Maximum Entropy Network Ensembles

LTCC Course
Lesson 5

Ginestra Bianconi

School of Mathematical Sciences Queen Mary University of London

Queen Mary
University of London

Lesson 5

- Networks with given expected degree correlations
- 2 star model and Strauss model
- Models with hidden variables
- Inference of block models

References

Books

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

Articles

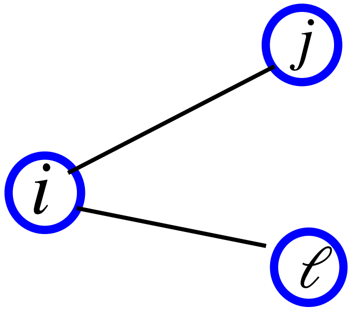
- Park, J. and Newman, M.E., 2004. Statistical mechanics of networks. *Physical Review E*, *70*(6), p.066117.
- Bianconi, G., 2007. The entropy of randomized network ensembles. *EPL (Europhysics Letters)*, *81*(2), p.28005.
- Bianconi, G., 2009. Entropy of network ensembles. Physical Review E, 79(3), p.036114.
- Peixoto, T.P., 2017. Bayesian stochastic blockmodeling. *arXiv preprint arXiv:1705.10225*.
- Caldarelli, G., Capocci, A., De Los Rios, P. and Munoz, M.A., 2002. Scale-free networks from varying vertex intrinsic fitness. *Physical review letters*, *89*(25), p.258702.
- Krioukov D, Papadopoulos F, Kitsak M, Vahdat A, Boguná M. Hyperbolic geometry of complex networks. Physical Review E. 2010 Sep 9;82(3):036106.

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

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

and the expected total number of wedges as a soft constraint

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

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

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

with Hamiltonian given by

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

where λ and γ are Lagrangian multipliers enforcing the constraints

Mean-field approximation

In the mean field approximation we neglect correlations and we put

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

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.

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

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

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

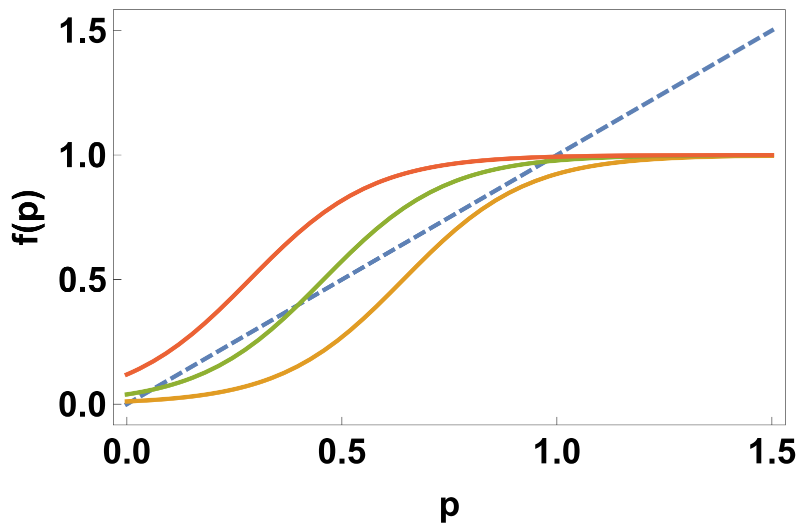
We can calculate the marginal which leads to

the self-consistent equation for p given by

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

Phase transition in the 2-star model

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



For some values of the Lagrangian multipliers

there are two stable solutions at

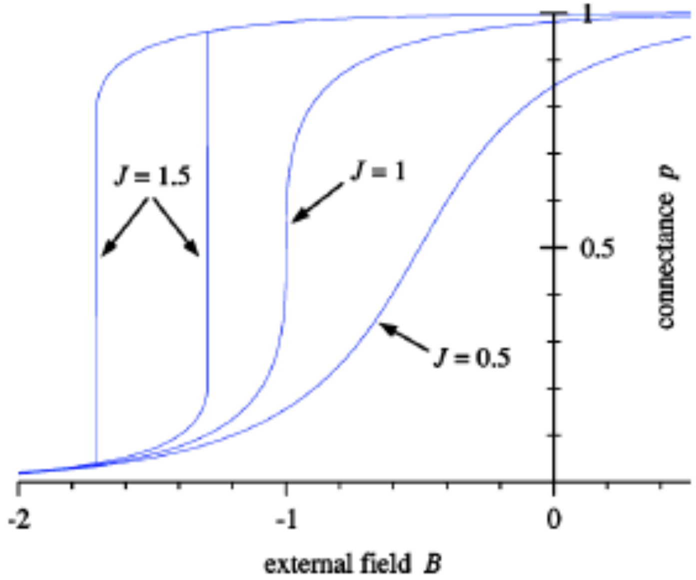
high density (high value of p) and low density (low values of p)

Phase transition in the 2-star model

By putting

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

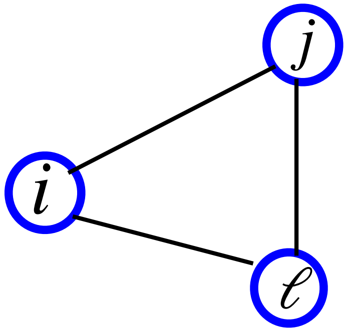
the phase diagram of p as a function of B is given by



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

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

and the expected total number of triangles as a soft constraint

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

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.

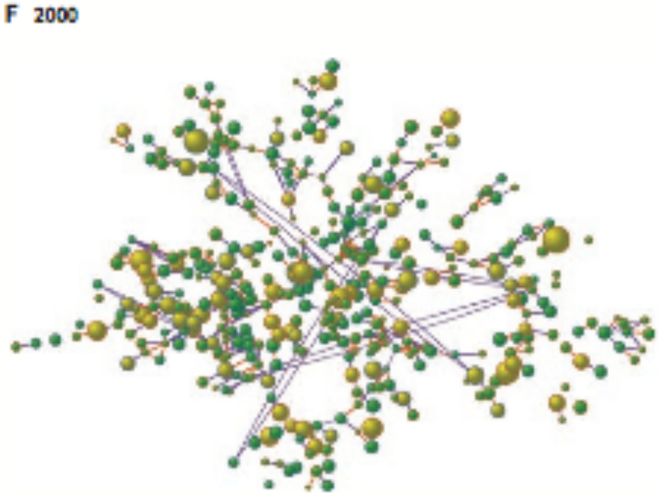
Feature of the nodes
And
Hidden variables

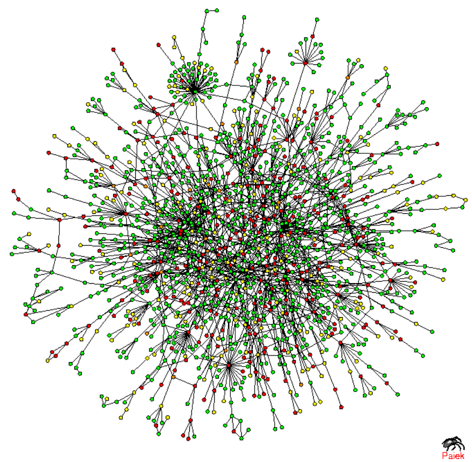
Often network data includes MetaData

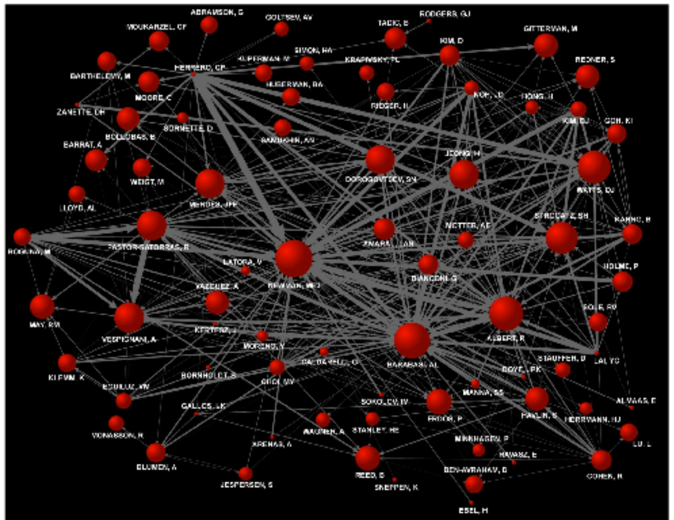
SOCIAL NETWORKS
Metadata about agents

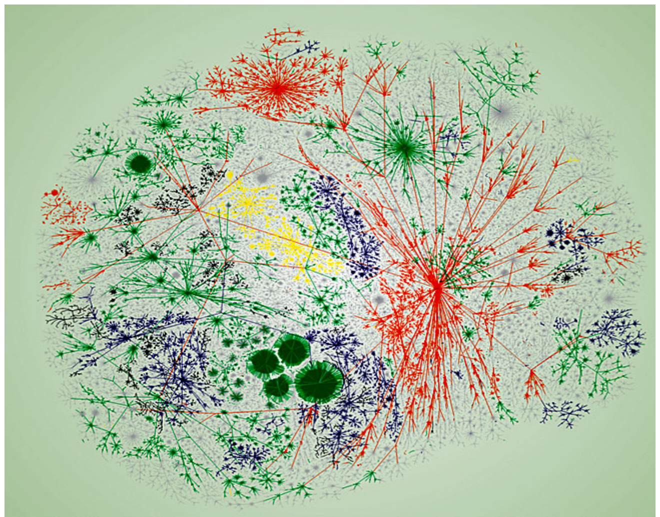
COMMUNICATION NETWORKS
Locations

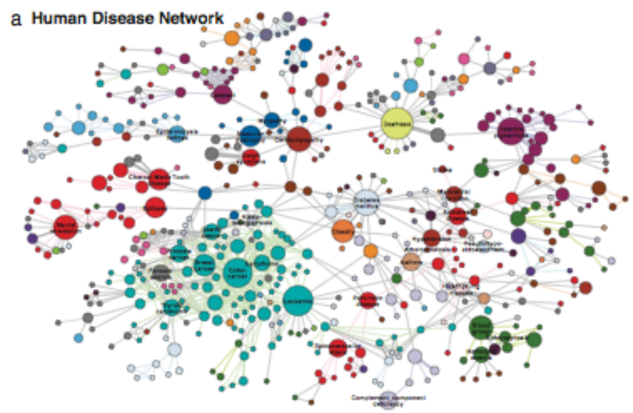
BIOLOGICAL NETWORKS
Metadata about nodes





