Breaking of ensemble equivalence in complex networks

Frank den Hollander
Mathematical Institute, Leiden University

P.C. Mahalanobis Lectures 2015–2016,

Indian Statistical Institute, Delhi & Bangalore & Kolkata.

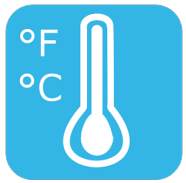
§ STATISTICAL PHYSICS

Systems consisting of a very large number of interacting components can be described by statistical ensembles, i.e., probability distributions on spaces of configurations.

Two important examples are:

I.	micro-canonical ensemble
II.	canonical ensemble

The former fixes the energy of the system, the latter fixes the average energy of the system, with temperature as the control parameter.



N particles described by a configuration space $\mathcal{C}_N$ and an energy function $E\colon \mathcal{C}_N \to \mathbb{R}$.

- Micro-canonical ensemble:

$$P_N^{\mathsf{mic}}(c) = \begin{cases} 1/\Omega_{E^*} & \text{if } E(c) = E^*, \\ 0 & \text{else,} \end{cases}$$

where $\Omega_{E^*} = |\{c \in \mathcal{C}_N\colon\ E(c) = E^*\}|$.

- Canonical ensemble:

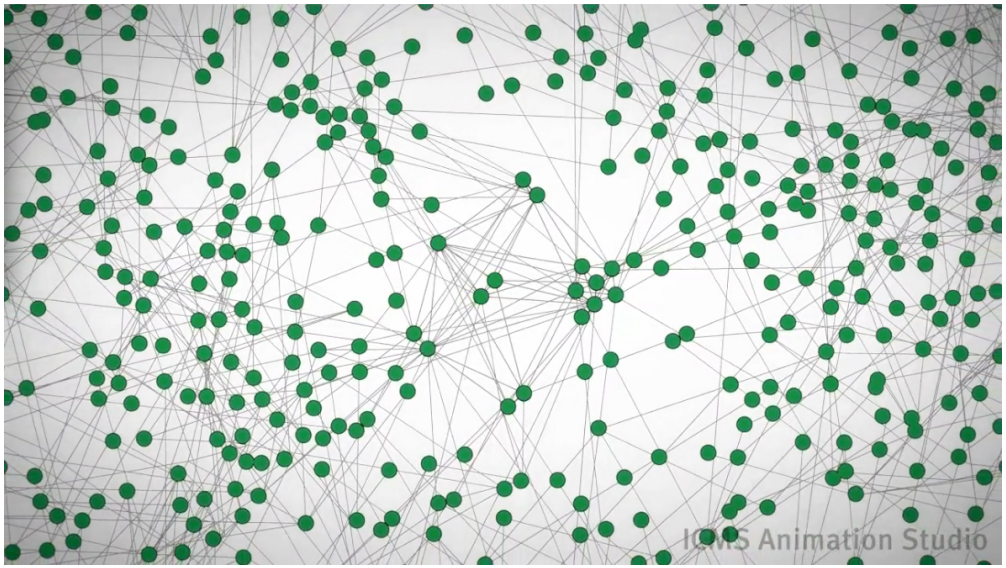
$$P_N^{\mathsf{can}}(c) = \frac{1}{\mathcal{N}(\beta^*)}\, \mathrm{e}^{-\beta^* E(c)},$$

where the inverse temperature β^* is to be chosen such that the average of $E(c)$ equals E^* (β^* is a Lagrange multiplier).

In statistical physics textbooks the two ensembles are often assumed (!) to be thermodynamically equivalent, the idea being that for large systems the energy is always close to its average value.

This is reasonable only for systems with interactions that are short-ranged. But counter examples have been found for systems with interactions that are long-ranged.

In this talk we will be interested in large random graphs, i.e., the two ensembles of interest live on the set $\mathcal{G}_N$ of all graphs with N vertices where $N \to \infty$.



A realisation of a large random graph

§ DEFINITIONS

Given are a vector-valued function $\vec{C}$ on $\mathcal{G}_N$, and a specific vector $\vec{C}^*$ called the constraint.

