DENSE RANDOM GRAPHS:

BREAKING OF ENSEMBLE EQUIVALENCE FOR CONSTRAINED GRAPHONS

Frank den Hollander
Mathematical Institute, Leiden University

Spring School in Probability: Complex Networks,
Technische Universität Darmstadt, Germany, 2-6 March 2020.

Michel Mandjes

Andrea Roccaverde

Nicos Starreveld

Electronic Journal of Probability 2018

Random Structures and Algorithms 2020

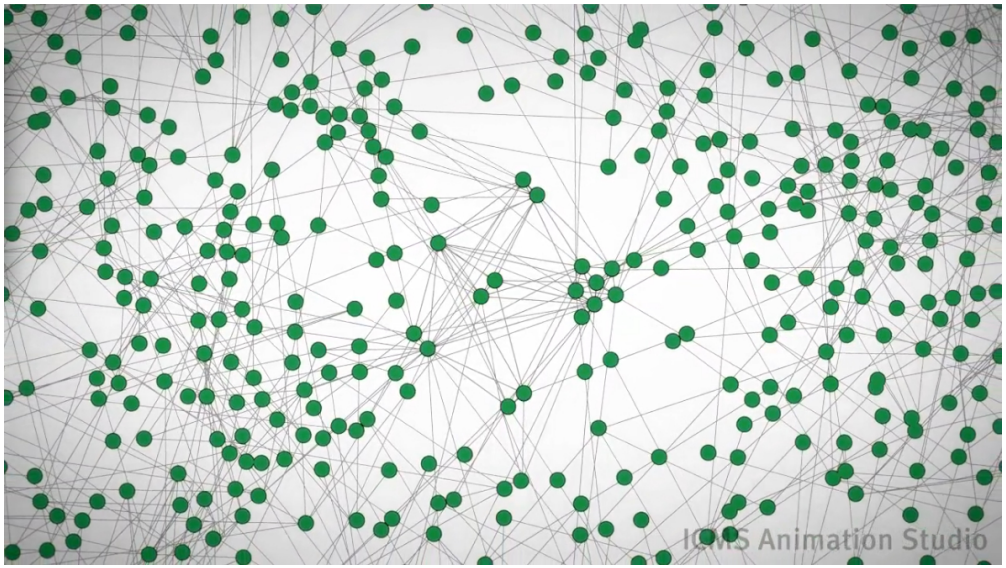
§ CONSTRAINED NETWORKS

The goal of this presentation is to exhibit possible ways to incorporate constraints in the description of large networks.

Networks are modelled as graphs, consisting of vertices connected by edges. Large networks must be modelled as random graphs, where the edges are chosen randomly.

Large networks are typically so complex, that it is both appropriate and effective to use randomness: the network can be viewed as the outcome of a 'probabilistic experiment'.

We will be interested in large random graphs, drawn at random from the set $\mathcal{G}_n$ of all simple graphs with n vertices where $n \to \infty$.



A realisation of a large random graph

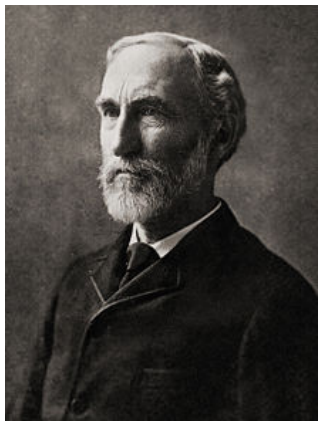
Typically, some a priori information is available about the network, e.g. the degrees of the nodes or the total number of edges and triangles. How should this a priori information be properly incorporated when we choose the probability distribution according to which we build the random graph?

We will focus on two possible choices:

I. hard constraints: true always.
II. soft constraints: true on average.

The two choices capture different situations. In most of the literature on model selection it is assumed (!) that the two choices are asymptotically equivalent, the idea being that for large random graphs all relevant quantities are close to their average value. However, this turns out to be wrong (!).

Care needs to be taken with the choice of 'statistical ensemble'.