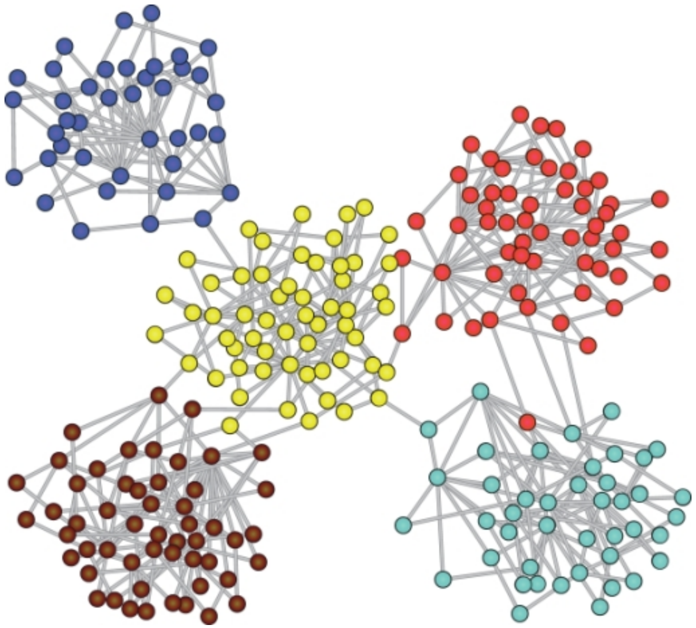




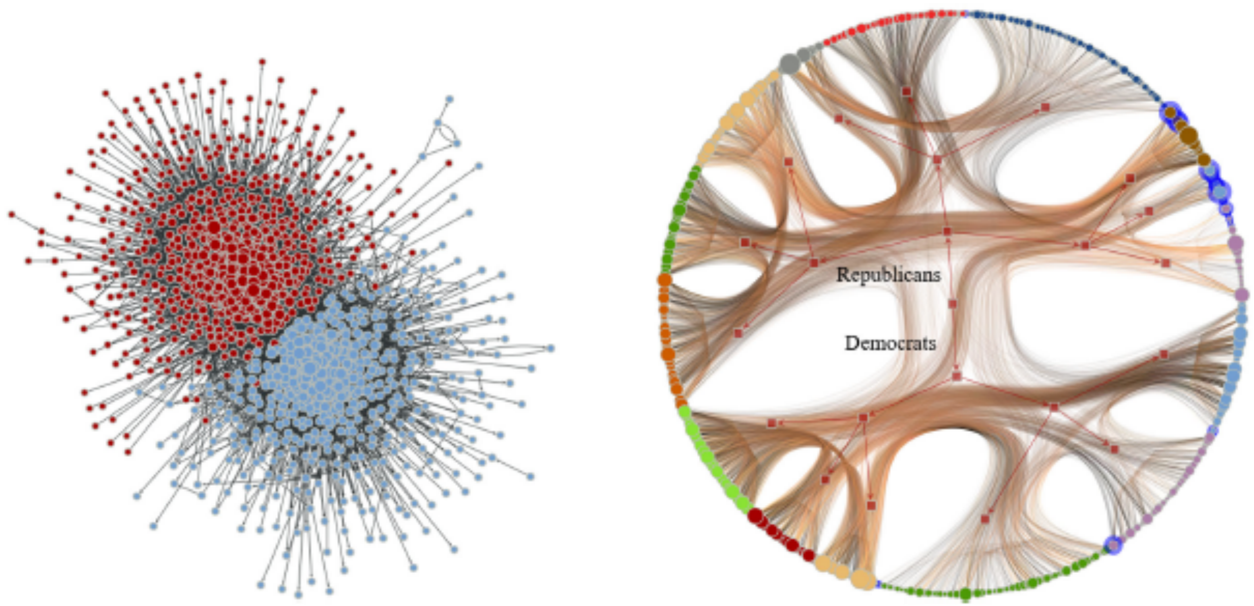


Some time we aim at inferring metadata

COMMUNITY DETECTION



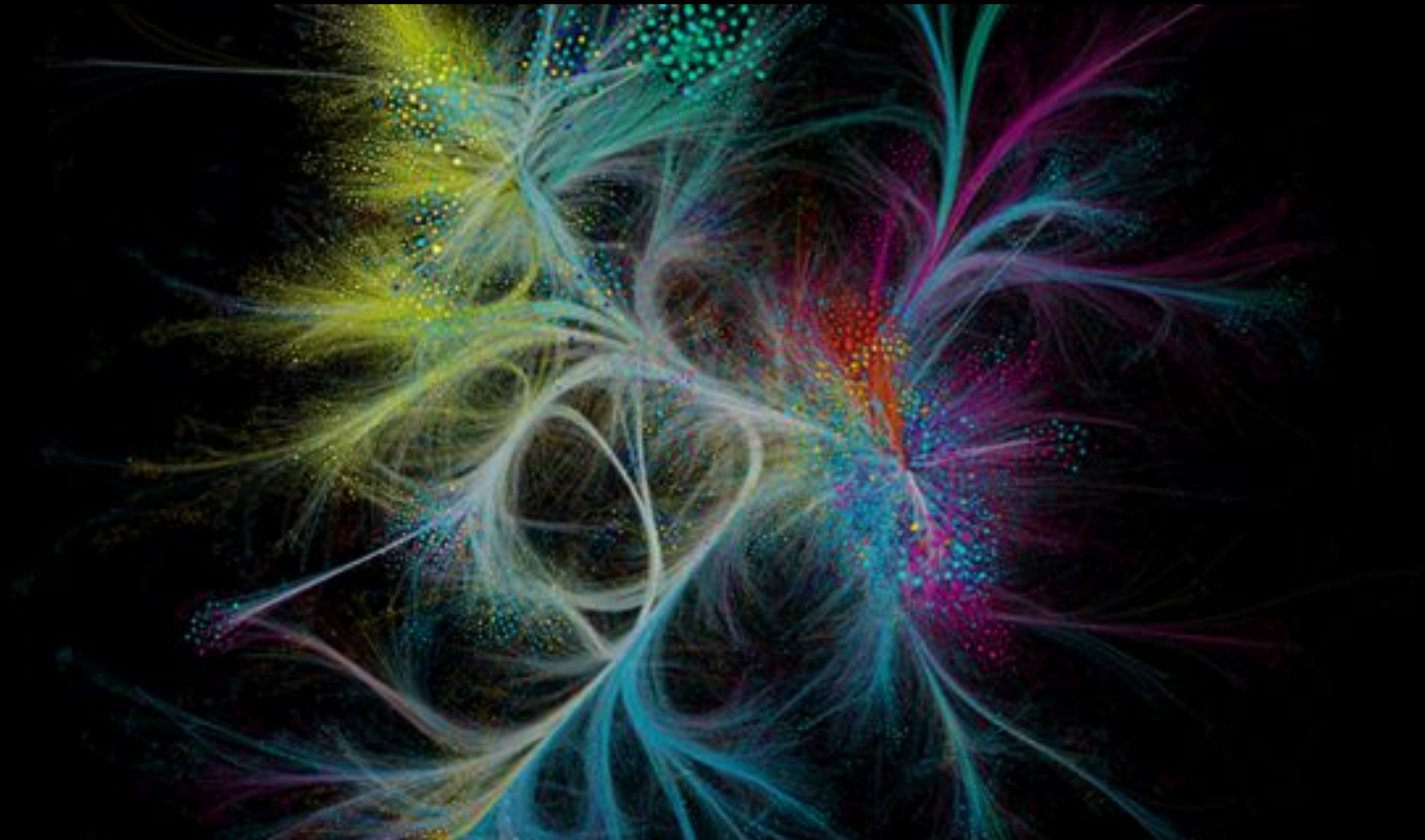
COMMUNITY DETECTION



T. Peixoto PRX (2014)

Network visualisation

INFERRING NETWORK EMBEDDINGS
THE 150 YEARS NATURE COVER



Hidden variables
The direct problem

- In order to have a well defined model for inferring hidden variables we need to approach and solve the direct problem.
- The direct problem is the problem of modelling networks under the assumption that the hidden variable are known
- Here we will describe this approach within the theory of the canonical network ensembles.

