Random Connection Hypergraphs

Morten Brun, Christian Hirsch, Peter Juhasz, Moritz Otto

AARHUS UNIVERSITET

peter.juhasz@math.au.dk

Collaboration Networks

- Higher-order networks describe group interactions.
- Goal: understand real-world higher-order networks by higher-order network models.
- The models could shed light on the high-level structure of scientific collaborations.

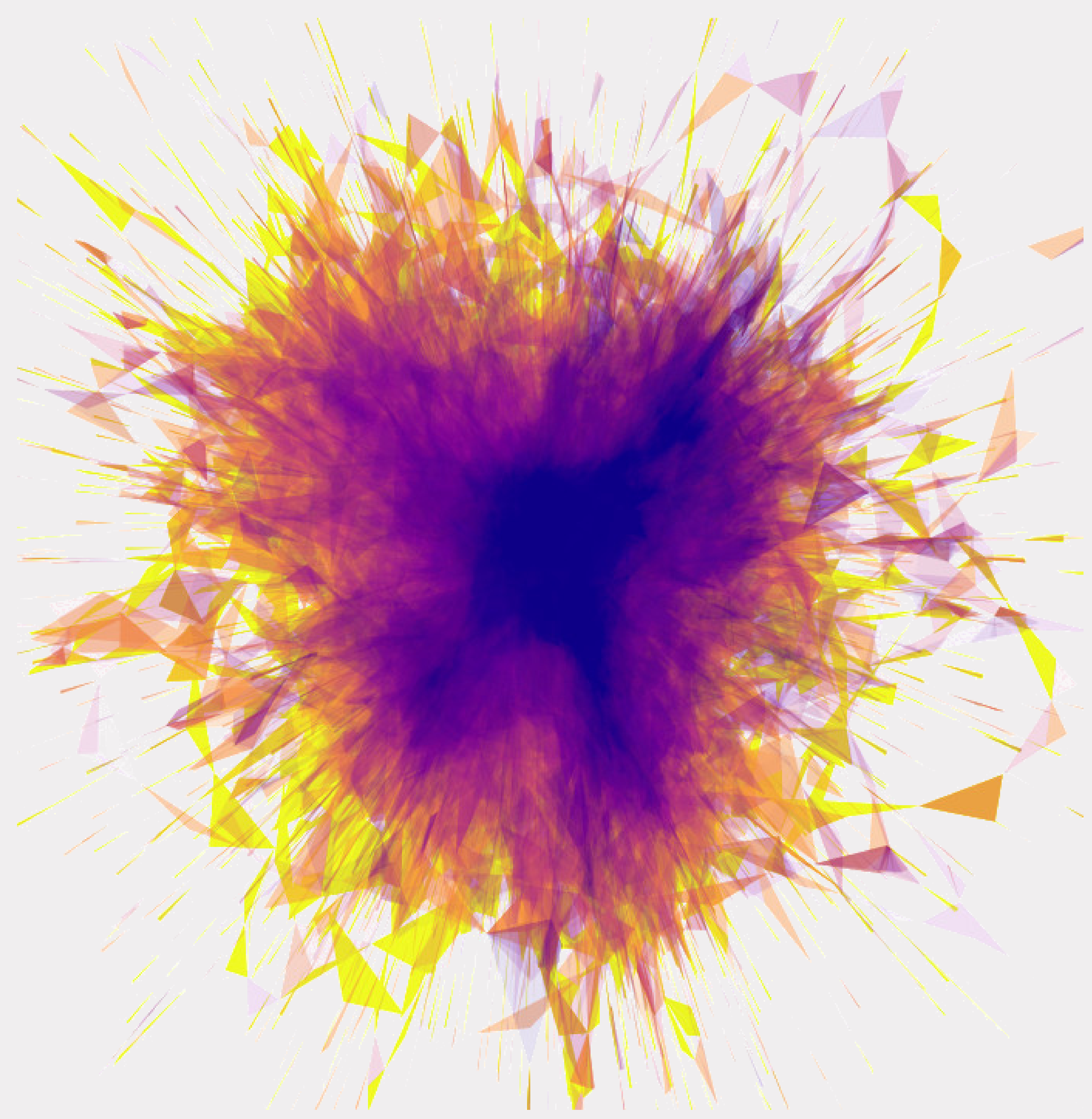

Figure 1: Collaboration of scientists in the field of engineering

- Idea: use simplicial complexes where each interaction is represented by a simplex.
- Explaining simplex counts in various dimensions describes high-level community structures.