Gibbs

§ DEFINITIONS

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

I. The hard-constraint ensemble is defined by

$$P_n^{\text{hard}}(G) = \begin{cases} \frac{1}{\Omega_{\vec{C}^*}} & \text{if } \vec{C}(G) = \vec{C}^*, \\ 0 & \text{else,} \end{cases}$$

where $\Omega_{\vec{C}^*} = |\{G \in \mathcal{G}_n \colon \vec{C}(G) = \vec{C}^*\}| > 0$.

II. The soft-constraint ensemble is defined by Jaynes 1957

$$P_n^{\text{soft}}(G) = \frac{1}{Z(\vec{\theta}^*)} \, \mathrm{e}^{(\vec{\theta}^*, \vec{C}(G))},$$

where $\vec{\theta}^*$ is a control parameter that must be chosen such that $\sum_{G \in \mathcal{G}_n} \vec{C}(G) P_n^{\text{soft}}(G) = \vec{C}^*$, and $Z(\vec{\theta}^*)$ plays the role of a normalisation constant.

INTERPRETATION

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

Which of the two should be used to model a specific real-world network depends on the a priori information that is available about the network.

§ ENSEMBLE EQUIVALENCE Touchette 2015

P_n^{hard} and P_n^{soft} are said to be equivalent when their relative entropy defined by

$$S_n\left(P_n^{\text{hard}} \mid P_n^{\text{soft}}\right) = \sum_{G \in \mathcal{G}_n} P_n^{\text{hard}}(G) \log\left(\frac{P_n^{\text{hard}}(G)}{P_n^{\text{soft}}(G)}\right)$$

grows slower than the number of vertices or the number of edges, depending on whether the random graph is sparse or dense.

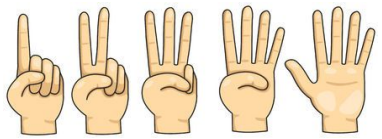
Because in both ensembles all $G \in \mathcal{G}_n$ such that $\vec{C}(G) = \vec{C}^*$ have the same probability, we get the simpler formula

$$S_n\left(P_n^{\text{hard}} \mid P_n^{\text{soft}}\right) = \log\left(\frac{P_n^{\text{hard}}(G^*)}{P_n^{\text{soft}}(G^*)}\right)$$

for any G^* such that $\vec{C}(G^*) = \vec{C}^*$. This greatly simplifies the computation, since we need not carry out the sum over $\mathcal{G}_n$ and only need to compute with a single graph G^*.

In the remainder we illustrate breaking of ensemble equivalence via examples that are tuned to graphons.

§ SUBGRAPH COUNTS



Label the simple graphs in any order:

$$F_1 \text{ edge, } F_2 \text{ wedge, } F_3 \text{ triangle, etc.}$$

Let $C_k(G)$ be the number of labelled subgraphs F_k in $G \in \mathcal{G}_n$. In the dense regime, $C_k(G)$ grows like n^{V_k}, where $V_k = |V(F_k)|$ is the number of vertices in F_k.

For $m \in \mathbb{N}$, consider the scaled vector-valued function

$$\vec{C}(G) := \left(\frac{C_k(G)}{n^{V_k-2}} \right)_{k=1}^{m} = n^2 \left(\frac{C_k(G)}{n^{V_k}} \right)_{k=1}^{m}.$$

The term $C_k(G)/n^{V_k}$ represents a subgraph density in the graph G. The additional n^2 guarantees that the full vector scales like n^2, the rate in the LDP for the Erdős-Rényi random graph, which will be needed later.

For F_k we define the homomorphism density as

$$t(F_k, G) := \frac{\mathsf{hom}(F_k, G)}{n^{V_k}} = \frac{C_k(G)}{n^{V_k}}.$$

Hence

$$(\vec{\theta}, \vec{C}(G)) = n^2 \sum_{k=1}^{m} \theta_k \, t(F_k, G) = n^2 (\vec{\theta} \cdot \vec{T}(G)),$$

where

$$\vec{T}(G) := (t(F_k, G))_{k=1}^{m} .$$

The soft ensemble with parameter $\vec{\theta}$ thus takes the form

$$P_n^{\mathsf{soft}}(G \mid \vec{\theta}) := \mathsf{e}^{n^2 \left[\vec{\theta} \cdot \vec{T}(G) - \psi_n(\vec{\theta}) \right]},$$

where ψ_n replaces the normalisation constant:

$$\psi_n(\vec{\theta}) := \frac{1}{n^2} \log \sum_{G \in \mathcal{G}_n} \mathsf{e}^{n^2 (\vec{\theta} \cdot \vec{T}(G))}.$$