Local constraints dependent on structural properties and hidden variables

Spatial networks

Consider a simple network of N nodes

Let us assume that each node i is assigned a set of position of nodes in space $\mathbf{x}_i$

This position can be

an actual geographical position of nodes in space

or a position of nodes in a generalised space

(social distance in social networks or any suitable hidden embedding space)

Spatial networks

We bin the set of all possible distances d in a vector such that

$$\mathbb{1}\left(d(\mathbf{x}_i, \mathbf{x}_j), d\right) = \begin{cases} 1 \text{ \textbf{if} } d(\mathbf{x}_i, \mathbf{x}_j) \in [d, d + \Delta d] \\ 0 \text{ \textbf{otherwise}} \end{cases}$$

We consider the set of two different types of soft constraints enforcing respectively the degree of each nodes and the expected number of links at distance n

Expected degree of each node

$$F_i(G) = \sum_{j=1}^{N} a_{ij}$$

$$\bar{k}_i = \sum_{G \in \Omega_G} \left[P(G \mid \mathbf{x}) \left(\sum_{j=1}^{N} a_{ij} \right) \right]$$

Expected total number of links at distance d

$$F_\mu(G) = \sum_{i<j}^{N} a_{ij} \, \mathbb{1}(d(\mathbf{x}_i, \mathbf{x}_j), d)$$

$$\bar{L}(d) = \sum_{G \in \Omega_G} \left[P(G \mid \mathbf{x}) \left(\sum_{i<j}^{N} a_{ij} \, \mathbb{1}(d(\mathbf{x}_i, \mathbf{x}_j), d) \right) \right]$$

Probability of a spatial network

According to the general theory of canonical network ensembles (lesson 2)

the probability of a spatial network

in which we enforce

the expected degree sequence

and the expected number of links at distance d

reads

$$P(G\,|\,\mathbf{x}) = \frac{1}{Z}\exp\left[-\sum_{i=1}^{N}\lambda_i\sum_{j=1}^{N}a_{ij} - \sum_{d}\omega_d\sum_{i<j}a_{ij}\mathbb{1}\,\delta(d(\mathbf{x}_i,\mathbf{x}_j),d)\right]$$

The marginal probability of the spatial network ensemble

The marginal probability a link is given by

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

Where λ_i **and** ω_d are the Lagrangian multipliers enforcing the constraints

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

$$\bar{L}(d) = \sum_{i<j} \frac{e^{-\lambda_i - \lambda_j - \omega_d}}{1 + e^{-\lambda_i - \lambda_j - \omega_d}} \mathbb{1}(d(\mathbf{x}_i, \mathbf{x}_j), d)$$

(left as an exercise)

Marginal of the spatial network ensemble

Given the marginal

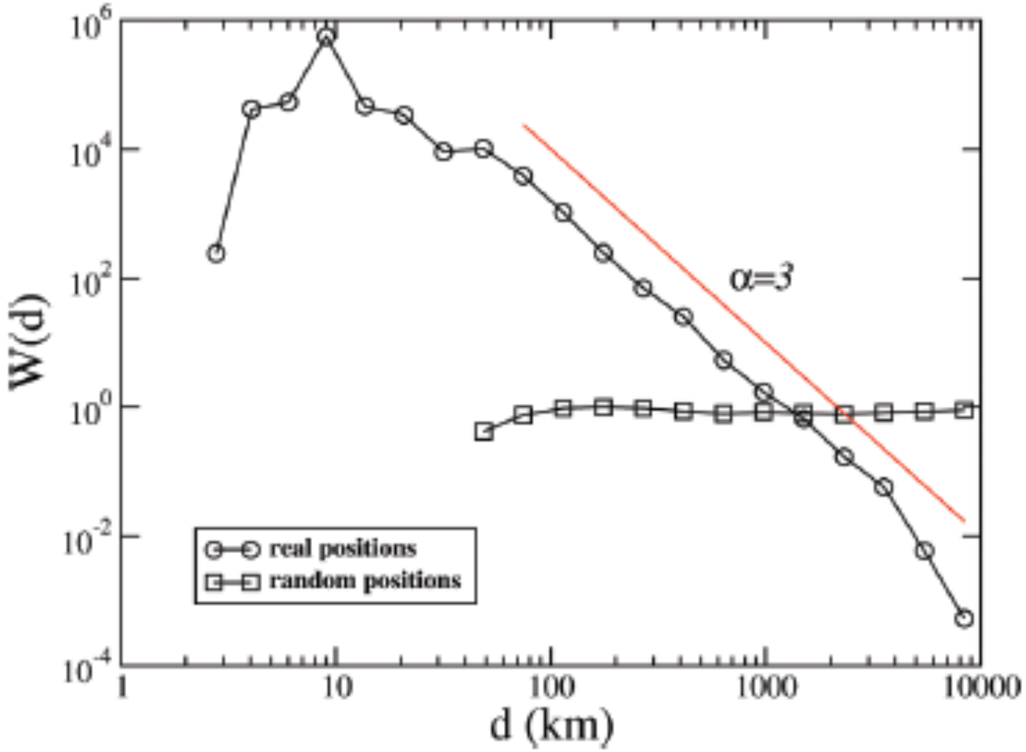
$$\boxed{p_{ij} = \frac{e^{-\lambda_i-\lambda_j-\omega_{d_{ij}}}}{1+e^{-\lambda_i-\lambda_j-\omega_{d_{ij}}}}}$$

by putting

$$W(d_{ij}) = e^{-\omega_{d_{ij}}} \qquad \theta_i = e^{-\lambda_i}$$

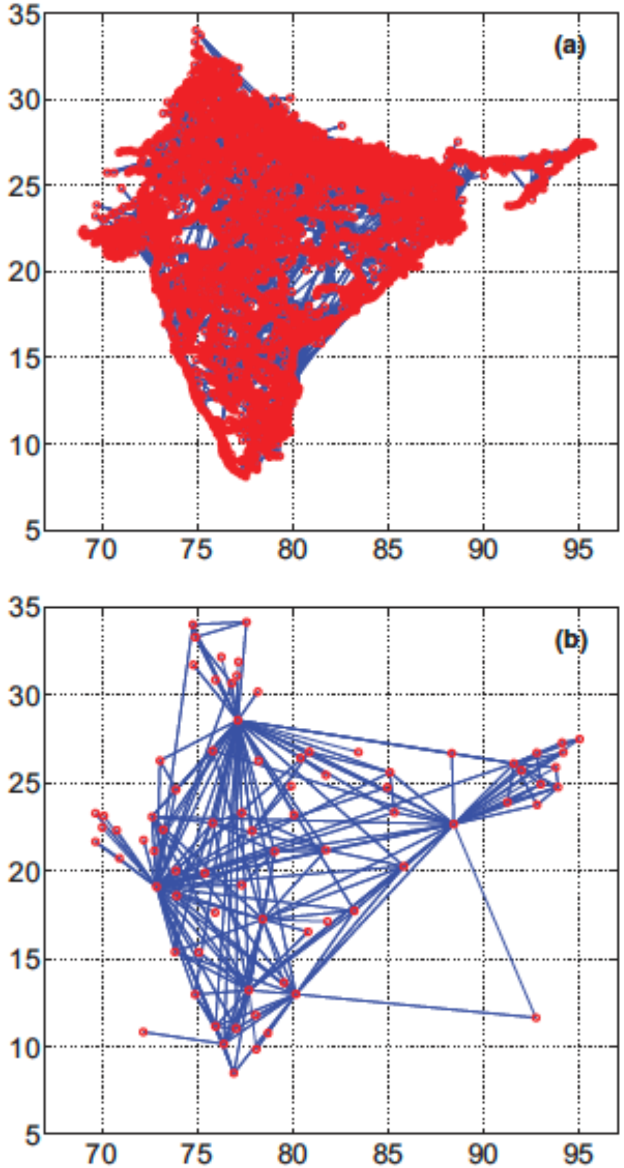
we have

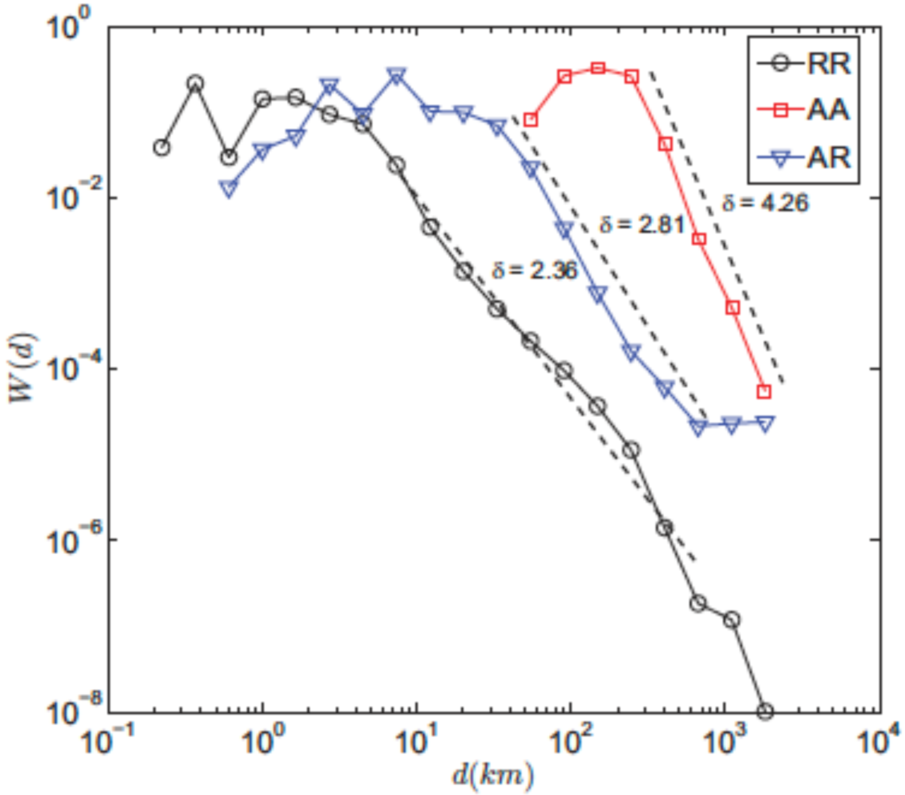
$$\boxed{p_{ij} = \frac{e^{-\lambda_i-\lambda_j-\omega_{d_{ij}}}}{1+e^{-\lambda_i-\lambda_j-\omega_{d_{ij}}}} = \frac{\theta_i\theta_j W(d_{ij})}{1+\theta_i\theta_j W(d_{ij})}}$$