I. The micro-canonical ensemble is defined by

$$P_N^{\text{mic}}(G) = \begin{cases} 1/\Omega_{\vec{C}^*} & \text{if } \vec{C}(G) = \vec{C}^*, \\ 0 & \text{else,} \end{cases}$$

where $\Omega_{\vec{C}^*} = |\{G \in \mathcal{G}_N \colon \vec{C}(G) = \vec{C}^*\}|$.

II. The canonical ensemble is defined by

$$P_N^{\text{can}}(G) = \frac{1}{\mathcal{N}(\vec{\theta}^*)} \, \mathrm{e}^{-\vec{\theta}^* \cdot \vec{C}(G)},$$

where $\mathcal{N}(\vec{\theta}^*)$ is the normalising constant and $\vec{\theta}^*$ is to be chosen such that the average of $\vec{C}(G)$ equals $\vec{C}^*$.

INTERPRETATION

- P_N^{mic} models a random graph of which no information is available other than the constraint.

- P_N^{can} models a random graph of which no information is available other than the average constraint.

The latter can be viewed as the solution to the problem of statistical inference on the basis of the partial information provided by the average constraint: maximum likelihood.

§ EQUIVALENCE

The two ensembles P_N^{mic} and P_N^{can} are said to be equivalent when their relative entropy per vertex defined by

$$\boxed{s_N\left(P_N^{\text{mic}} \mid P_N^{\text{can}}\right) = \frac{1}{N} \sum_{G \in \mathcal{G}_N} P_N^{\text{mic}}(G) \log\left(\frac{P_N^{\text{mic}}(G)}{P_N^{\text{can}}(G)}\right)}$$

tends to zero as $N \to \infty$.

Complex networks form a new class of systems where the breaking of ensemble equivalence manifests itself through the presence of global constraints.

In this talk we illustrate this phenomenon via a number of examples based on the so-called configuration model.

Literature:

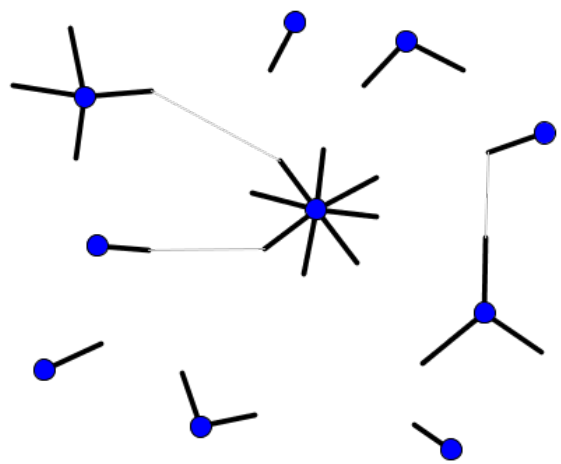
D. Garlaschelli, F. den Hollander, J. de Mol, T. Squartini,
Phys. Rev. Lett. 115 (2015) 268701.

D. Garlaschelli, F. den Hollander, A. Roccaverde,
Nieuw Archief voor Wiskunde 5/16 (2015) 207–209.

D. Garlaschelli, F. den Hollander, A. Roccaverde,
work in progress.

§ THE CONFIGURATION MODEL

Vertices are assigned degrees. This is done by assigning to each vertex a number of half-edges and matching these up randomly.



