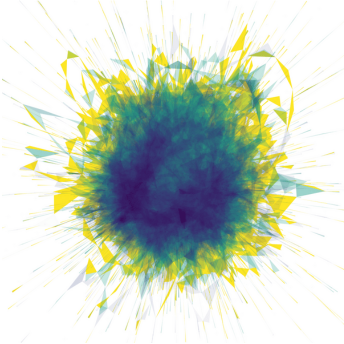
Random Connection Hypergraphs

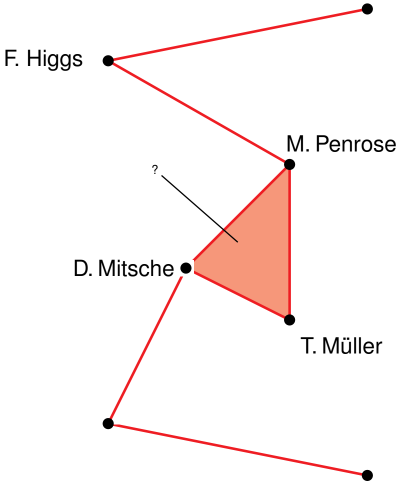
Christian Hirsch

joint work with M. Brun, P. Juhász & M. Otto

Stochastic Geometry in Action, University of Bath, Sep 12, 2024

Collaboration of arXiv math.stat

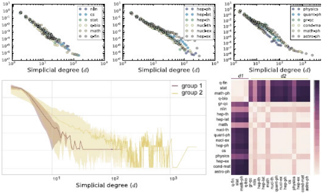



F. Higgs
M. Penrose
?
D. Mitsche
T. Müller

(Petri, Scolamiero, Donato &Vaccarino, 2013)

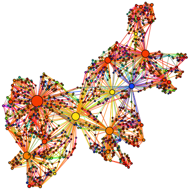
(Patania, Petri, & Vaccarino, 2017)

- ▷ ***Topological data analysis*** of coauthor networks
- ▷ Empirical investigation



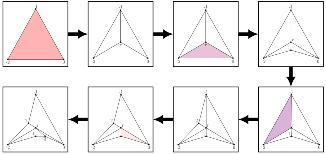
(Bianconi & Rahmede, 2016)

- ▷ ***Network geometry with flavor***
- ▷ Difficulty creating non-trivial homology



(Fountoulakis, Iyer, Mailler & Sulzbach, 2022)

- ▷ ***General model of random simplicial complexes***
- ▷ ***Degree distribution***
- ▷ Higher-order quantities?

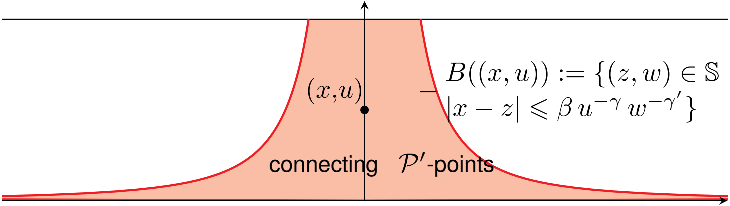


- Develop spatial model for higher-order connectivities
- Define ***bipartite random connection model*** between authors and papers
- Incorporate ***weights*** to realize scale-freeness
- Combines ***weighted RCMs*** and ***AB random geometric graphs***

Model definition

- $\mathcal{P}, \mathcal{P}' :=$ Poisson point processes on $\mathbb{R} \times [0,1]$ with intensities λ, λ'
- $(X,U) \in \mathcal{P}$, $(Z,W) \in \mathcal{P}'$ are connected by edge iff

$$|X - Z| \leqslant \beta U^{-\gamma} W^{-\gamma'}, \quad \gamma, \gamma' \in (0,1), \beta > 0$$



- Define hypergraph $G(\mathcal{P}, \mathcal{P}')$ with set of ***hyperedges*** of cardinality $m+1$ as

$$\Sigma_m := \Big\{\Delta \subseteq \mathcal{P} \colon \#\Delta = m+1, \mathcal{P}' \cap \bigcap_{P \in \Delta} B(P) \neq \emptyset\Big\}$$

Age-dependent random connection models

(Gracar, Grauer, Lüchtrath & Mörters, 2019)