10^6
10^4
10^2
10^0
10^{-2}
10^{-4}
W(d)
1
10
100
1000
10000
d (km)
α=3
real positions
random positions

Indian train and airport network





Halu et. al. PRE (2014)

Block model

Block assignment

Consider a simple network of N nodes

Let us assume that each node i is assigned a set of node classification or "block " $b_i \in \{1,2,\ldots,B\}$

This classification of nodes can come

from metadata about the nodes or

can be an hidden property of nodes reflected in the network structure

(for instance the network model might enforce the tendency of nodes of the same block to be more likely to link to each other or less likely to link to each other)

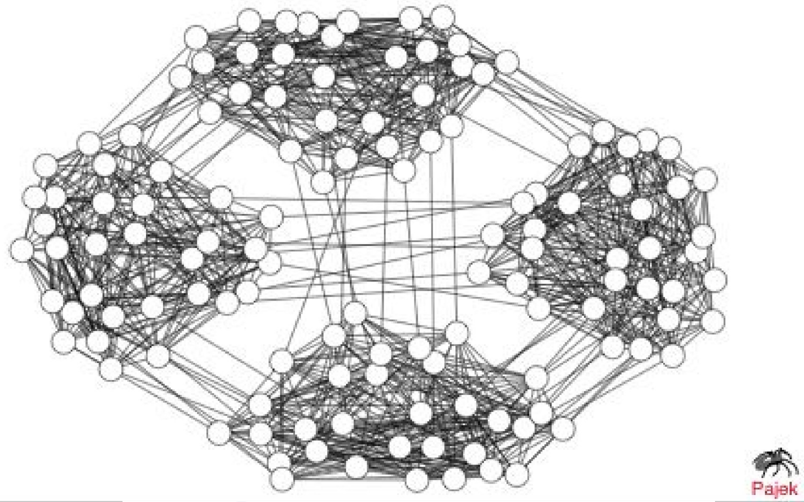
Examples

Girvan-Newman benchmark

Regular network with 4 communities

where nodes are more likely to be connected

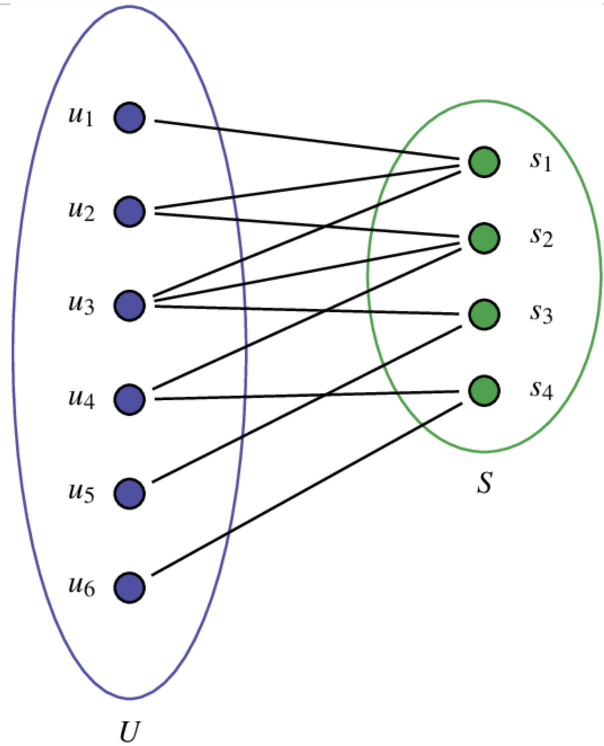
within their community than across communities



Bipartite network

Network formed by two blocks with links

allowed only between nodes of different blocks



Canonical ensemble

We consider the canonical network ensemble with soft constraints of the type

$$\bar{C}_\mu = \sum_{G \in \Omega_G} \left[P(G \mid \mathbf{b}) F_\mu(G) \right]$$

Where the soft constraints enforce a given expected total number of links and given expected total cost of the link

Expected total number of links between nodes of block b and nodes of block b'

$$F_{b,b'}(G) = \sum_{i,j}^{N} a_{ij} \delta(b_i, b) \delta(b_j, b')$$

$$\bar{e}(b, b') = \sum_{G \in \Omega_G} \left[P(G \mid \mathbf{b}) \left(\sum_{i,j}^{N} a_{ij} \delta(b_i, b) \delta(b_j, b') \right) \right] \quad \textbf{for } b \leq b'$$

Probability of a network given the block assignment

According to the general treatment of canonical network ensembles the probability of the network reads

$$\boxed{P(G\,|\,\mathbf{b}) = \frac{1}{Z}\exp\left[-\sum_{b\le b'}\lambda_{b,b'}\sum_{i,j} a_{ij}\delta(b_i,b)\delta(b_j,b')\right]}$$

where $\{\lambda_{b,b'}\}$ are the Lagragian multipliers enforcing the soft constraint

$$\bar{e}(b,b') = \sum_{G\in\Omega_G}\left[P(G\,|\,\mathbf{b})\left(\sum_{i,j}^{N} a_{ij}\delta(b_i,b)\delta(b_j,b')\right)\right] \quad \textbf{for } b\le b'$$

Marginal

The marginal probability of a link is given by

$$\boxed{p_{ij} = \frac{e^{-\lambda_{b_i,b_j}\left[1+\delta(b_i,b_j)\right]}}{1+e^{-\lambda_{b_i,b_j}\left[1+\delta(b_i,b_j)\right]}}}$$

Where the Lagrangian multipliers are given by

$$\boxed{e^{-\lambda_{b,b'}} = \frac{\pi_{b,b'}}{1-\pi_{b,b'}} \textbf{ for } b \neq b'}$$

$$\boxed{e^{-2\lambda_{b,b}} = \frac{\pi_{b,b}}{1-\pi_{b,b}} \textbf{ for } b = b'}$$

with

$$\boxed{\pi_{b,b'} = \frac{\bar{e}_{b,b'}}{n_{b,b'}}}$$

$$\boxed{n_{b,b'} = \begin{cases} n_b n_{b'} \textbf{ if } & b \neq b' \\ n_b(n_b-1) \textbf{ if } & b = b' \end{cases}}$$

Proof

The Gibbs measure for the block model is given by

$$P(G\,|\,\mathbf{b}) = \frac{1}{Z}\sum_{\mathbf{a}} \exp\left[-\sum_{b\le b'}\lambda_{b,b'}\left(\sum_{i,j} a_{ij}\delta(b_i,b)\delta(b_j,b')\right)\right] = \frac{e^{-H(G)}}{Z}$$

with Hamiltonian

$$H(G) = \sum_{b\le b'}\lambda_{b,b'}\left(\sum_{i,j} a_{ij}\delta(b_i,b)\delta(b_j,b')\right) = \sum_{ij} a_{ij}\sum_{b\le b'}\lambda_{b,b'}\delta(b_i,b)\delta(b_j,b')$$

$$= \sum_{i<j} a_{ij}\sum_{b\le b'}\lambda_{b,b'}\delta(b_i,b)\delta(b_j,b') + \sum_{j>i} a_{ij}\sum_{b\le b'}\lambda_{b,b'}\delta(b_i,b)\delta(b_j,b')$$

$$= \sum_{i<j} a_{ij}\sum_{b\le b'}\lambda_{b,b'}\left[\delta(b_i,b)\delta(b_j,b') + \delta(b_i,b')\delta(b_j,b)\right]$$

$$= \sum_{i<j} a_{ij}\lambda_{\underline{b_{ij}},\overline{b_{ij}}}\left[1+\delta(b_i,b_j)\right] = \sum_{i<j} a_{ij}\left[1+\delta(b_i,b_j)\right]\lambda_{b_i,b_j}$$

