

# MATHEMATICAL METHODS AND NETWORKS

Frank den Hollander  
Mathematical Institute, Leiden University



*Networks Matchmaking Event,*  
Doorn, The Netherlands, 20–22 January 2020

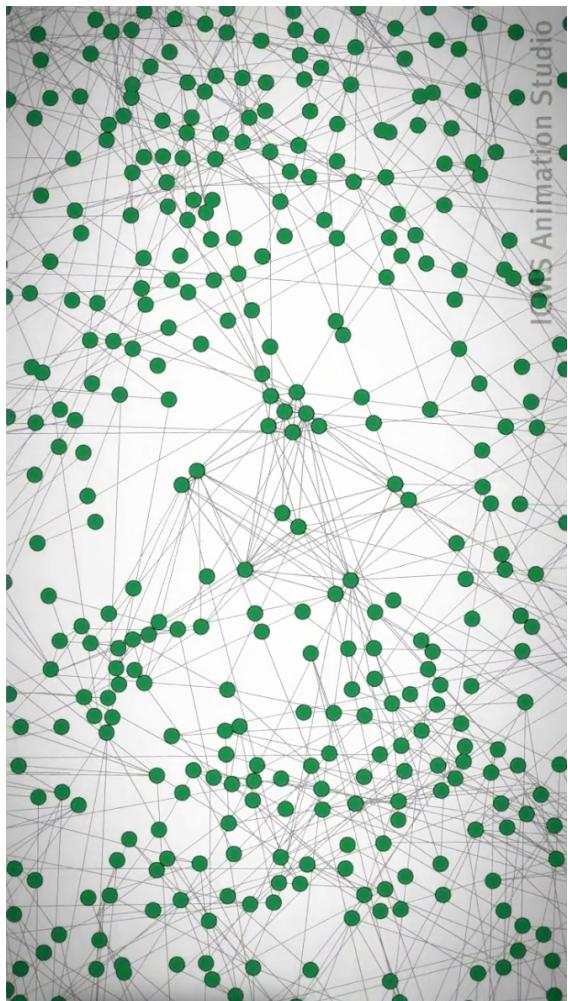
## § CONSTRAINED NETWORKS

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

Networks can be modelled as graphs, consisting of nodes connected by links. Large networks should be modelled as random graphs, where the links 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  $N \rightarrow \infty$ .



A realisation of a **large random graph**

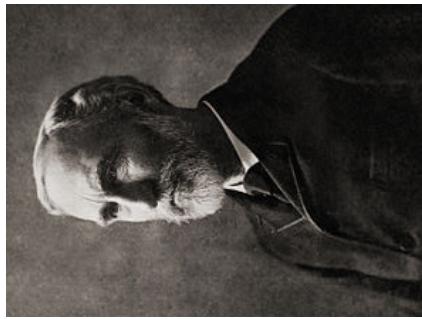
Typically, some a priori information is available about the network, e.g. the degrees of the nodes or the total number of edges and triangles. How should this a priori information be properly incorporated when we choose the probability distribution according to which we build the random graph?

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

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

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

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  $\mathcal{N}(\vec{\theta}^*)$  is a constant that **normalises**.

## INTERPRETATION

- $P_N^{\text{hard}}$  models a random graph of which no information is available other than the constraint.
- $P_N^{\text{soft}}$  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

Tourette

$P_N^{\text{hard}}$  and  $P_N^{\text{soft}}$  are said to be equivalent when their relative entropy defined by

$$s_N(P_N^{\text{hard}} \mid P_N^{\text{soft}}) = \sum_{G \in \mathcal{G}_N} P_N^{\text{hard}}(G) \log \left( \frac{P_N^{\text{hard}}(G)}{P_N^{\text{soft}}(G)} \right)$$