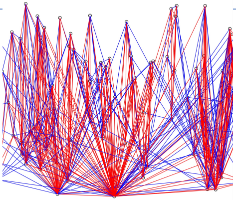
- ▷ Degree distribution, clustering coefficient

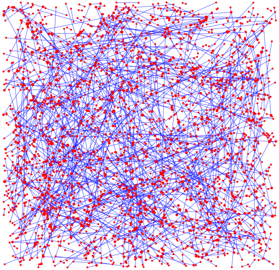
(van der Hofstad, van der Hoorn & Maitra, 2022)

- ▷ ***Scaling of clustering*** conditioned on typical degree
- ▷ Results for interpolation kernel
- ▷ ***No results on bipartite setting***

(van der Hofstad et. al., 2020)

- ▷ ***Alpha-stable limit*** of clustering coefficient
- ▷ ***No geometry***





https://petergracar.github.io/

- Special case, where $\gamma = \gamma' = 0$
- Motivated by applications in telecommunication networks.

(Iyer & Yogeshwaran, 2012)

- Investigation of percolation properties
- Special focus $d = 2$

(Penrose, 2014)

- Percolation properties in general dimension
- Upper bounds on critical intensities

(Dereudre & Penrose, 2018)

- Lower bounds on critical intensities

$\rightsquigarrow$ ***No heavy-tailed degree distribution***

Idea. Generalize degree to higher-order adjacencies

- Definition of ***typical** m**-simplex*** Δ_m^* via ***Palm theory***

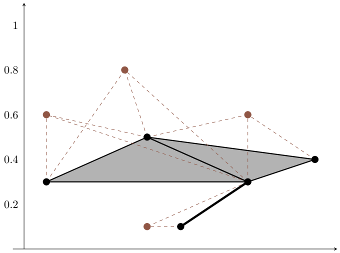
$$\mathbb{E}\big[f(\Delta_m^*, \mathcal{P}, \mathcal{P}')\big] = \frac{1}{\lambda_m}\mathbb{E}\Big[\sum_{\substack{\Delta_m \in \Sigma_m \\ c(\Delta_m)\in[0,1]}} f(\Delta_m - c(\Delta_m), \mathcal{P} - c(\Delta_m), \mathcal{P}' - c(\Delta_m))\Big]$$

- $\deg(\Delta_m^*) := \mathcal{P}'\big(\Delta_m^*\big) := \mathcal{P}'\big(\bigcap_{p\in\Delta_m^*} B(p)\big)$
- $d_{m,k} := \mathbb{P}(\deg(\Delta_m^*) \geqslant k) =$ tail of typical simplex degree
- $d_{1,k} =$ typical node-degree

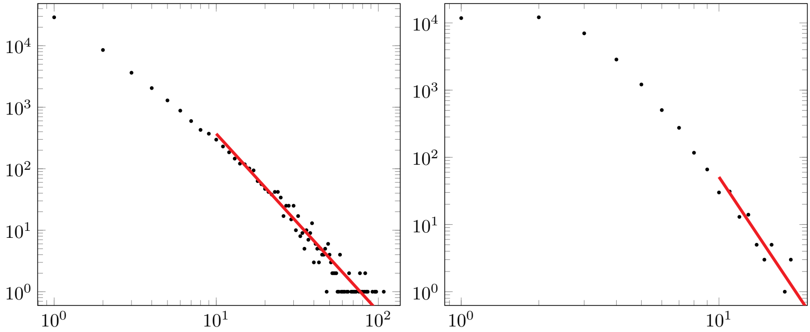
Theorem (Power-law degrees)

Let $\gamma' < 1/(m+1)$*. Then,*

$$\lim_{k\uparrow\infty} \log(d_{m,k})/\log(k) = m - (m+1)/\gamma$$



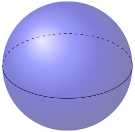
Vertex degrees

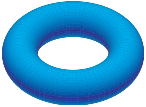


Δ_0 degree (left) and Δ_0' degree (right) in `math.stat`

Is the model a good fit with the data?