In the sequel we take $\vec{\theta}$ equal to a specific value $\vec{\theta}^*$ so as to meet the soft constraint, i.e.,

$$\langle \vec{T} \rangle = \sum_{G \in \mathcal{G}_n} \vec{T}(G)\, P_n^{\mathsf{soft}}(G) = \vec{T}^*.$$

The soft ensemble becomes

$$P_n^{\mathsf{soft}}(G) = P_n^{\mathsf{soft}}(G \mid \vec{\theta}^*),$$

a notation that emphasises the fact that the constraint is controlled by $\vec{\theta}^*$, which is a function of $\vec{T}^*$.

KEY ASSUMPTION:

The constraint $\vec{T}^*$ and the Lagrange multiplier $\vec{\theta}^*$ in general depend on n, i.e., $\vec{T}^* = \vec{T}^*_n$ and $\vec{\theta}^* = \vec{\theta}^*_n$. We consider constraints that converge:

$$\lim_{n\to\infty} \vec{T}^*_n = \vec{T}^*_\infty.$$

Consequently, we expect that

$$\lim_{n\to\infty} \vec{\theta}^*_n = \vec{\theta}^*_\infty.$$

In what follows we assume that the latter hold, and drop the subscript ∞.

If convergence fails, then we may still consider convergence along subsequences.

§ BREAKING OF ENSEMBLE EQUIVALENCE

THEOREM 1:

den Hollander, Mandjes, Roccaverde, Starreveld 2018

For every subgraph constraint $\vec{T}^*$,

$$s_\infty = \lim_{n\to\infty} \binom{n}{2}^{-1} S_n\Big(P_n^{\text{hard}} \mid P_n^{\text{soft}}\Big)$$
$$= \sup_{\tilde{h}\in\tilde{W}} \Big[2\vec{\theta}^* \cdot \vec{T}(\tilde{h}) - I(\tilde{h})\Big] - \sup_{\tilde{h}\in\tilde{W}^*} \Big[2\vec{\theta}^* \cdot \vec{T}(\tilde{h}) - I(\tilde{h})\Big],$$

where $\vec{\theta}^* = \vec{\theta}^*(\vec{T}^*)$ *is the Lagrange multiplier,*

$$I(\tilde{h}) = \int_{[0,1]^2} \mathrm{d}x\mathrm{d}y\, I(h(x,y)),$$
$$I(u) = u\log u + (1-u)\log(1-u),$$

is (minus) the entropy functional and

$$\tilde{W}^* = \{\tilde{h}\in\tilde{W}\colon\ \vec{T}(\tilde{h}) = \vec{T}^*\}.$$

is the set of contrained graphons.

PROOF:

- We have

$$s_\infty = \lim_{n\to\infty} \binom{n}{2}^{-1} [\log P_n^{\text{hard}}(G^*) - \log P_n^{\text{soft}}(G^*)],$$

where G^* is any graph in $\mathcal{G}_n$ such that $\vec{T}(G^*) = \vec{T}^*$.

- For the hard ensemble, write

$$\log P_n^{\text{hard}}(G^*) = -\log \Omega_{\vec{T}^*} = -\log \left[(\tfrac{1}{2})^{\binom{n}{2}} \Omega_{\vec{T}^*} \times 2^{\binom{n}{2}} \right]$$

$$= -\log \mathbb{P}_{\frac{1}{2},n} \left(\{G \in \mathcal{G}_n \colon \vec{T}(G) = \vec{T}^*\} \right) - \binom{n}{2} \log 2,$$

where

$$\Omega_{\vec{T}^*} = |\{G \in \mathcal{G}_n \colon \vec{T}(G) = \vec{T}^*\}| > 0.$$

Consider the operator $\vec{T} \colon W \to \mathbb{R}^m$ that maps the graphon h to the subgraphs densities $(t(F_k, h))_{k=1}^m$. This operator can be extended to an operator $\vec{T}$ on $(\tilde{W}, \delta_\square)$ by defining $\vec{T}(\tilde{h}) = \vec{T}(h)$ with $h \in \tilde{h}$.