Where we are introducing the notation $\lambda_{b,b'} = \lambda_{b',b}$

Proof

Given the expression of the probability of the graph given the block assignment

$$P(G\,|\,\mathbf{b}) = \frac{e^{-H(G)}}{Z} = \frac{1}{Z}\sum_{\mathbf{a}} \exp\left[-\sum_{i<j} a_{ij}\lambda_{b_i,b_j}\left[1+\delta(b_i,b_j)\right]\right]$$

the partition function can be calculated to be

$$Z = \sum_{\mathbf{a}} \exp\left[-\sum_{i<j} a_{ij}\lambda_{b_i,b_j}\left[1+\delta(b_i,b_j)\right]\right] = \prod_{i<j}\left(1+e^{-\lambda_{b_i,b_j}\left[1+\delta(b_i,b_j)\right]}\right)$$

and the marginal distributions are given by

$$p_{ij} = \frac{1}{Z}\sum_{\mathbf{a}} a_{ij}\exp\left[-\sum_{r<s} a_{rs}\lambda_{b_r,b_s}\left[1+\delta(b_r,b_s)\right]\right] = \frac{e^{-\lambda_{b_i,b_j}\left[1+\delta(b_i,b_j)\right]}}{1+e^{-\lambda_{b_i,b_j}\left[1+\delta(b_i,b_j)\right]}}$$

Proof (continuation)

The constraints on the degree correlations for $b \neq b'$ can be written alternatively as

$$\sum_{G \in \Omega_G} P(G \,|\, \mathbf{b}) \left(\sum_{i,j} a_{ij} \delta(b_i, b) \delta(b_j, b') \right) = \bar{e}_{b,b'} \quad \text{or as} \quad \sum_{i,j} p_{ij} \delta(b_i, b) \delta(b_j, b') = \bar{e}_{b,b'}$$

Considering the expression of the marginal probability

$$p_{ij} = \frac{e^{-\lambda_{b,b'}}}{1 + e^{-\lambda_{b,b'}}} \delta(b_i, b) \delta(b_j, b')$$

The constraint reads

$$\frac{e^{-\lambda_{b,b'}}}{1 + e^{-\lambda_{b,b'}}} n_{b,b'} = \bar{e}_{b,b'} \quad \textbf{with} \quad n_{b,b'} = \begin{cases} n_b n_{b'} & \textbf{if} \quad b \neq b' \\ n_b(n_b - 1) & \textbf{if} \quad b = b' \end{cases}$$

Or equivalently

$$\frac{e^{-\lambda_{b,b'}}}{1 + e^{-\lambda_{b,b'}}} = \pi_{b,b'} = \frac{\bar{e}_{b,b'}}{n_{b,b'}}$$

Therefore the Lagrangian multipliers are given by

$$\boxed{e^{-\lambda_{b,b'}} = \frac{\pi_{b,b'}}{1 - \pi_{b,b'}}}$$

Proof (continuation)

The constraints on the degree correlations for $b = b'$ can be written alternatively as

$$\sum_{G \in \Omega_G} P(G \mid \mathbf{b}) \left(\sum_{i,j} a_{ij} \delta(b_i, b) \delta(b_j, b') \right) = \bar{e}_{b,b'} \text{ or as } \sum_{i,j} p_{ij} \delta(b_i, b) \delta(b_j, b') = \bar{e}_{b,b'}$$

Considering the expression of the marginal probability

$$p_{ij} = \frac{e^{-2\lambda_{b,b}}}{1 + e^{-2\lambda_{b,b}}} \delta(b_i, b) \delta(b_j, b)$$

The constraint reads

$$\frac{e^{-2\lambda_{b,b}}}{1 + e^{-2\lambda_{b,b}}} n_{b,b'} = \bar{e}_{b,b} \quad \textbf{with} \quad n_{b,b'} = \begin{cases} n_b n_{b'} & \textbf{if} \quad b \neq b' \\ n_b(n_b - 1) & \textbf{if} \quad b = b' \end{cases}$$

Or equivalently

$$\frac{e^{-2\lambda_{b,b}}}{1 + e^{-2\lambda_{b,b}}} = \pi_{b,b} = \frac{\bar{e}_{b,b}}{n_{b,b}}$$

Therefore the Lagrangian multipliers are given by

$$\boxed{e^{-2\lambda_{b,b}} = \frac{\pi_{b,b}}{1 - \pi_{b,b}}}$$

Microcanonical ensemble

We consider the microcanonical network ensemble

$$\bar{C}_\mu = \sum_{G \in \Omega_G} \left[P(G \,|\, \mathbf{b}) F_\mu(G) \right]$$

Where the soft constraints enforce a given expected total number of links and given expected total cost of the link

Total number of links between nodes of block b and nodes of block b'

$$\boxed{e_{b,b'} = \sum_{i,j}^{N} a_{ij} \delta(b_i, b) \delta(b_j, b') \text{ for } b \leq b'}$$

Microcanonical block model

According to the theory of micro canonical network ensembles (lesson 2) the probability of a network in the microcanonical block modes is given by

$$\boxed{P(G\,|\,\mathbf{b}) = \frac{1}{Z_M}\prod_{b\leq b'}\delta\left(e_{b,b'}, \sum_{ij} a_{ij}\delta(b_i, b)\delta(b_j, b')\right)}$$

where the normalisation constant is given by the total number of network in the ensembles, i.e.

$$\boxed{Z_M = \prod_{b<b'}\binom{n_b n_{b'}}{e_{b,b'}}\prod_b\binom{n_b(n_b-1)/2}{e_{b,b}/2}}$$

The entropy of the micro canonical block model is given by

$$\boxed{\Sigma = \ln\left[\prod_{b<b'}\binom{n_b n_{b'}}{e_{b,b'}}\prod_b\binom{n_b(n_b-1)/2}{e_{b,b}/2}\right]}$$

Equivalence of the microcanonical and canonical block models

As long as the number of constraints is not extensive,

i.e the number of communities satisfies

$$B \ll \sqrt{N}$$

the microcanonical and canonical block models

are equivalent

Block model with constrained degree of the nodes

We consider the canonical network ensemble with two soft constraints of the type

$$C_\mu = \sum_{G \in \Omega_G} \left[P(G \mid \mathbf{b}) F_\mu(G) \right]$$

Where the soft constraints enforce a given expected total number of links and given expected total cost of the link

Expected degree of each node

$$F_i(G) = \sum_{j=1}^{N} a_{ij}$$

Expected total number of links at distance d

$$F_\mu(G) = \sum_{i,j}^{N} a_{ij} \delta(q_i, q) \delta(q_j, q')$$

$$\bar{k}_i = \sum_{G \in \Omega_G} \left[P(G \mid \mathbf{b}) \left(\sum_{j=1}^{N} a_{ij} \right) \right] \quad \bar{e}(q, q') = \sum_{G \in \Omega_G} \left[P(G \mid \mathbf{b}) \left(\sum_{i,j}^{N} a_{ij} \delta(q_i, q) \delta(q_j, q') \right) \right]$$

Block model with constrained degree of the nodes

According to the general treatment of canonical network ensembles the probability of the network reads

$$\boxed{P(G\,|\,\mathbf{b}) = \frac{1}{Z}\exp\left[-\sum_{i=1}^{N}\lambda_i\sum_{j=1}^{N}a_{ij} - \sum_{b\leq b'}\lambda_{b,b'}\sum_{i,j}a_{ij}\delta(b_i,b)\delta(b_j,b')\right]}$$

where $\{\lambda_{b,b'}\}$ are the Lagragian multipliers enforcing the soft constraint on the block structure and where $\{\lambda_i\}$ are the Lagrangian multipliers enforcing the constraints on the node degrees

Marginal

The marginal probability of a link is given by

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

Where the Lagrangian multipliers satisfying the constraints

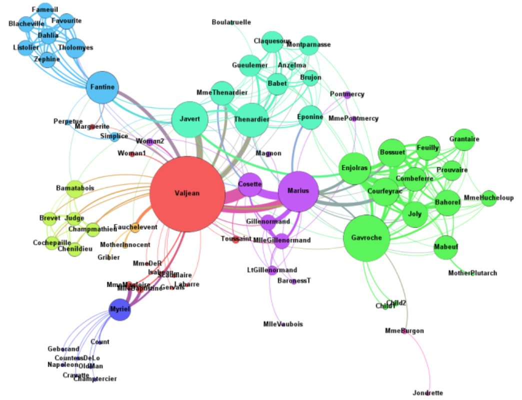
$$\boxed{\bar{k}_i = \sum_{j=1}^{N} p_{ij}}$$

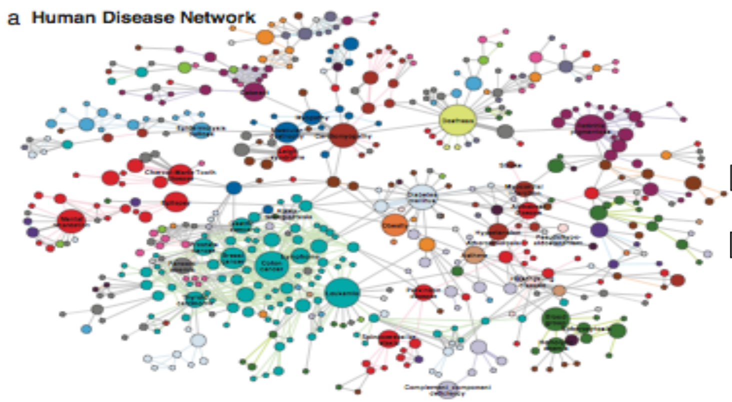
$$\boxed{\bar{e}_{b,b'} = \sum_{ij}^{N} p_{ij}\delta(b_i, b)\delta(b_j, b')}$$

