{2L}{N(N-1)}}$

Sparse regime

In the sparse regime

the total number of links satisfies

$$\boxed{2\bar{L} = \langle k \rangle N}$$

where the average degree is constant .

Therefore the marginal probability of a link

can be written as

$$\boxed{p = \frac{2\bar{L}}{N(N-1)} = \frac{\langle k \rangle}{N-1} \simeq \frac{\langle k \rangle}{N}}$$

Degree distribution of the G(N,p) ensemble

The degree distribution of the G(N,p) ensemble is given by the binomial distribution

$$\boxed{\mathbb{P}(k_i = k) = \mathbb{P}\left(\sum_{j=1}^{N} a_{ij} = k\right) = \binom{N-1}{k}\left(\frac{\langle k\rangle}{N}\right)^k\left(1-\frac{\langle k\rangle}{N}\right)^{N-1-k}}$$

That in the large network limit converges to the Poisson distribution

$$\boxed{\mathbb{P}(k_i = k) = \frac{1}{k!}\langle k\rangle^k e^{-\langle k\rangle}}$$

Entropy of the G(N,p) ensemble

The entropy of the G(N,p) ensemble

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

defined by the distribution

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

is given by

$$S = -\frac{N(N-1)}{2} \left[p \ln p + (1-p) \ln (1-p) \right]$$

By inserting the explicit expression of the marginal we get

$$\boxed{S = -\frac{N(N-1)}{2} \left[\frac{\langle k \rangle}{N} \ln \left(\frac{\langle k \rangle}{N} \right) + \left(1 - \frac{\langle k \rangle}{N} \right) \ln \left(1 - \frac{\langle k \rangle}{N} \right) \right]}$$

Scaling of the entropy of the random graph G(N,p)

The entropy of the ensemble G(N,p)

obeys the following scaling with the total number of nodes N

$$S = \underbrace{\frac{1}{2}\langle k\rangle N \ln\langle k\rangle N}_{\mathcal{O}(N \ln N)} - \underbrace{N\langle k\rangle \ln\langle k\rangle + \frac{N}{2}\langle k\rangle}_{\mathcal{O}(N)} + \mathcal{O}(1)$$

The entropy is not extensive

Proof

- Starting from the expression of the entropy

$$S = -\frac{N(N-1)}{2}\left[\frac{\langle k\rangle}{N}\ln\left(\frac{\langle k\rangle}{N}\right) + \left(1-\frac{\langle k\rangle}{N}\right)\ln\left(1-\frac{\langle k\rangle}{N}\right)\right]$$

- By expanding in the limit for $N \to \infty$

$$S = \frac{N}{2}\langle k\rangle \ln N - \frac{N}{2}\langle k\rangle \ln[\langle k\rangle] + \frac{N}{2}\langle k\rangle + \mathcal{O}(1)$$

- By rearranging the terms we get

$$\boxed{S = \frac{1}{2}\langle k\rangle N \ln\langle k\rangle N - N\langle k\rangle \ln\langle k\rangle + \frac{N}{2}\langle k\rangle + \mathcal{O}(1)}$$

Microcanonical random graph ensemble G(N,L)

The microcanonical ensemble G(N,L) where we enforce the hard constraints on the total number of links is determined by the distribution

$$\boxed{P(G) = \frac{1}{Z_M}\delta\left(L, \sum_{i<j} a_{ij}\right)}$$

where

$$\boxed{Z_M = \sum_{G\in\Omega_G} \delta\left(L, \sum_{i<j} a_{ij}\right) = \binom{N(N-1)/2}{L}}$$

indicates the total number of simple networks of N nodes and L links

Entropy of the G(N,L) ensemble

The entropy of the G(N,L) ensemble

is given by

$$\boxed{\Sigma = \ln Z_M = \ln \left[\binom{N(N-1)/2}{L} \right]}$$

Equivalence of the random graph ensembles

The random graph ensembles G(N,p) and G(N,L) are asymptotically equivalent.