Random Connection Hypergraph Model

- Hypergraph model: bipartite graph with two marked Poisson point processes $\mathcal{P}$ and $\mathcal{P}'$ on $\mathbb{R} \times [0, 1]$:
  - Vertices: $\mathcal{P} = \{(x_i, u_i)\}_{i \geq 1}$
  - Hyperedges: $\mathcal{P}' = \{(y_i, v_i)\}_{i \geq 1}$
- A vertex–hyperedge pair $\{(x, u), (y, v)\} \in \mathcal{P} \times \mathcal{P}'$ is connected iff

$$|x - y| \leq \beta\, u^{-\gamma} v^{-\gamma'} \qquad \beta > 0,\ \gamma, \gamma' \in (0, 1)$$

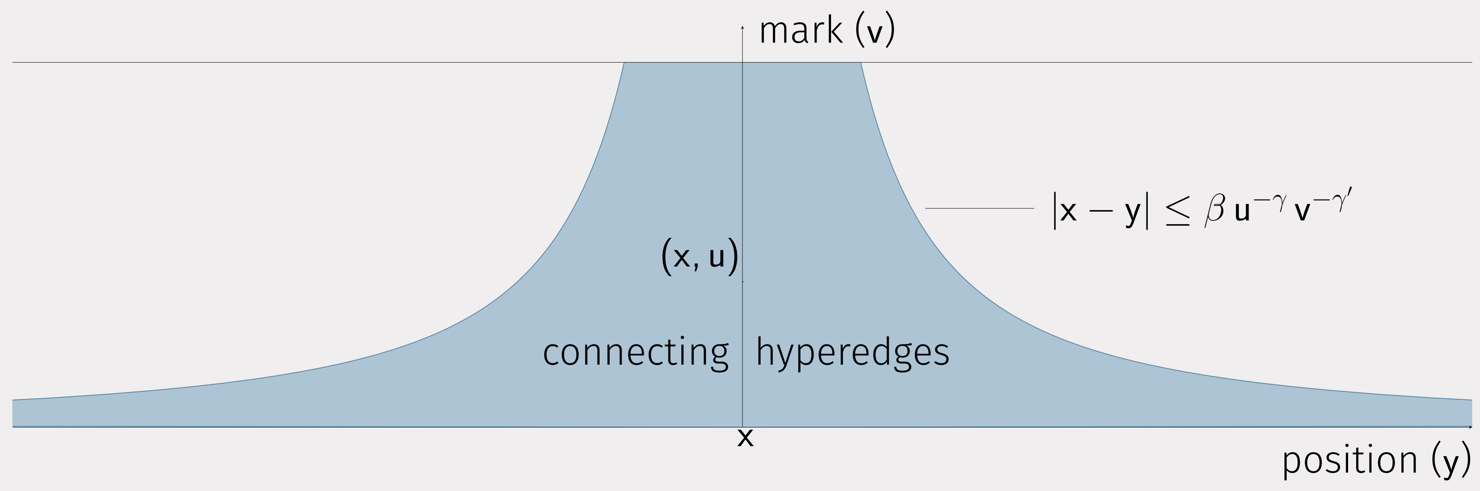

Figure 3: Connection condition

- Result: preferential attachment + spatially induced clustering.