In what follows we will consider graphs with a community structure. We will be considering three examples:

1. Uni-partite community graphs.

2. Bi-partite community graphs.

3. Multi-partite community graphs.

§ UNI-PARTITE COMMUNITY GRAPHS

Consider a graph $G = (V, E)$ with vertex set $V = \{1, \ldots, N\}$ and edge set E such that all the vertices have prescribed degrees $k_1, \ldots, k_N \in \mathbb{N}_0$, i.e., $\vec{C}^* = (k_1, \ldots, k_N)$. Under the sparseness condition

$$k_{\max} = \max_{1 \le i \le N} k_i = o(\sqrt{N}), \qquad N \to \infty,$$

the number of such graphs is given by the formula

$$\Omega_N(k_1, \ldots, k_N)$$
$$= \frac{\sqrt{2}(Nm_1/e)^{2Nm_1}}{\prod_{1 \le i \le N} k_i!} \exp\left[-\frac{(m_2)^2 - (m_1)^2}{4(m_1)^2} + o\left(\frac{m_3}{N}\right)\right]$$

with

$$m_l = \frac{1}{N} \sum_{1 \le i \le N} (k_i)^l, \quad l = 1, 2, 3.$$

I. Under the micro-canonical ensemble the distribution of G on the set $\mathcal{G}_N$ is uniform:

$$P_N^{\mathsf{mic}}(G) = \frac{1}{\Omega_N(k_1, \ldots, k_N)}.$$

II. Under the canonical ensemble the distribution of G on the set $\mathcal{G}_N$ is tilted exponential:

$$\begin{aligned} P_N^{\mathsf{can}}(G) &= \frac{1}{\mathcal{N}(\vec{C}^*)} \mathrm{e}^{-\sum_{i \in V} \theta_i^* k_i(G)} \\ &= \prod_{e \in E} (p_e^*)^{a_e(G)} (1 - p_e^*)^{1 - a_e(G)}. \end{aligned}$$

Here, $a_e(G) = 1_{\{e \in E\}}$ is the adjacency matrix and $(\theta_i^*)_{i \in V}$ is replaced by $(p_e^*)_{e \in E}$ with $p_{(ij)}^* = \mathrm{e}^{-\theta_i^* - \theta_j^*} / (1 + \mathrm{e}^{-\theta_i^* - \theta_j^*})$.

But $(p_e^*)_{e\in E}$ must be chosen such that the average degree of vertex i is k_i. It turns out that in the sparse regime

$$p^*_{(ij)} \sim \frac{k_i k_j}{N m_1}, \qquad N \to \infty.$$

Combining the above formulas, we find after some simple calculation that

$$\boxed{s = \lim_{N\to\infty} s_N\left(P_N^{\mathsf{mic}} \mid P_N^{\mathsf{can}}\right) = \lim_{N\to\infty} \frac{1}{N} \sum_{1\le i\le N} g(k_i)}$$

with

$$g(k) = \log\left(\frac{k!}{k^k e^{-k}}\right), \qquad k \in \mathbb{N}_0.$$

The formula for s has an interesting interpretation. Let

$$f_N = \frac{1}{N} \sum_{i=1}^{N} \delta_{k_i}$$

denote the empirical distribution of the degrees. Let POI_k denote the Poisson distribution with average k. Then

$$g(k) = s\Big(\delta_k \mid \mathsf{POI}_k\Big).$$

Hence, if $\lim_{N\to\infty} \|f_N - f\|_{\ell^1(g)} = 0$ for some $f \in \mathcal{M}_1(\mathbb{N}_0)$ with $\|f\|_{\ell^1(g)} < \infty$, then the above formula becomes

$$s = \sum_{k\in\mathbb{N}_0} f(k)g(k) = \|f\|_{\ell^1(g)},$$

i.e., each vertex with degree k contributes an amount $g(k)$ to the relative entropy.

In summary:

$$\boxed{\text{Under the canonical ensemble the degrees are asymptotically Poisson.}}$$

In other words, in the limit as $N \to \infty$ the vertices have a random degree with a Poisson distribution that has a mean compatible with the constraint.

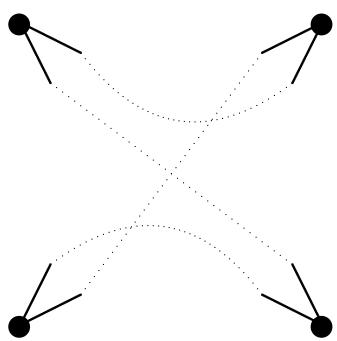
Consequently, there is breaking of ensemble equivalence for all $f \neq \delta_0$.