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**

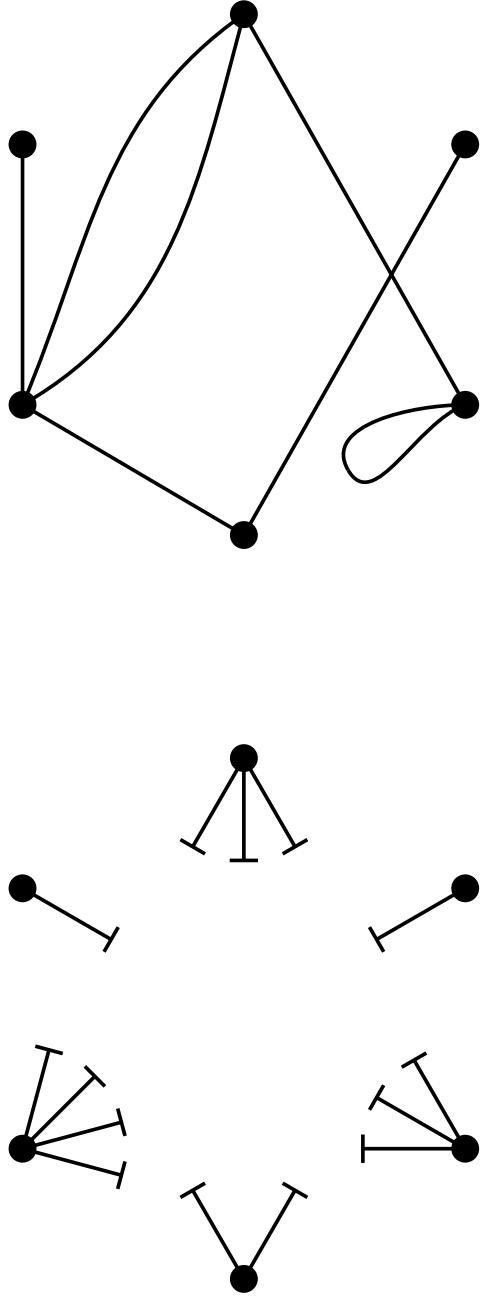
$$S_N(P_N^{\text{hard}} \mid P_N^{\text{soft}}) = \log \left( \frac{P_N^{\text{hard}}(G^*)}{P_N^{\text{soft}}(G^*)} \right)$$

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 two examples.

## § CONSTRAINT ON THE DEGREE SEQUENCE

Each vertex gets a prescribed number of **half-edges**, which are **paired off randomly** to form edges.



Example with  $N = 6$  and  $\vec{d}_N = (1, 3, 1, 3, 2, 4)$

All the vertices have **prescribed degrees**. In other words, the constraint is

$$\vec{C}^* = (d_1^*, \dots, d_N^*).$$

Suppose that the degrees are *moderate*, i.e.,

$$\max_{1 \leq i \leq N} d_i^* = o(\sqrt{N}), \quad N \rightarrow \infty.$$

sparseness

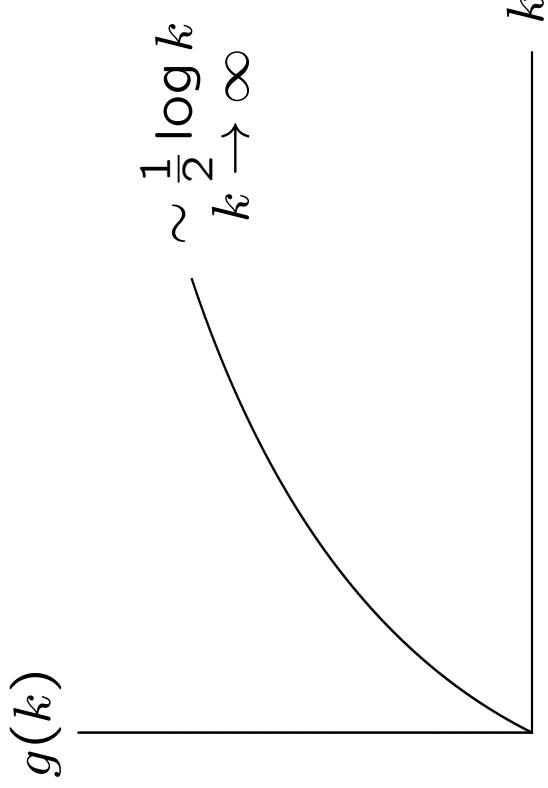
For  $k = 0, 1, 2 \dots$  let

$$f_N(k) = \frac{1}{N} \sum_{i=1}^N \mathbf{1}_{\{d_i^* = k\}}$$

= fraction of vertices with degree  $k$

and define

$$g(k) = \log \left( \frac{k!}{k^k e^{-k}} \right).$$



*Suppose that*

$$\lim_{N \rightarrow \infty} f_N = f \quad \text{in } L^1(g)$$

*for some limiting function  $f$ . Then*

$$s_\infty = \sum_{k=0}^{\infty} f(k)g(k).$$

Garlaschelli, den Hollander, Roccaverde, Squartini

**There is breaking of ensemble equivalence for all  $f \not\equiv 0$ .**



The above says that, in the limit as  $N \rightarrow \infty$ ,

- hard-constraint ensemble:  
vertices have a **fixed degree**.
- soft-constraint ensemble:  
vertices have a **random degree**.