- The considered hypergraph is a ***simplicial complex***
- ***Betti numbers*** as test statistic





- $\beta_0 =$ number of components
- $\beta_1 =$ number of loops
- $\beta_2 =$ number of cavities

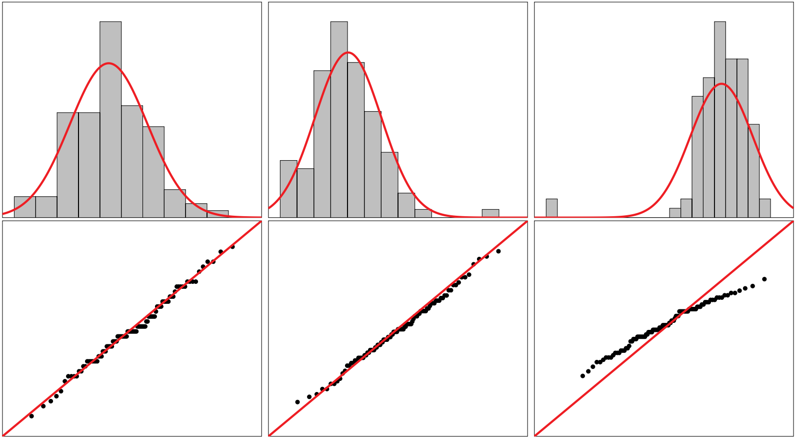
- $\mathcal{P}_n := \mathcal{P} \cap [0, n] :=$ Poisson process on $[0, n]$
- $\beta_{n,q} :=$ qth Betti number of $G_n := G\big(\mathcal{P}_n, \mathcal{P}'_n\big)$.

Theorem (CLT for Betti numbers)

Let $q \geqslant 0$, $\gamma < 1/4$ *and* $\gamma' < 1/(4(m+1))$. *Then,*

$$\frac{\beta_{n,q} - \mathbb{E}[\beta_{n,q}]}{\sqrt{\mathrm{Var}(\beta_{n,q})}} \xrightarrow{d} \mathcal{N}(0,1).$$

Betti numbers, $\gamma' = 0.1$



$\gamma = 0.25$ (left), $\gamma = 0.5$ (middle), $\gamma = 0.75$ (right)

What if $\gamma > 1/2$?

Conjecture: Alpha-stable limit

Let $q \geqslant 0$, $\gamma \in (1/2, 1)$, and γ' ***sufficiently small***. Then,

$$n^{-\gamma}(\beta_{n,q} - \mathbb{E}[\beta_{n,q}]) \xrightarrow{d} \mathcal{S}(1/\gamma),$$

where $\mathcal{S}(1/\gamma)$ is a $(1/\gamma)$-stable random variable

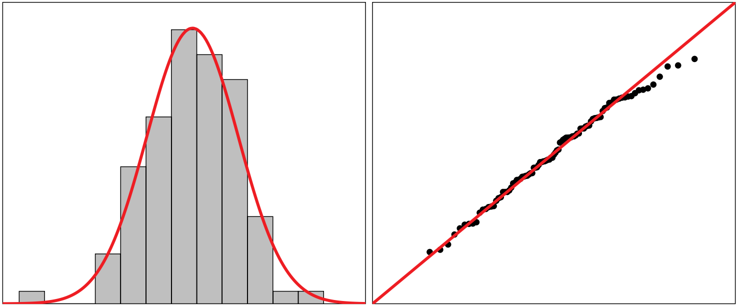
- $S_n :=$ Edge count of bipartite graph in $[0, n]$

Theorem (Limit law for the edge count)

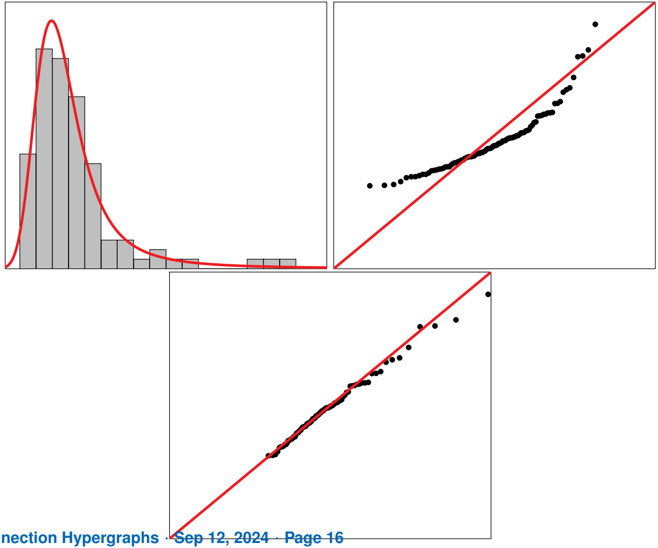
- *Let* $\gamma \in (1/2, 1), \gamma' < 1/3$. *Then,* $n^{-\gamma}(S_n - \mathbb{E}[S_n]) \xrightarrow{d} \mathcal{S}(1/\gamma)$
- *Let* $\gamma < 1/2, \gamma' < 1/3$. *Then,* $n^{-1/2}(S_n - \mathbb{E}[S_n]) \xrightarrow{d} \mathcal{N}(0, \sigma^2)$

Extends to simplex count.