Indeed for $N \gg 1$ their entropies satisfy

$$\boxed{\Sigma \simeq S}$$

(left as an exercise)

Average clustering coefficient

The average clustering coefficient

of the nodes of a Poisson network is given by

$$\boxed{\langle C_i | k_i \rangle = \frac{\langle k \rangle}{N}}$$

Therefore it is vanishing in the large network limit

Diameter of random graphs

The diameter of the G(N,p) ensemble scales like

$$\boxed{D = \mathcal{O}(\ln N)}$$

therefore we say that random graphs have infinite Hausdorff dimension

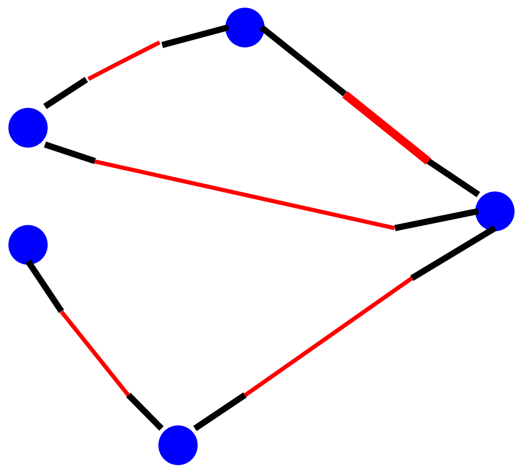
$$\boxed{d_H = \infty}$$

Degree sequence as constraint

Network ensemble with given degree sequence

Microcanonical ensemble

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

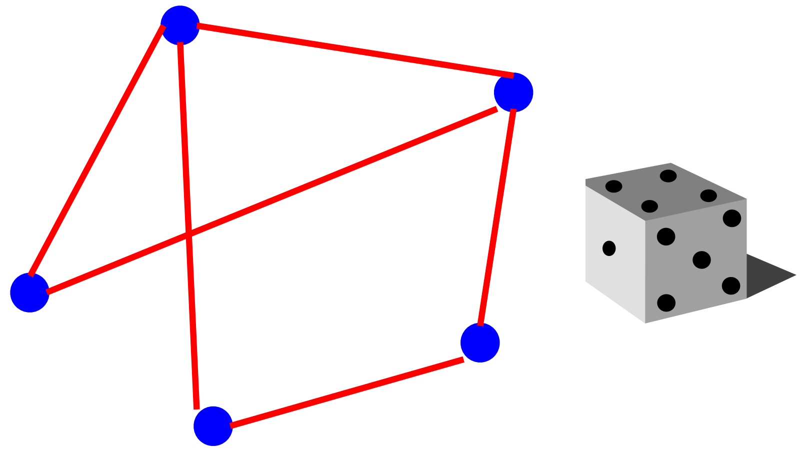


Ensemble of network with exact degree sequence

Configuration model

Canonical ensemble

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



Ensemble of networks given expected degree sequence

Exponential random graph

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

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

Canonical ensemble

Proposition

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

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

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

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

Hamiltonian and Partition Function

- The Hamiltonian of the canonical ensemble with given expected degree sequence is given by

$$\boxed{H(G) = \sum_{i=1}^{N} \lambda_i \left(\sum_{j=1}^{N} a_{ij} \right) = \sum_{i<j} (\lambda_i + \lambda_j) a_{ij}}$$

- The partition function is given by

$$\boxed{Z = \prod_{i<j} \left(1 + e^{-(\lambda_i + \lambda_j)} \right)}$$

Proof

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

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

that can be equivalent expressed as

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

Indeed the Hamiltonian can be written as

$$H(G) = \sum_{i=1}^{N} \lambda_i \sum_{j=1}^{N} a_{ij} = \frac{1}{2} \left[\sum_{i,j=1,\ldots,N} a_{ij}\lambda_i + \sum_{i,j=1,\ldots,N} a_{ji}\lambda_j \right] = \frac{1}{2} \sum_{i,j=1,N} a_{ij}(\lambda_i + \lambda_j) = \sum_{i<j} a_{ij}(\lambda_i + \lambda_j)$$