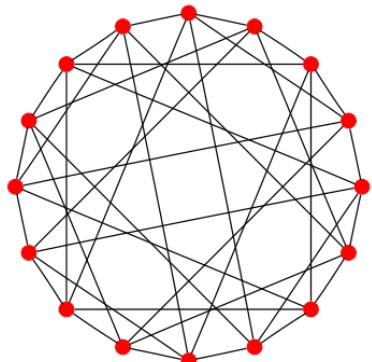
Breaking of ensemble equivalence occurs as soon as the number of constraints is **proportional** to the number of nodes.

**Example 1:**

$f_N(k) = 1_{\{k=\ell\}}$  with  $\ell = o(\sqrt{N})$ .

For  $k$ -regular graphs:

$$s_\infty = g(\ell) > 0.$$



5-regular graph

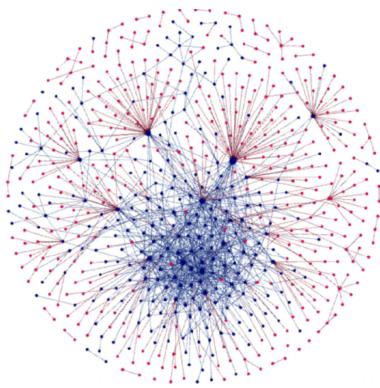
### Example 2:

$$f_N(k) = C_N k^{-\tau}, \quad 1 \leq k = o(\sqrt{N}),$$

with  $\tau \in (1, \infty)$  a tail exponent.

For scale-free graphs:

$$s_\infty \approx \frac{1}{2(\tau - 1)}.$$



graph with hubs

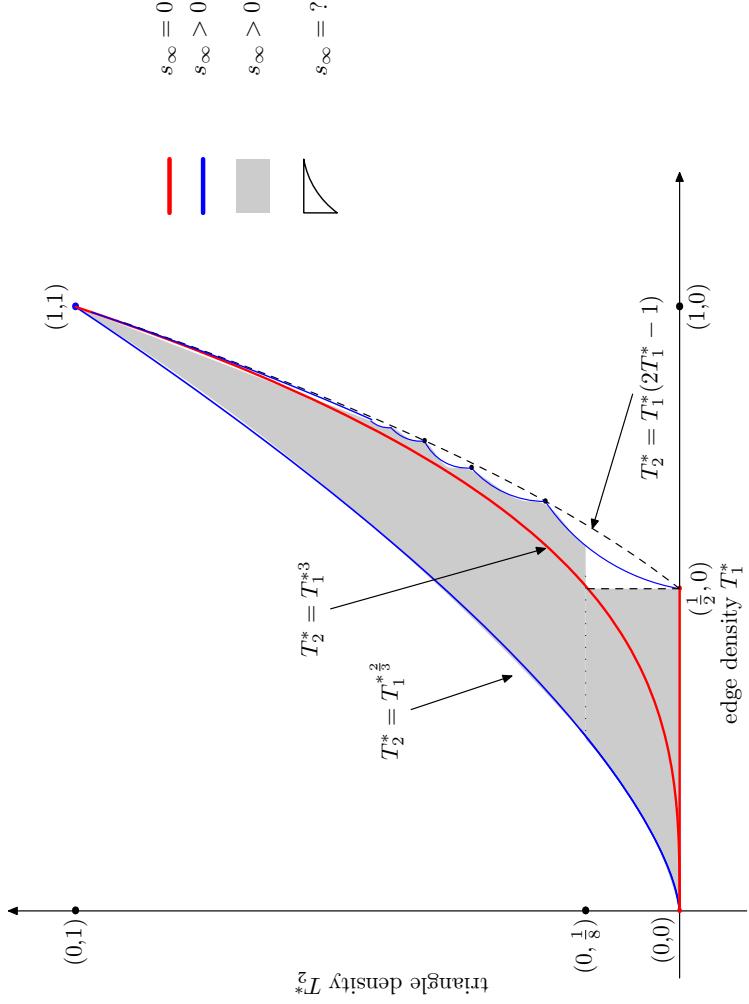
## § CONSTRAINT ON THE TOTAL NUMBER OF EDGES AND TRIANGLES

For the constraint we pick

$$\begin{aligned}\vec{C}^* &= (\text{number of edges}, \text{number of triangles}) \\ &= \left( T_1^* \binom{N}{2}, T_2^* \binom{N}{3} \right), \quad T_1^*, T_2^* \in (0, 1).\end{aligned}$$

This corresponds to the **dense regime**, in which the number of edges is of order  $N^2$ . The quantity of interest is

$$s_\infty = \lim_{N \rightarrow \infty} \frac{1}{N^2} \log S_N \left( P_N^{\text{hard}} \mid P_N^{\text{soft}} \right).$$



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

den Hollander, Mandjes, Roccaverde, Starreveld

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

## § 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.
- The approach based on ensembles provides a flexible framework.