Example 1:

$$k_1 = \cdots = k_N = k = o(\sqrt{N}).$$

For k-regular graphs:

$$s = g(k) > 0.$$



$N = 4,\ k = 2$

Example 2:

$N^{-1}\,|\{1 \le i \le N\colon\ k_i = k\}| \approx C\,k^{-\tau},\ 1 \le k \le k_{\text{cutoff}},$
with $k_{\text{cutoff}} = o(\sqrt{N})$ and $\tau \in (1,\infty)$ a tail exponent.

For scale-free graphs:

$$s \approx \frac{1}{2(\tau-1)} + \log\sqrt{2\pi} > 0.$$

§ BI-PARTITE COMMUNITY GRAPHS

Suppose that we are given two sets of vertices V, V' of sizes N, N'. We fix the two degree sequences

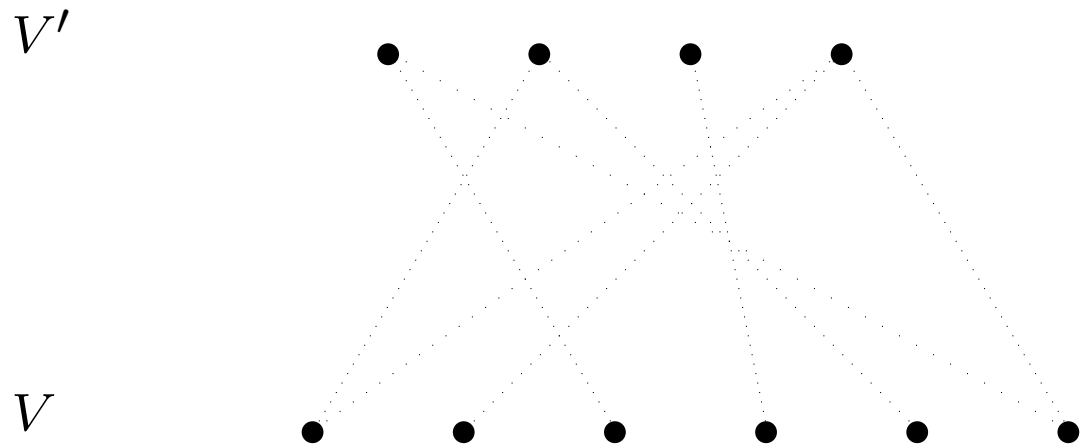
$$(k_1, \ldots, k_N), \qquad (k'_1, \ldots, k'_{N'}),$$

and we only allow edges between V and V'. In particular,

$$\sum_{i=1}^{N} k_i = \sum_{j=1}^{N'} k'_j = L$$

with L the total number of edges. The definition of $P^{\mathsf{mic}}_{N,N'}$ and $P^{\mathsf{can}}_{N,N'}$ is adapted accordingly, with

$$\vec{C}^* = (k_1, \ldots, k_N, k'_1, \ldots, k'_{N'}).$$



$N = 6,\ N' = 4$

Under the sparseness condition

$$k_{\mathsf{max}}\, l_{\mathsf{max}} = o(L^{2/3})$$

it can be shown that

$$s = \lim_{N,N'\to\infty} \frac{s_{N,N'}(P^{\mathsf{mic}}_{N,N'} \mid P^{\mathsf{can}}_{N,N'})}{N+N'}$$

exists and equals

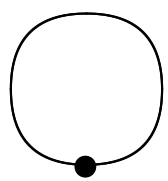
$$\begin{aligned} s &= \alpha \sum_{k\in\mathbb{N}_0} f_V(k)g(k) + (1-\alpha) \sum_{k\in\mathbb{N}_0} f_{V'}(k)g(k) \\ &= \alpha \|f_V\|_{\ell^1(g)} + (1-\alpha)\|f_{V'}\|_{\ell^1(g)} \end{aligned}$$

when $f_V, f_{V'}$ are the limits of the two empirical distributions of the degrees and

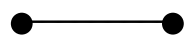
$$\lim_{N,N'\to\infty} \frac{N}{N+N'} = \alpha \in [0,1].$$

§ MULTI-PARTITE COMMUNITY GRAPHS

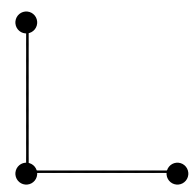
It is natural to consider constraints that are controlled by a multi-partite graph.



uni-partite



bi-partite



etc.