Proof (continuation)

Given the expression for the Gibbs measure

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

The partition function of the ensemble can be written as

$$\boxed{Z = \sum_{\mathbf{a}} e^{-\sum_{i<j} a_{ij}(\lambda_i+\lambda_j)} = \prod_{i<j} \sum_{a_{ij}=0,1} e^{-a_{ij}(\lambda_i+\lambda_j)} = \prod_{i<j} \left(1 + e^{-(\lambda_i+\lambda_j)}\right)}$$

Marginal and equation for the Lagrangian multipliers

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

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

is given by

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

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

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

Proof

The partition function of the ensemble can be written as

$$Z = \sum_{\mathbf{a}} e^{-\sum_{r<s} a_{rs}(\lambda_r+\lambda_s)} = \prod_{r<s} \sum_{a_{rs}=0,1} e^{-a_{rs}(\lambda_r+\lambda_s)} = \prod_{r<s} \left(1 + e^{-(\lambda_r+\lambda_s)}\right)$$

The marginal probability of the link (i,j) can be calculated as

$$p_{ij} = \frac{1}{Z} \sum_{\mathbf{a}} a_{ij} e^{-\sum_{r<s} a_{rs}(\lambda_r+\lambda_s)} = \frac{1}{Z} e^{-\lambda_i-\lambda_j} \prod_{r<s|(r,s)\neq(i,j))} \sum_{a_{rs}=0,1} e^{-a_{rs}(\lambda_r+\lambda_s)}$$

$$p_{ij} = \frac{e^{-(\lambda_i+\lambda_j)}}{Z} \prod_{r<s|(r,s)\neq(i,j))} \left(1 + e^{-(\lambda_r+\lambda_s)}\right) = \frac{e^{-\lambda_i-\lambda_j}}{1 + e^{-\lambda_i-\lambda_j}}$$

Proof (continuation)

The Lagrangian multipliers are fixed by the constraints,

$$\bar{k}_i = \sum_{\mathbf{a}} P(\mathbf{a}) \left(\sum_{j=1}^{N} a_{ij} \right) = \sum_{j=1}^{N} \left(\sum_{\mathbf{a}} P(\mathbf{a}) a_{ij} \right) = \sum_{j=1}^{N} p_{ij}$$

i.e.

$$\bar{k}_i = \sum_{j=1}^{N} p_{ij}$$

Therefore by substituting the expression of marginal in terms of the Lagrangian multipliers we obtain

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

The probability of a graphs in terms of the marginals

When the canonical network model involves only constraints linear on the adjacency matrix like the expected degree sequence than the probability of a network can be written as

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

In the case in which we have $P(\mathbf{a}) = \frac{1}{Z} e^{-\sum_{i=1}^{N} \lambda_i \sum_{j=1}^{N} a_{ij}}$

This expression follows directly from the equations

$$p_{ij} = \frac{e^{-\lambda_i - \lambda_j}}{1 + e^{-\lambda_i - \lambda_j}} \qquad Z = \prod_{i<j} \left(1 + e^{-\lambda_i - \lambda_j}\right)$$

Entropy of canonical ensemble

The entropy

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

of the canonical ensemble with Gibbs measure

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

can be expressed as

$$\boxed{S = \sum_{i=1}^{N} \lambda_i \bar{k}_i + \ln Z = \sum_{i=1}^{N} \lambda_i \bar{k}_i + \sum_{i<j} (1 + e^{-\lambda_i - \lambda_j})}$$

Entropy of canonical ensemble

Alternatively the entropy

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

of the canonical ensemble can be expressed as

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

where we have used

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

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

Log-likelihood

Given a network G our aim to to model it with a canonical network model

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

depending on the parameters $\overrightarrow{\lambda}$

The log-likelihood of the parameters is given by

$$\boxed{\mathscr{L}(\overrightarrow{\lambda}\,|\,\mathbf{a}) = -\ln P_{\overrightarrow{\lambda}}(\mathbf{a}) = \sum_{i=1}^{N} \lambda_i k_i + \sum_{i<j} \ln(1 + e^{-\lambda_i - \lambda_j})}$$