Define the sets

$$\tilde{W}^* := \left\{\tilde{h} \in \tilde{W} \colon\ T(\tilde{h}) = \vec{T}^*\right\},$$
$$\tilde{W}_n^* := \left\{\tilde{h} \in \tilde{W}^* \colon\ \tilde{h} = \tilde{h}^G \text{ for some } G \in \mathcal{G}_n\right\}.$$

Since $\vec{T}$ is Lipschitz continuous, it follows that $\tilde{W}^*$ is a closed set in $(\tilde{W}, \delta_\square)$. Since the latter is a compact space, it follows that

$$\lim_{n\to\infty} \binom{n}{2}^{-1} \log \mathbb{P}_{\frac{1}{2},n}\left(\{G \in \mathcal{G}_n \colon\ \vec{T}(G) = \vec{T}^*\}\right)$$
$$= -\inf_{\tilde{h}\in\tilde{W}^*} I_{\frac{1}{2}}(\tilde{h}) = -\inf_{\tilde{h}\in\tilde{W}^*} I(\tilde{h}) - \log 2.$$

The large deviation principle therefore yields

$$\lim_{n\to\infty} \binom{n}{2}^{-1} \log P_n^{\mathsf{hard}}(G^*) = \inf_{\tilde{h}\in\tilde{W}^*} I(\tilde{h}).$$

▶ For the soft ensemble, consider a graph G_n^* such that $\vec{T}(G_n^*) = \vec{T}^*$. We may suppose that $(G_n^*)_{n\in\mathbb{N}}$ converges to a graphon h^*. Since $\vec{T}$ is continuous, $\vec{T}(G_n^*)$ converges to $\vec{T}(h^*) = \vec{T}^*$. Hence

$$\lim_{n\to\infty} \binom{n}{2}^{-1} \log P_n^{\text{soft}}(G_n^*) = 2\vec{\theta}^* \cdot \vec{T}^* - \psi(\vec{\theta}^*),$$

where the normalisation constant equals (as shown in the computation carried out earlier)

$$\psi(\vec{\theta}^*) = \sup_{\tilde{h}\in\tilde{W}} \Big[2\vec{\theta}^* \cdot \vec{T}(\tilde{h}) - I(\tilde{h})\Big].$$

Chatterjee & Diaconis 2013

- Combining the above expressions, we get

$$s_\infty = \lim_{n\to\infty} \binom{n}{2}^{-1} [\log P_n^{\text{hard}}(G_n^*) - \log P_n^{\text{soft}}(G_n^*)]$$

$$= \inf_{\tilde{h}\in\tilde{W}^*} I(\tilde{h}) - 2\vec{\theta}^* \cdot \vec{T}^* + \sup_{\tilde{h}\in\tilde{W}} \Big[2\vec{\theta}^* \cdot \vec{T}(\tilde{h}) - I(\tilde{h})\Big].$$

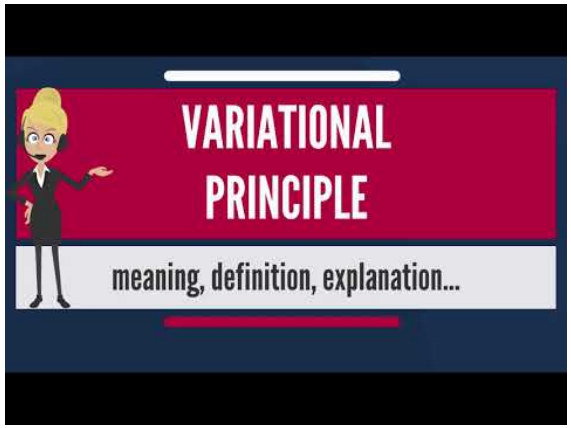
By definition, all $\tilde{h} \in \tilde{W}^*$ satisfy $\vec{T}(\tilde{h}) = \vec{T}^*$. Hence the expression in the right-hand side can be written as