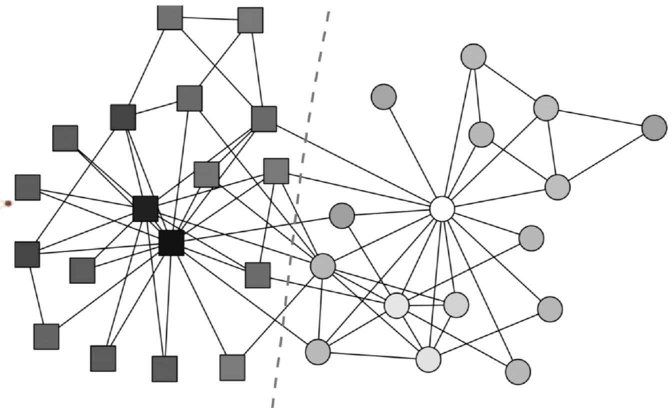
(left as an exercise)

Inference of block structure

Communities





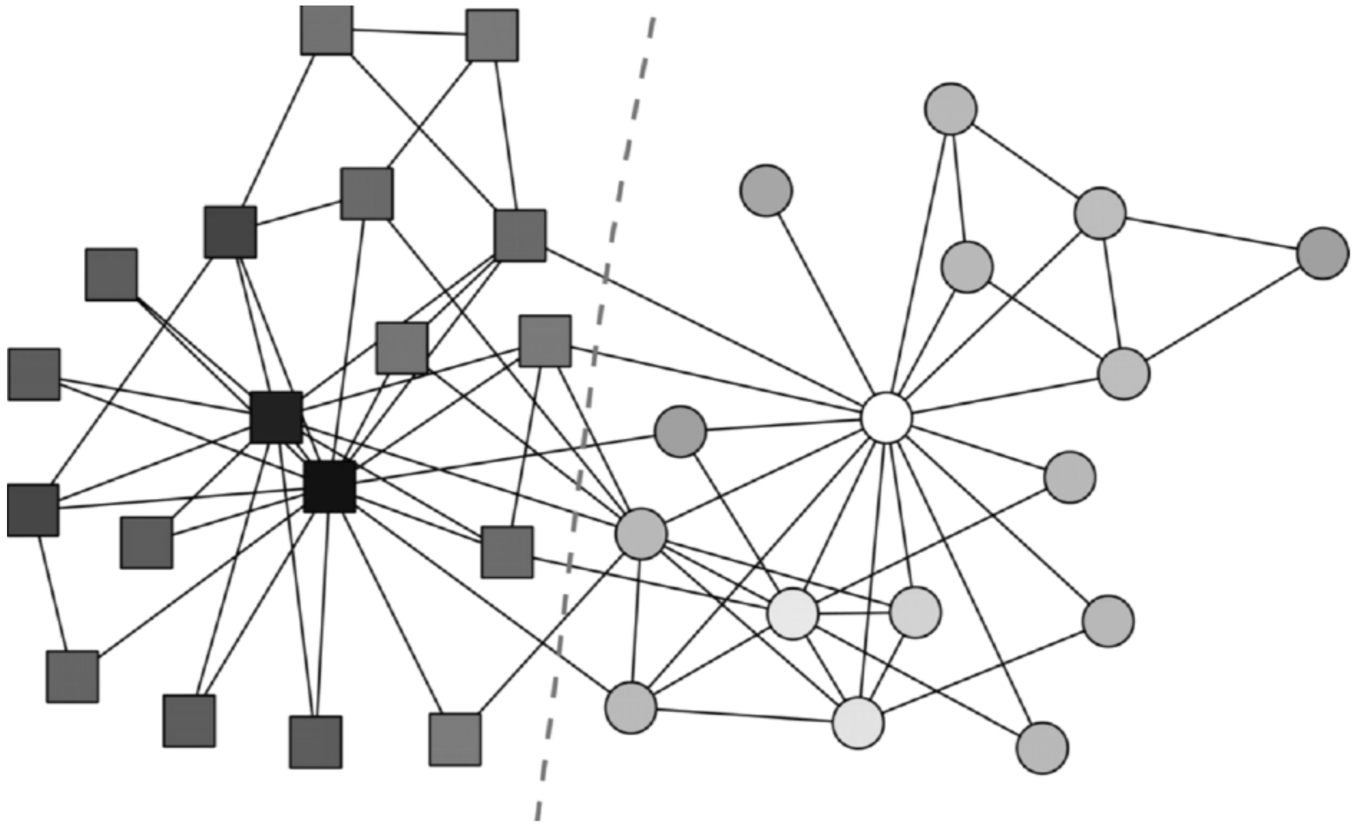


describe

the mesoscale organization of

Social, Biological and Technological systems

The Zackary-Karate Club Network



The Zackary-Karate Club Club

Zackary Karate Club Club Trophy

The first scientist at any conference on networks who uses Zackary's karate Club as an example is inducted into the Zackary Karate Club Club and awarded a prize.

Inference

Until now we have assumed to know the block assignment and we have modelled the network using the probability

$$P(G \mid \mathbf{b})$$

If we want to infer the block assignment given the network G

we need to consider the **posterior distribution**

$$\boxed{P(\mathbf{b} \mid G)}$$

of a given block assignment

given our priors beliefs encoded in the **prior distribution**

$$\boxed{P(\mathbf{b})}$$

indicating our assumptions on the distribution of the block assignment and the statistical network model from which we believe the model is sampled.

Posterior distribution

Using Bayes rule

we can express the posterior distribution as

$$\boxed{P(\mathbf{b} \,|\, G) = \frac{P(G \,|\, \mathbf{b}) P(\mathbf{b})}{P(G)}}$$

where

$$\boxed{P(G) = \sum_{\mathbf{b}} P(G \,|\, \mathbf{b}) P(\mathbf{b})}$$

Maximum a posteriori (MAP) estimator

The maximum a posterior (MAP)

Estimator of the block assignment

infers the block assignment

by maximising the prior distribution

$$\boxed{\mathbf{b}^{\star} = \arg\max_{\mathbf{b}} P(\mathbf{b} \,|\, G)}$$

Minimum description length

The description length is defined as

$$\boxed{L(G, \mathbf{b}) = -\log_2 \left[P(G \,|\, \mathbf{b}) P(\mathbf{b})\right]}$$

and indicates the number of bits needed to communicate the data within the chosen model

The prior can be written as

$$\boxed{P(\mathbf{b} \,|\, G) = \frac{P(G \,|\, \mathbf{b}) P(\mathbf{b})}{P(G)} = \frac{2^{-L(G,\mathbf{b})}}{P(G)}}$$

Since the probability P(G) does not depend on the block assignment optimising the prior distribution is equivalent to minimise the description length

Priors

The choice of priors might depend on the particular problem to study.

However in general is better to consider unbiased priors as the one described below.

The prior probability on block assignment can be factored in terms

$$\boxed{P(\mathbf{b}) = P(\mathbf{b} \mid \mathbf{N}) P(\mathbf{N} \mid B) P(B)}$$

Where B is the total number of non-empty blocks and **N** is the distribution of number of nodes for each block b

Priors

For instance one can choose

$$P(B) = \frac{1}{N}$$

probability of having B blocks

$$P(\mathbf{N} \mid B) = \left[\binom{N-1}{B-1} \right]^{-1}$$

uniform probability over distributions N_b
of nodes per block forbidding empty blocks

$$P(\mathbf{b} \mid \mathbf{N}) = \left[\frac{N!}{\prod_b N_b!} \right]^{-1}$$

probability of each block assignement$\mathbf{b}$
constaining N_b nodes per block b

Peixoto “Stochatic block modelling”

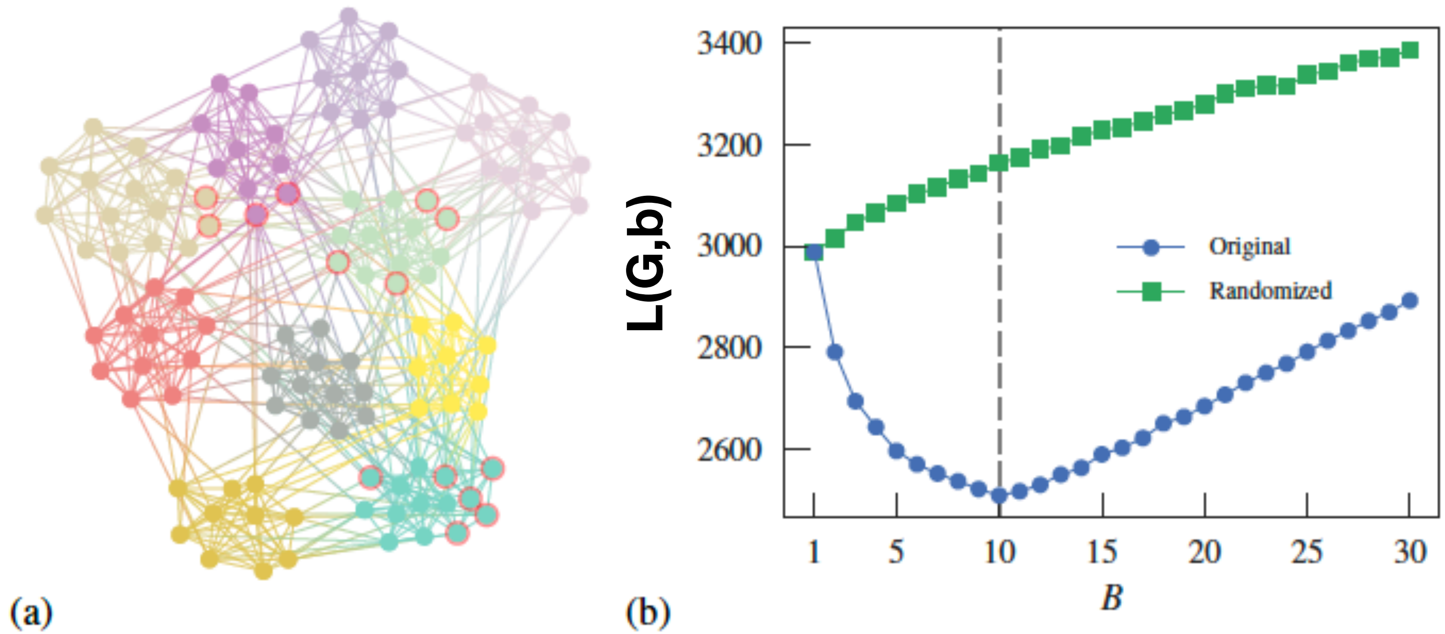
The importance of the prior

The prior distribution is important

In fact maximising only the likelihood could result in an overfitting where the “best” model is the model with a number of blocks equal to the number of nodes

The prior presence in the minimum description length however typically prevents from overfitting because models dependent on more parameters are associated to a longer description length.