Log-likelihood

Given the alternative expression of the probability of a network

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

where the marginal probabilities are given by

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

The log-likelihood can be also expressed as

$$\boxed{\mathscr{L}(\overrightarrow{\lambda} \,|\, \mathbf{a}) = -\sum_{i<j} \left[a_{ij} \ln p_{ij} + (1 - a_{ij}) \ln(1 - p_{ij}) \right]}$$

Maximum-likelihood Estimation of the parameters

Given the log-likelihood of the parameters $\overrightarrow{\lambda}$

$$\mathcal{L}(\overrightarrow{\lambda}\,|\,G) = \sum_{i=1}^{N} \lambda_i k_i + \sum_{i<j} \ln(1 + e^{-\lambda_i - \lambda_j})$$

The maximum likelihood estimation of the parameters $\overrightarrow{\lambda^\star}$ **gives**

$$0 = \frac{\partial \mathcal{L}(\overrightarrow{\lambda}\,|\,G)}{\partial \lambda_i} = k_i + \frac{\partial}{\partial \lambda_i} \sum_{r<s} \ln(1 + e^{-\lambda_r - \lambda_s}) = k_i - \sum_{j=1}^{N} \frac{e^{-\lambda_i - \lambda_j}}{1 + e^{-\lambda_i - \lambda_j}}$$

or equivalently

$$\boxed{k_i = \langle k_i \rangle_{ENSEMBLE} = \sum_{j=1}^{N} \frac{e^{-\lambda_i - \lambda_j}}{1 + e^{-\lambda_i - \lambda_j}}}$$

Algorithm to generate networks in the canonical ensemble

- Given the sequence of expected degrees calculate the Lagrangian multipliers solving the equations

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

- For every pair of nodes *(i,j)* draw a link, i.e. put

$$\boxed{a_{ij} = \begin{cases} 1 \ \textbf{with probability} \ p_{ij} \\ 0 \ \textbf{with probability} \ 1-p_{ij} \end{cases}}$$

with

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

Metropolis-Hastings algorithm

- Start from a given network of N nodes. Calculate the Lagrangian multipliers as in the previous algorithm.

- Iterate the following procedure until convergence of observables

1. Pick randomly a pair of nodes (i, j)

2. Perform the transition $\mathbf{a} \to \mathbf{a}'$ with probability $\boxed{\Pi_{\mathbf{a}\to\mathbf{a}'} = \textbf{min}\left[1, \frac{P(\mathbf{a}')}{P(\mathbf{a})}\right]}$

where $\mathbf{a}'$ has elements

$$\boxed{a'_{rs} = \begin{cases} a_{rs} & \textbf{if } (r,s) \neq (i,j) \textbf{ and } \quad (r,s) \neq (j,i) \\ 1 - a_{ij} & \textbf{if } (r,s) = (i,j) \textbf{ or } \quad (r,s) = (j,i) \end{cases}}$$