$$\sup_{\tilde{h}\in\tilde{W}} \Big[2\vec{\theta}^* \cdot \vec{T}(\tilde{h}) - I(\tilde{h})\Big] - \sup_{\tilde{h}\in\tilde{W}^*} \Big[2\vec{\theta}^* \cdot \vec{T}(\tilde{h}) - I(\tilde{h})\Big],$$

which settles the claim. □

REMARK: We have obtained a variational characterisation of ensemble equivalence:

$s_\infty = 0$ if and only if at least one of the maximisers of $2\vec{\theta}^* \cdot \vec{T}(\tilde{h}) - I(\tilde{h})$ in $\tilde{W}$ also lies in $\tilde{W}^* \subset \tilde{W}$.



§ EDGE-TRIANGLE MODEL

As an example we pick the two-fold constraint

$$\begin{aligned}\vec{C}^* &= (\text{number of edges}, \text{number of triangles}) \\ &= \left(T_1^* \binom{n}{2}, T_2^* \binom{n}{3}\right)\end{aligned}$$

with

$$(T_1^*, T_2^*) = \vec{T}^* \in [0,1]^2.$$

The quantity of interest is

$$s_\infty = \lim_{n\to\infty} \binom{n}{2}^{-1} S_n \left(P_n^{\text{hard}} \mid P_n^{\text{soft}}\right).$$

The corresponding operator on $\tilde{W}$ is

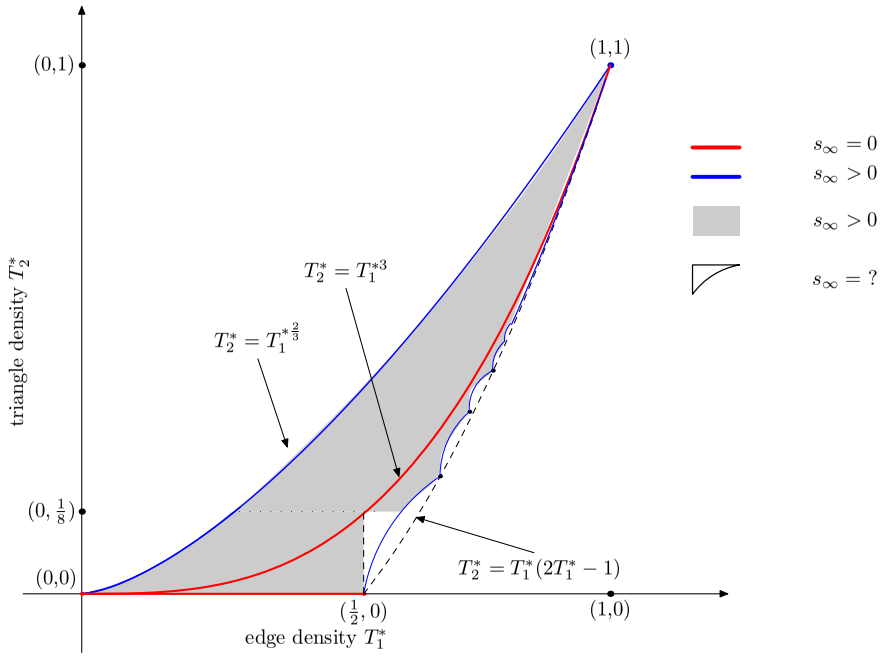
$$\vec{T}(\tilde{h}) = (T_1(\tilde{h}), T_2(\tilde{h}))$$

with

$$T_1(\tilde{h}) = \int_{[0,1]} \mathrm{d}x\, \mathrm{d}y\, h(x,y),$$
$$T_2(\tilde{h}) = \int_{[0,1]} \mathrm{d}x\, \mathrm{d}y\, \mathrm{d}z\, h(x,y)h(y,z)h(z,x),$$

THEOREM 2:

den Hollander, Mandjes, Roccaverde, Starreveld 2018



Between the blue curves the edge-triangle densities are admissible.