Bayesian inference of the SBM for a network of American college football team



Minimizing the description lengths gives 10 communities

Peixoto "Stochastic block modelling"

Inference using Markov-Chain MonteCarlo (MCMC)

- The maximum a posteriori estimator (MAP) can lead often to NP-hard problems.
- An alternative way to estimate the block assignment is to sample from the posterior distribution performing a MCMC algorithm
- Where we consider a move proposal $\mathbf{b} \rightarrow \mathbf{b}'$
- And we accept it with probability $\boxed{\Pi(\mathbf{b} \rightarrow \mathbf{b}') = \min\left(1, \frac{P(\mathbf{b}' \mid G) P(\mathbf{b} \mid \mathbf{b}')}{P(\mathbf{b} \mid G) P(\mathbf{b}' \mid \mathbf{b})}\right)}$

Global constraints only dependent on hidden variables

Hidden variables

- Consider a simple network of N nodes
- Let us assume that each node i is assigned a set of hidden variables $\mathbf{x}_i$

These hidden variables can indicate any kind of metadata

i.e. position in space, classification,

proxy for the degree of the node

“Payoff” of a link

Let us consider a given function

$$f(\mathbf{x}, \mathbf{y})$$

indicating the “payoff”

(“cost” or “benefit”)

of a link between a node with hidden variable $\mathbf{x}$

and a node with hidden variable $\mathbf{y}$

Canonical ensemble

We consider the canonical network ensemble with two soft constraints of the type

$$C_\mu = \sum_{G \in \Omega_G} \left[P(G) F_\mu(G) \right]$$

Where the soft constraints enforce a given expected total number of links and given expected total cost of the link.

Note that for such global constraints

the canonical and micro canonical ensembles are equivalent

Expected total number of links

$$F_1(G) = \sum_{i<j}^{N} a_{ij}$$

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

Expected total payoff of the links

$$F_2(G) = \sum_{i<j}^{N} a_{ij} f(\mathbf{x}_i, \mathbf{x}_j)$$

$$\bar{C} = \sum_{G \in \Omega_G} \left[P(G) \left(\sum_{i<j}^{N} a_{ij} f(\mathbf{x}_i, \mathbf{x}_j) \right) \right]$$

Gibbs measure

Applying the general theory of canonical network ensemble it is immediate to derive the Gibbs measure of this ensemble given by

$$\boxed{P(G) = \frac{1}{Z} \exp \left[-\beta \sum_{i<j} a_{ij} f(\mathbf{x}_i, \mathbf{x}_j) - \mu \sum_{i<j} a_{ij} \right]}$$

Where μ and β are Lagrangian multipliers fixed by the constraints

Expected total number of links

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

$$\bar{L} = \sum_{i<j}^{N} p_{ij}$$

Expected total cost of the links

$$\bar{C} = \sum_{G \in \Omega_G} \left[P(G) \left(\sum_{i<j}^{N} a_{ij} f(\mathbf{x}_i, \mathbf{x}_j) \right) \right]$$

$$\bar{C} = \sum_{i<j}^{N} p_{ij} f(\mathbf{x}_i, \mathbf{x}_j)$$

Marginal probability

The marginal probability that two nodes

i with hidden variables $\mathbf{x}_i$

and

j with hidden variables $\mathbf{x}_j$

are linked

is only a function of the hidden variables

and is given by

$$\boxed{p_{ij} = p(\mathbf{x}_i, \mathbf{x}_j) = \frac{e^{-\beta f(\mathbf{x}_i, \mathbf{x}_j) - \mu}}{1 + e^{-\beta f(\mathbf{x}_i, \mathbf{x}_j) - \mu}}}$$

(left as an exercise)

Scalar hidden variable

If the hidden variable is a scalar the marginal probability reads

$$\boxed{p_{ij} = p(x_i, x_j) = \frac{e^{-\beta f(x_i,x_j)-\mu}}{1 + e^{-\beta f(x_i,x_j)-\mu}}}$$

All the nodes with hidden variable $x_i = x$ have the same expected degree given by

$$\boxed{\bar{k}_i = N \int dx \rho(x) p(x_i, x)}$$

where $\rho(x)$ is the density distribution function of the nodes with hidden variables $x_i = x$

Therefore the hidden variable is a proxy for the degree

Scalar hidden variable

The marginal probability for scalar hidden variables is given by

$$\boxed{p(x_i, x_j) = \frac{e^{-\beta f(x_i,x_j)-\mu}}{1 + e^{-\beta f(x_i,x_j)-\mu}}}$$

If we take

$$\boxed{\beta f(x_i, x_j) = \lambda_{x_i} + \lambda_{x_j} \qquad \theta_{x_i} = e^{-\lambda_{x_i}} \qquad z = e^{\mu}}$$

We have

$$\boxed{p(x_i, x_j) = \frac{e^{-(\lambda_{x_i}+\lambda_{x_j})-\mu}}{1 + e^{-(\lambda_{x_i}+\lambda_{x_j})-\mu}} = \frac{\theta_{x_i}\theta_{x_j}/z}{1 + \theta_{x_i}\theta_{x_j}/z}}$$

Where μ or equivalently z fix the expected total number of links

The scalar hidden variables

The scalar hidden variables can be:

- The GDP of a country in the World Trade Networks
- Assets and liabilities of a bank in financial networks
- Any variable that can be considered as a good proxy for the degree of a node

Spatial networks

The marginal probability for scalar hidden variables is given by

$$\boxed{p(x_i, x_j) = \frac{e^{-\beta f(x_i,x_j)-\mu}}{1+e^{-\beta f(x_i,x_j)-\mu}}}$$

If we take a constraints depending on the distance between the hidden variables

$$\boxed{\beta f(x_i, x_j) = \lambda_{\mathbf{x}_i} + \lambda_{\mathbf{x}_j} + \beta\omega(d(\mathbf{x}_i, \mathbf{x}_j))}$$

we have

$$\boxed{p(x_i, x_j) = \frac{e^{-(\lambda_{\mathbf{x}_i}+\lambda_{\mathbf{x}_j})-\mu}W(d_{ij})}{1+e^{-(\lambda_{\mathbf{x}_i}+\lambda_{\mathbf{x}_j})-\mu}W(d_{ij})} = \frac{\theta_{\mathbf{x}_i}\theta_{\mathbf{x}_j}W(d_{ij})/z}{1+\theta_{\mathbf{x}_i}\theta_{\mathbf{x}_j}W(d_{ij})/z}}$$

where

$$\boxed{W(d_{ij}) = e^{-\beta\omega(d_{ij})} \qquad \theta_{\mathbf{x}_i} = e^{-\lambda_{x_i}} \qquad z = e^{\mu}}$$

Spatial networks

With the choice

$$\beta f(x_i, x_j) = \lambda_{\mathbf{x}_i} + \lambda_{\mathbf{x}_j} + \beta\omega(d(\mathbf{x}_i, \mathbf{x}_j))$$

The payoff of a link can depend on the distance in different ways .For instance

If $\omega(d_{ij}) = \ln d_{ij}$ **payoff is linear with the order of magnitude of the distance**

If $\omega(d_{ij}) = d_{ij}$ **payoff is linear with the distance**

The marginal reads

$$p(x_i, x_j) = \frac{e^{-(\lambda_{\mathbf{x}_i}+\lambda_{\mathbf{x}_j})-\mu-\beta\omega(d(\mathbf{x}_i,\mathbf{x}_j))}}{1 + e^{-(\lambda_{\mathbf{x}_i}+\lambda_{\mathbf{x}_j})-\mu-\beta\omega(d(\mathbf{x}_i,\mathbf{x}_j))}} = \frac{\theta_{\mathbf{x}_i}\theta_{\mathbf{x}_j}W(d_{ij})/z}{1 + \theta_{\mathbf{x}_i}\theta_{\mathbf{x}_j}W(d_{ij})/z}$$

with

If $\omega(d_{ij}) = \ln d_{ij}$ $\qquad W(d_{ij}) = d_{ij}^{-\beta}$ **power-law decay with distance**