Final remarks

In this first second of the second lesson we have covered

A. Random graphs

B. Canonical ensembles of networks with given expected degree sequence

In the next lesson we will introduce

Degree Correlations and Natural cutoffs

We will discuss the microcanonical ensemble with given degree sequence

We will expand on non-equivalence of ensembles

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

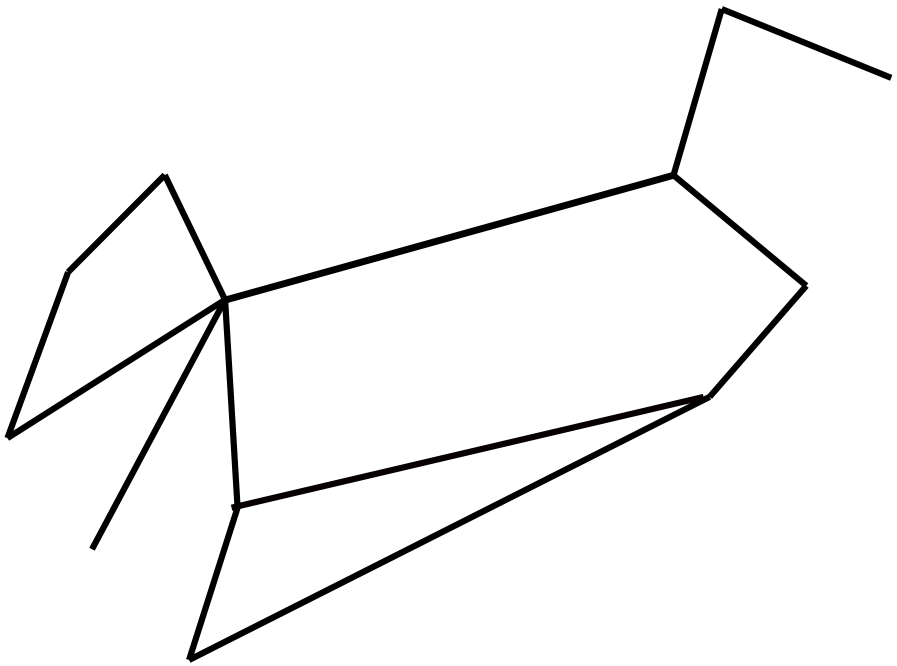
The most direct measure of the matrix $\pi_{k,k'}$ is the direct measure of the probability

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.

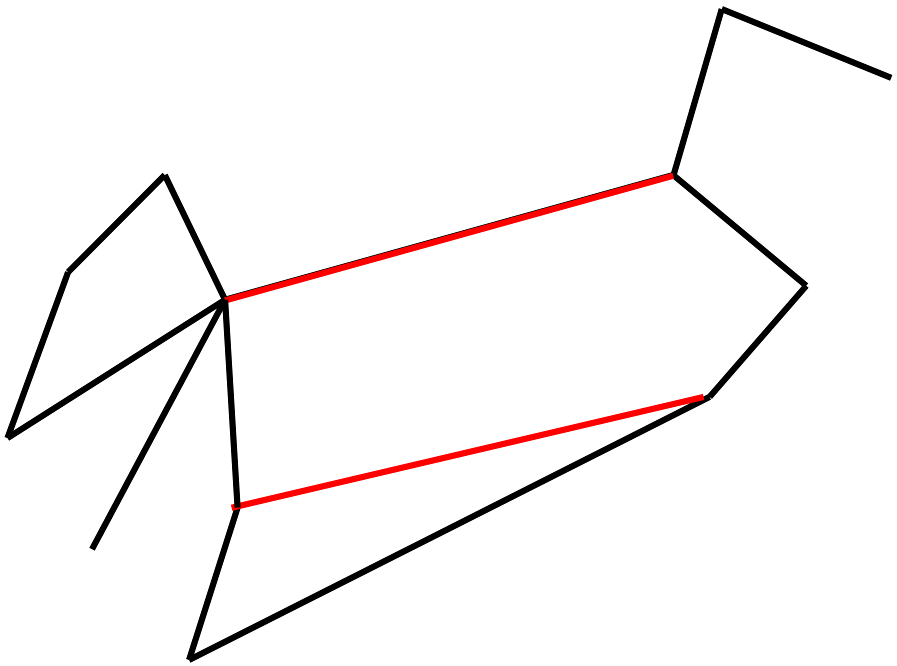
Randomization of a network swap of connections