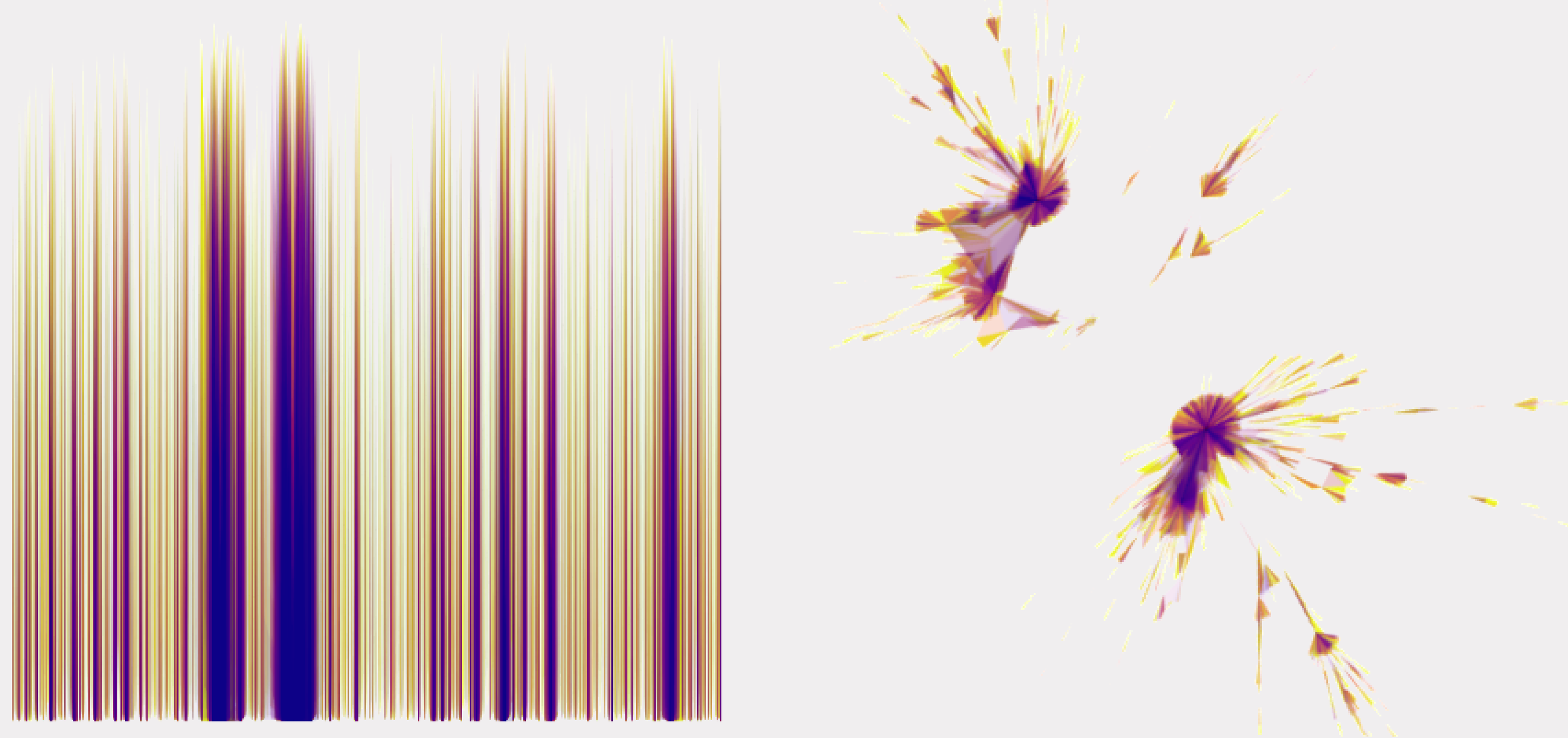

Figure 4: Hypergraph generated by the random connection hypergraph model. **Left**: Fixed vertex positions. **Right**: Force directed layout.

Stable Distribution of Simplex Counts

Theorem (Stable limit for simplex counts). *Let $S_{n,m}$ be the m-simplex count in the window $[0, n]$. If $\gamma > 1/2$, and $\gamma' < 1/(2m + 1)$, then*

$$n^{-\gamma}(S_{n,m} - \mathbb{E}[S_{n,m}]) \xrightarrow[n\uparrow\infty]{d} \mathcal{S}(1/\gamma).$$

- For $\gamma = 0.75$, $\gamma' = 0.1$ we simulated 100 networks of size $n = 10^5$.
- The Q-Q plot for the triangle counts shows a fat right tail of the distribution.