If $\omega(d_{ij}) = d_{ij}$ $\qquad W(d_{ij}) = e^{-\beta d_{ij}}$ **exponential decay with distance**

The observed power-law decay of W(d)

for the American airport network and the Indian train and airport network

can be interpreted as the outcome

of an effective “payoff” of the connection

growing proportionally

to the order of magnitude of distance between the nodes

Soft random geometric networks

If we take a "payoff" function of the form

$$\boxed{f(\mathbf{x}_i, \mathbf{x}_j) = d(\mathbf{x}_i, \mathbf{x}_j) - r_0}$$

The marginal takes the Fermi-Dirac form

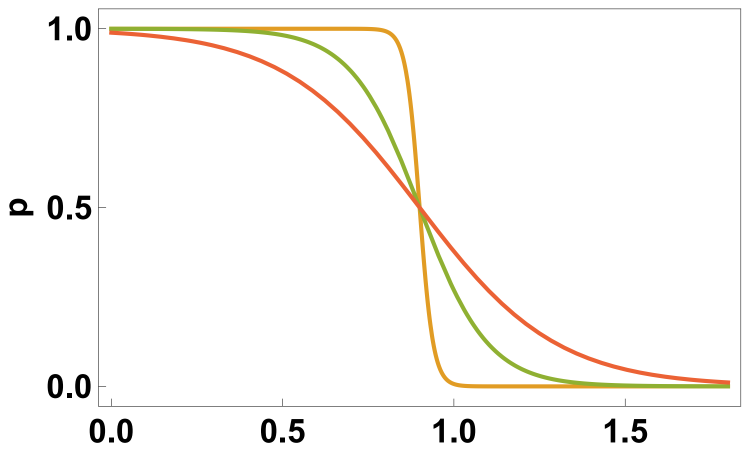
$$\boxed{p_{ij} = \frac{e^{-\beta(d_{ij}-r_0)-\mu}}{1+e^{-\beta(d_{ij}-r_0)-\mu}} = \frac{1}{e^{\beta(d_{ij}-r_0)+\mu}+1}}$$

which has functional form

For μ=0 and $\beta \to \infty$

Two nodes at distance d are connected if and only if

$d \leq r_0$

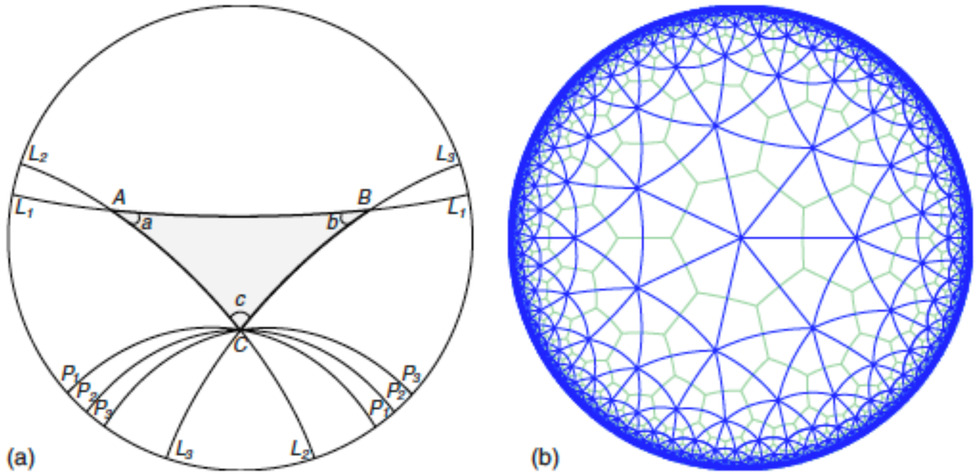


Here μ=0

and β=20,10,5

The larger is β the steepest is the function

Random geometric hyperbolic networks



Consider the hyperbolic plane $\mathbb{H}^2$ with curvature $R = -\zeta^2$

where two nodes of polar coordinates (r, θ) **and** (r', θ')

have distance d satisfying

$$\cosh \zeta d = \cosh \zeta r \cosh \zeta r' - \sinh \zeta r \sinh \zeta r' \cos \Delta\theta$$

which can be approximated as

$$\boxed{d \simeq r + r' + \frac{2}{\zeta} \ln \sin \frac{\Delta\theta}{2} \simeq r + r' + \frac{1}{\zeta} \ln \Delta\theta}$$

Random geometric hyperbolic networks

Using the expression for the hyperbolic distances

$$\boxed{d \simeq r + r' + \frac{2}{\zeta} \ln \sin \frac{\Delta\theta}{2} \simeq r + r' + \frac{1}{\zeta} \ln \Delta\theta}$$

The hyperbolic network are geometric random graph with payoff of each link

given by

$$\boxed{f(\mathbf{x}_i, \mathbf{x}_j) = \zeta(d_{ij} - r_0) = \zeta r_i + \zeta r_j + \ln \Delta\theta_{ij} - r_0}$$

where r_0 is a suitable parameter that can tune the number of the links

By setting $\mu = 0$

the marginal probability reads

$$\boxed{p_{ij} = \frac{e^{-\zeta\beta(d_{ij} - r_0)}}{1 + e^{-\beta\zeta(d_{ij} - r_0)}} = \frac{1}{e^{\beta\zeta(d_{ij} - r_0)} + 1}}$$

Random geometric hyperbolic networks

Considering the marginal

$$\boxed{p_{ij} = \frac{e^{-\beta\zeta(d_{ij}-r_0)}}{1+e^{-\beta\zeta(d_{ij}-r_0)}}}$$

with the hyperbolic network choice

$$\zeta(d_{ij}-r_0) = \zeta r_i + \zeta r_j + \ln \Delta\theta_{ij} - \zeta r_0$$

By putting

$$\boxed{\begin{aligned} &\theta_{x_i} = e^{-\beta\zeta(r_i - r_0/2)} \\ &W(\Delta\theta_i) = e^{-\beta \ln \Delta\theta_{ij}} = (\Delta\theta_{ij})^{-\beta} = (\Delta\theta_{ij})^{-\beta} \end{aligned}}$$

We obtain the marginal

$$\boxed{p_{ij} = \frac{\theta_i\theta_j W(\Delta\theta_{ij})}{1+\theta_i\theta_j W(\Delta\theta_{ij})} = \frac{\theta_i\theta_j(\Delta\theta_{ij})^{-\beta}}{1+\theta_i\theta_j(\Delta\theta_{ij})^{-\beta}}}$$

Therefore the random geometric hyperbolic network model reduces

to a maximum entropy model in which we fix the degrees

and we have a cost of the link proportional to the

order of magnitude of the angular distance between the nodes

Degree of the nodes

The only difference between the random geometric hyperbolic network and the maximum entropy ensemble is that the degrees are fixed by the embedding of the nodes in space, i.e.

$$\boxed{\theta_{x_i} = e^{-\beta\zeta(r_i - r_0/2)}}$$

and

$$\boxed{\bar{k}_i = \sum_{j=1}^{N} p_{ij} = \sum_{j=1}^{N} \frac{\theta_i\theta_j(\Delta\theta_{ij})^{-\beta}}{1 + \theta_i\theta_j(\Delta\theta_{ij})^{-\beta}}}$$

Therefore the degree distribution is fixed by the distribution of nodes in space.

Degree of the nodes
In random geometric hyperbolic network

The only difference between the random geometric hyperbolic network and the maximum entropy ensemble is that the degrees are fixed by the embedding of the nodes in space, i.e.

$$\boxed{\theta_{x_i} = e^{-\beta\zeta(r_i - r_0/2)}}$$

and

$$\boxed{\bar{k}_i = \sum_{j=1}^{N} p_{ij} = \sum_{j=1}^{N} \frac{\theta_i\theta_j(\Delta\theta_{ij})^{-\beta}}{1 + \theta_i\theta_j(\Delta\theta_{ij})^{-\beta}}}$$

Therefore the degree distribution is fixed by the distribution of nodes in space

Power-law random geometric hyperbolic model

Considering the constraints $$\boxed{\bar{k}_i = \sum_{j=1}^{N} p_{ij} = \sum_{j=1}^{N} \frac{\theta_i\theta_j(\Delta\theta_{ij})^{-\beta}}{1+\theta_i\theta_j(\Delta\theta_{ij})^{-\beta}}}$$

we can assume in the sparse regime that

$$k \propto \theta \propto e^{-\beta\zeta r}$$

Therefore the radial distribution with that enforces a power-law degree distribution with exponent $\gamma \geq 2$ is exponential

$$\boxed{p(k) \simeq Ck^{-\gamma} \longrightarrow \rho(r) \propto e^{\beta\zeta(\gamma-1)r}}$$

In fact using the transformation due to the change of variables $\rho(r)dr = P(k)dk$

we get $$\rho(r) = P(k)\frac{dk}{dr}\bigg|_{k=Ae^{-\beta\zeta r}} = AC\beta\zeta k^{-\gamma}e^{-\beta\zeta r}\bigg|_{k=Ae^{-\beta\zeta r}} = AC\beta\zeta e^{\beta\zeta(\gamma-1)r}$$

Final remarks

In this lesson we have covered maximum entropy network ensembles that

1. -go beyond enforcing exclusively the degree sequence

2. -describe network ensembles with hidden variables

Moreover we have described the

3. -basic principles of inference from block models