- *Choose two random links linking four distinct nodes*

Maslov & Sneppen 2002

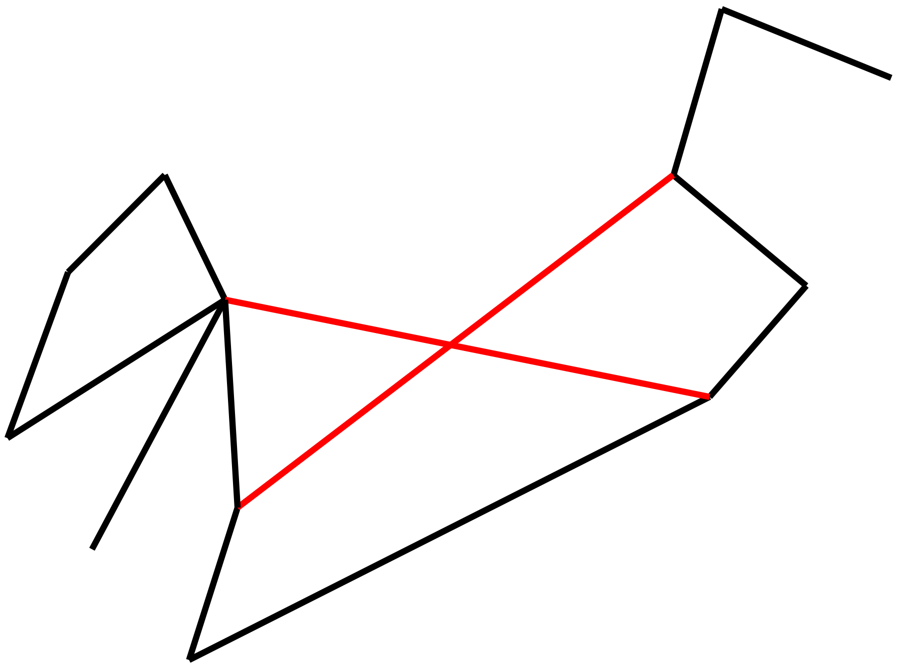
Randomization of a network swap of connections



- ➢ *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



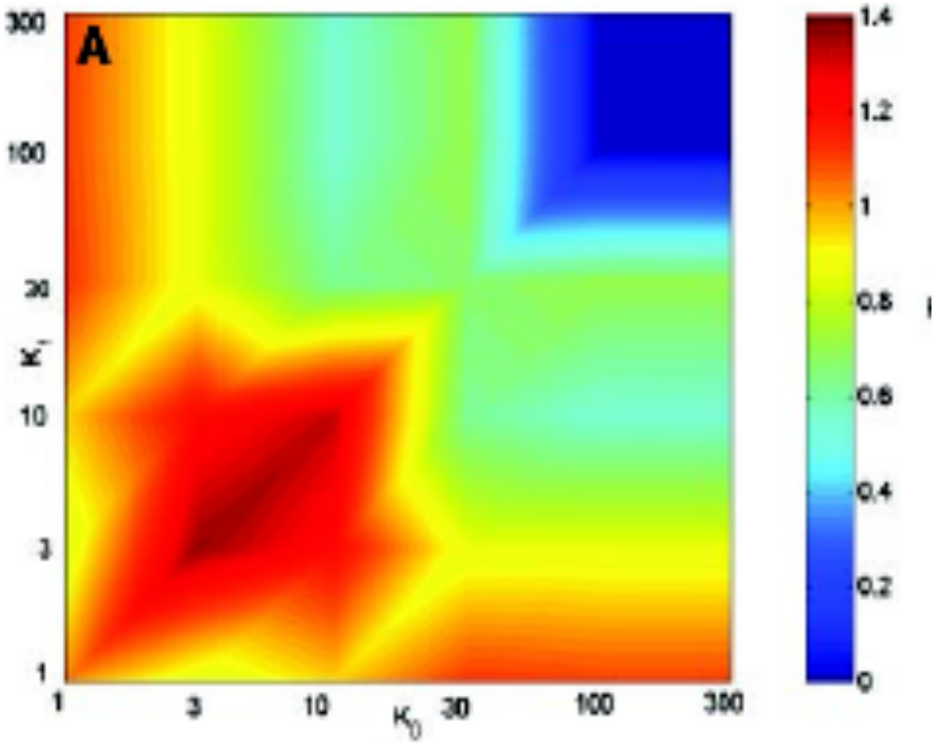
- *Choose two random links linking four distinct nodes*
- *If possible (not already existing links) swap the ends of the links*