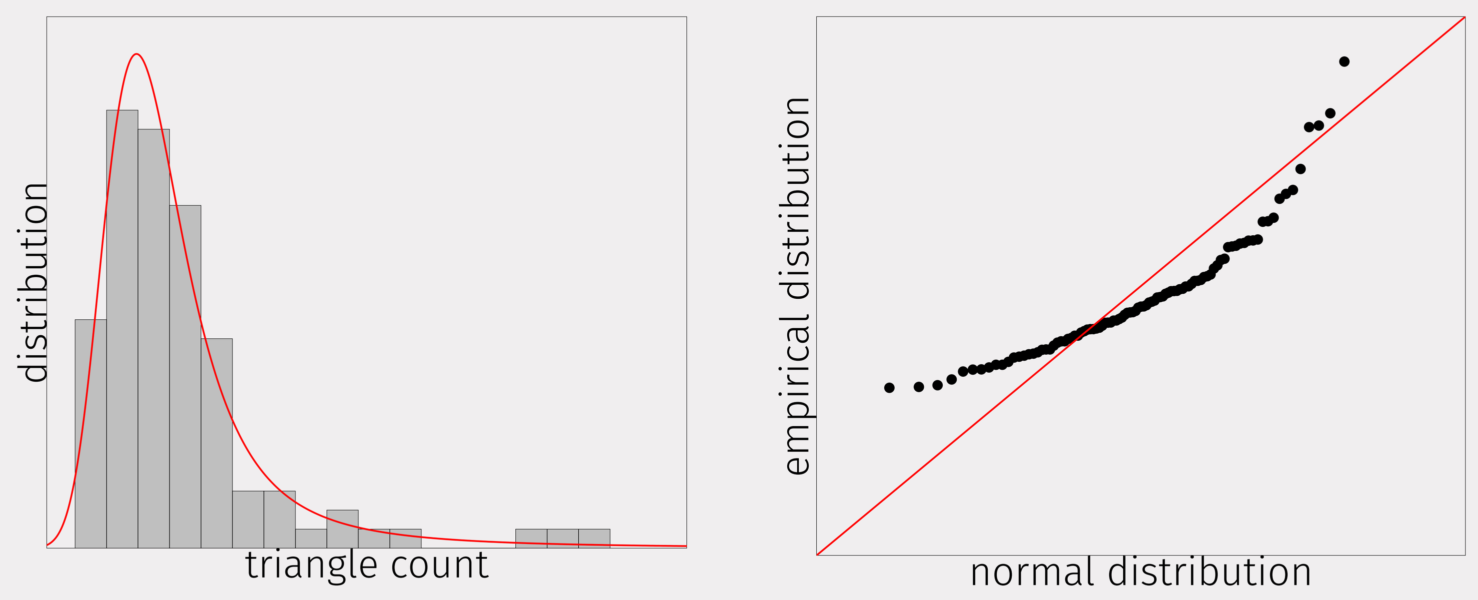

Figure 2: Left: stable distribution of the triangle counts. **Right**: Q-Q plot with the fitted normal distribution.

Hypothesis Test

- Model parameters β, γ, γ' are fitted to the empirical dataset.
- Parameters of a stable distribution are fitted to the simplex counts generated by the model.
- Hypothesis test is performed to determine if the model captures the triangle count of the dataset.

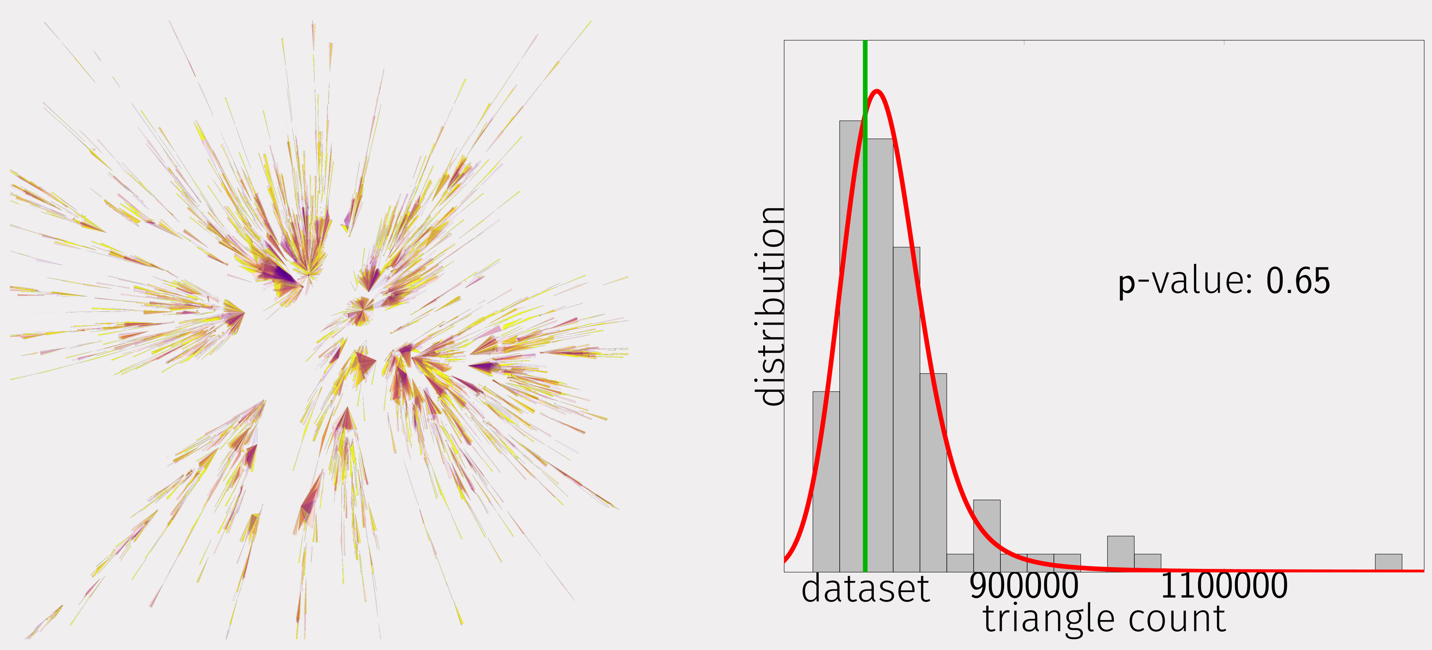

Figure 5: Left: Sample network with the fitted model parameters. **Right**: Hypothesis test for the triangle count.

This work was supported by the Danish Data Science Academy, which is funded by the Novo Nordisk Foundation (NNF21SA0069429) and VILLUM FONDEN (40516).