Radin, Sadun 2015

Breaking of ensemble equivalence occurs as soon as the constraints are frustrated.

WEAK FRUSTRATION

What happens close the line $T_2^* = T_1^{*3}$? It turns out that anomalous behaviour shows up:

THEOREM 3:

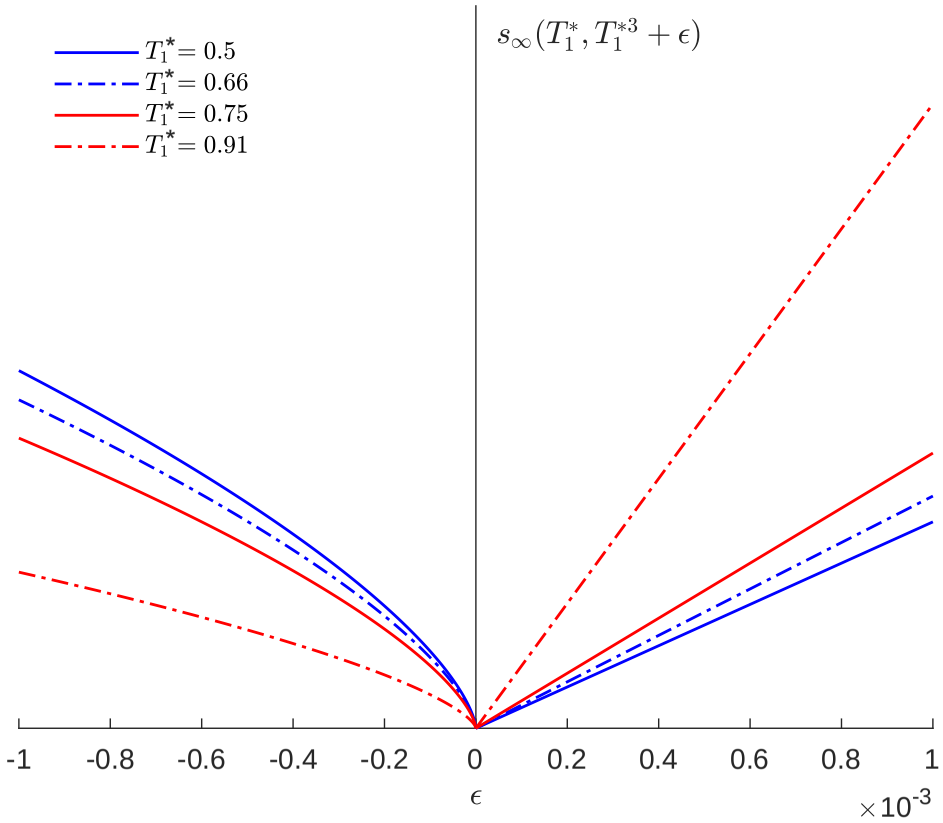
den Hollander, Mandjes, Roccaverde, Starreveld 2020

For $T_1^* \in (0,1)$,

$$\lim_{\epsilon \downarrow 0} \epsilon^{-1} \, s_\infty(T_1^*, T_1^{*3} + \epsilon) = C^+(T_1^*) \in (0,\infty),$$

$$\lim_{\epsilon \downarrow 0} \epsilon^{-2/3} \, s_\infty(T_1^*, T_1^{*3} - \epsilon) = C^-(T_1^*) \in (0,\infty),$$

where $C^+(T_1^*), C^-(T_1^*)$ *are computable functions of* T_1^*.



More details in the lecture by Nicos Starreveld!

§ CONCLUSION

- Care needs to be taken with the way in which the a priori information that is available about the network is used to choose the proper randomness.
- Hard constraints or soft constraints may lead to very different behaviour, even for very large networks.
- It turns out that breaking of ensemble equivalence is the rule rather than the exception for natural classes of constraints that are frustrated.
- The approach based on ensembles provides a flexible framework.


AAAHHHH!