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

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

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

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

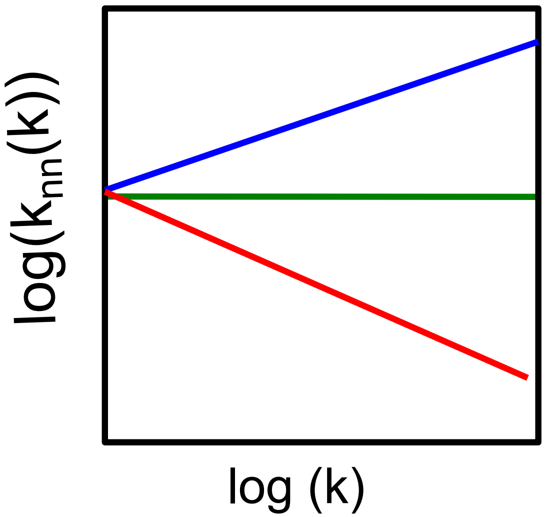
The average degree of neighbour nodes

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

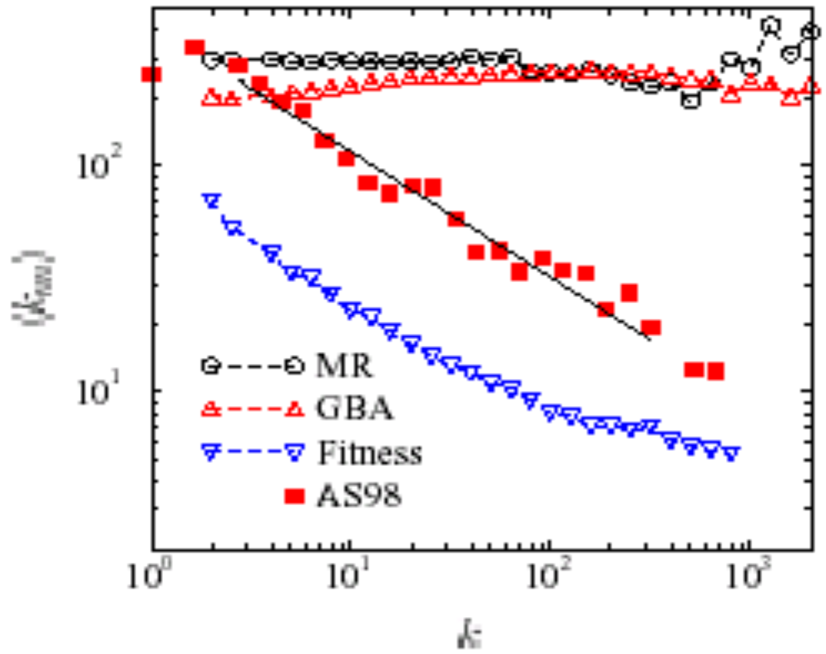
$$k_{nn}(k) = \left\langle \frac{1}{k_i} \sum_{j=1}^{N} a_{ij} k_j \right\rangle_{k_i=k}$$

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

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

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

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

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

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

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

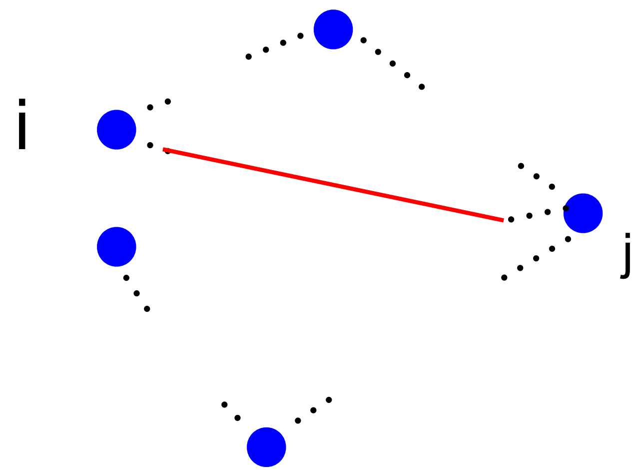
$$\pi_{i|j} = \frac{\pi_{ij}}{\sum_{j'=1}^{N} \pi_{jj'}} = \left(\frac{\bar{k}_i \bar{k}_j}{(\langle \bar{k} \rangle N)^2} \right) \left(\frac{\langle k \rangle N}{\bar{k}_j} \right) = \frac{\bar{k}_i}{\langle \bar{k} \rangle N} = q_i$$