Fix a number $M \in \mathbb{N}$ of communities. Each community C_s has n_s vertices. The set of corresponding graphs is denoted by $\mathcal{G}_{n_1,\ldots,n_M}$.

Any graph $G \in \mathcal{G}_{n_1,\ldots,n_M}$ is represented by its $n_1 \times \cdots \times n_M$ adjacency matrix with elements

$$a_{(i,j)}(G) = \begin{cases} 1 & \text{there is an edge between } i \text{ and } j, \\ 0 & \text{otherwise.} \end{cases}$$

Fix the degree sequence for edges between all pairs of communities (including a community and itself). Thus, the constraints are

$$\vec{C}^* = \left\{ \vec{k}^*_{s\to t} = \left(k^{*\,t}_{s_1}, \ldots, k^{*\,t}_{s_{n_h}} \right) : \; s,t = 1,\ldots,M \right\}.$$

Necessarily,

$$L^*_{s,t} = \sum_{i \in C_s} k_i^{*\,t} = \sum_{j \in C_t} k_j^{*\,s} \qquad \forall\, s, t.$$

We abbreviate

$$k^*_{s \to t} = \max_{i \in C_s} k_i^{*\,t}, \quad f^{(n_t)}_{s \to t} = n_s^{-1} \sum_{i \in C_s} \delta_{k_i^{*\,t}},$$

and assume the existence of

$$A_s = \lim_{n_1, \ldots, n_M \to \infty} \frac{n_s}{n} \quad \forall\, s,$$

where $n = \sum_{s=1}^{M} n_s$. The sparse regime corresponds to

$$\begin{aligned} k^*_{s \to t} k^*_{s \to t} &= o({L^*_{s,t}}^{2/3}), \quad && n_s, n_t \to \infty \;\forall\, s \neq t, \\ k^*_{s \to s} &= o(n_s^{1/2}), && n_s \to \infty \;\forall\, s. \end{aligned}$$

We further assume that for all s, t there exist an $f_{s\to t} \in \mathcal{M}_1(\mathbb{N}_0)$ with $\|f_{s\to t}\|_{\ell^1(g)} < \infty$ such that

$$\lim_{n_s\to\infty} \|f_{s\to t}^{(n_s)} - f_{s\to t}\|_{\ell^1(g)} = 0.$$

THEOREM

Subject to the above assumptions,

$$\begin{aligned} s &= \lim_{n_1,\ldots,n_M\to\infty} n^{-1} S\Big(P^{\mathsf{mic}}_{n_1,\ldots,n_M} \mid P^{\mathsf{can}}_{n_1,\ldots,n_M}\Big) \\ &= \sum_{1\le s<t\le M} \Big\{A_s \|f_{s\to t}\|_{\ell^1(g)} + A_t \|f_{t\to s}\|_{\ell^1(g)}\Big\} \\ &\qquad + \sum_{s=1}^{M} A_s \|f_{s\to s}\|_{\ell^1(g)}. \end{aligned}$$

§ CONCLUSION

We have obtained a complete classification of breaking of ensemble equivalence in random graphs with a community structure subject to constraints on the degree sequences between communities.

Breaking occurs if and only if the number of constraints is extensive, i.e., linear in the number of vertices.

§ FUTURE CHALLENGES:

1. What is the formula for the relative entropy per vertex in the non-sparse regime?

2. What happens when constraints are put on the number of triangles, squares, etc. touching a vertex?

3. What is the effect of relaxing constraints, i.e., putting constraints on functions of the degrees?