Example

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

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

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

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

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

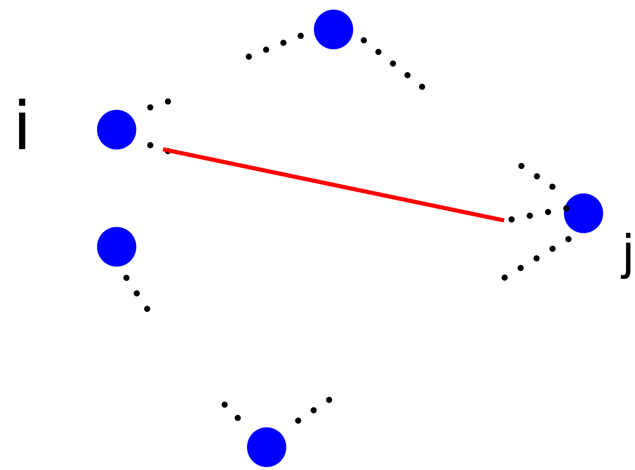
$$p_{ij} = \langle a_{ij} \rangle = n_{ij} = \frac{\bar{k}_i \bar{k}_j}{(\langle \bar{k} \rangle N)}$$

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

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

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

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

Proof

In uncorrelated network the probability that two nodes are connected is

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

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

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

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

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

Natural and structural cutoff of scale-free networks

For scale-free networks with degree distribution

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

the

natural cutoff is larger than the structural cutoff

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

for

$$\boxed{\gamma \leq 3}$$

Uncorrelated scale-free networks

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

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

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

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

Canonical ensemble

Proposition

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

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

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)

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

is given by

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

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

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

Natural correlations

Since the marginal probabilities

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

do not factorise in terms depending exclusively on single nodes,

the configuration model leads to

natural correlations

which are

disassortative

Evidence of disassortative correlations

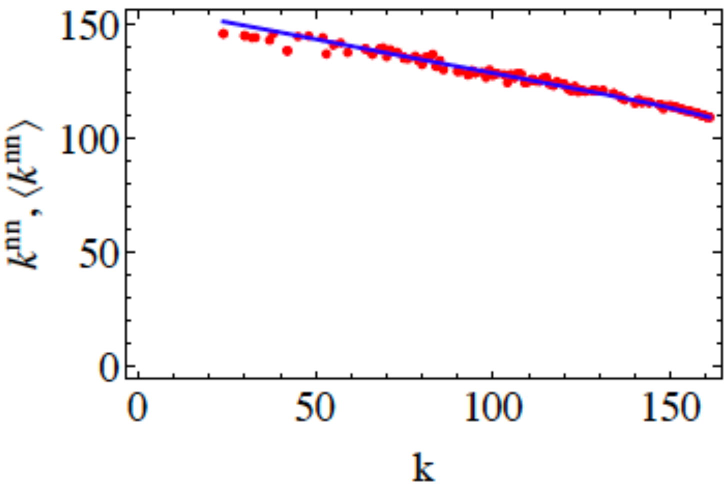
Average degree of the neighbour of a node in the data

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

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

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

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

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

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

We get that the initial hypothesis is only satisfied for

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

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

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

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

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

Only dependent on the average degree

Dependent on the degree distribution

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 $\quad x = \dfrac{\bar{k}_i\bar{k}_j}{\langle\bar{k}\rangle N}$

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

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

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

Only dependent on the average degree

Dependent on the degree distribution

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

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