$\gamma = 0.25, \gamma' = 0.10$: QQ-plot with normal distribution



$\gamma = 0.75, \gamma' = 0.10$: Stable Distribution



Step 1. Palm representation of typical m-simplex Δ_m^*.

- Represent Δ_m^* as $\boldsymbol{p}_m(u) := \big((0,u), \boldsymbol{p}_m\big) = \big((0,u), o_1, \dots, o_m\big)$
- $g_m(\boldsymbol{p}_m(u), \mathcal{P}') :=$ Indicator of event that $\boldsymbol{p}_m(u)$ forms an m-simplex $u \leqslant u_1 \leqslant \cdots \leqslant u_m$
- $\mathbb{S} := \mathbb{R} \times [0,1] :=$ Space-time domain

Palm representation

$$\mathbb{E}[f(\Delta_m^*, \mathcal{P}, \mathcal{P}')] = \frac{\int_0^1 \int_{\mathbb{S}^m} \mathbb{E}[f(\boldsymbol{p}_m(u), \mathcal{P}, \mathcal{P}')] g_m(\boldsymbol{p}_m(u), \mathcal{P}') \mathrm{d}\boldsymbol{p}_m \mathrm{d}u}{\int_0^1 \int_{\mathbb{S}^m} \mathbb{E}[g_m(\boldsymbol{p}_m(u), \mathcal{P}')] \mathrm{d}\boldsymbol{p}_m \mathrm{d}u}$$

for any bounded measurable f.

Proof idea.

- Application of ***Mecke formula.***
- Denominator finite?

Finiteness of denominator.

- ▷ $\mu_m(u) := \int_{\mathbb{S}^m} \mathbb{E}[g_m(\boldsymbol{p}_m(u), \mathcal{P}')]\mathrm{d}\boldsymbol{p}_m \mathrm{d}u.$
- ▷ $\mu_0(u) \leqslant$ Expected typical degree; $I_r(x) := [-r + x, r + x]$. That is,

$$\mu(u) := \mu_0(u) \leqslant \int_u^1 |I_{\beta u^{-\gamma} w^{-\gamma'}}| \mathrm{d}w = \frac{2\beta u^{-\gamma}}{1-\gamma'}(1 - u^{1-\gamma'}).$$

Sketch for general case.

$$\int_{\mathbb{S}} \mathbb{1}\{|z| \leqslant \beta u^{-\gamma} w^{-\gamma'}\} \int_{\mathbb{S}^m} \mathbb{1}\{|z - y_i| \leqslant \beta v_i^{-\gamma} w^{-\gamma'} : 1 \leqslant i \leqslant m\} \mathrm{d}\boldsymbol{p}_m \mathrm{d}p'.$$

Next, we integrate the spatial coordinates $\mathbf{y}$, the marks $\mathbf{v}$ as well as z, w, u

$$\begin{aligned}
&\int_{\mathbb{S}} \mathbb{1}\{|z| \leqslant \beta u^{-\gamma} w^{-\gamma'}\} \int_{[0,1]^m} \prod_{i=1}^m (2\beta v_i^{-\gamma} w^{-\gamma'}) \mathrm{d}\mathbf{v} \mathrm{d}p' \\
&= \left(\frac{2\beta}{1-\gamma}\right)^m \int_{\mathbb{S}} \mathbb{1}\{|z| \leqslant \beta u^{-\gamma} w^{-\gamma'}\} w^{-m\gamma'} \mathrm{d}p' \\
&= \frac{(2\beta)^{m+1}}{(1-\gamma)^m} u^{-\gamma} \int_0^1 w^{-(m+1)\gamma'} \mathrm{d}w = \frac{(2\beta)^{m+1}}{(1-\gamma)^m} \frac{u^{-\gamma}}{1-(m+1)\gamma'},
\end{aligned}$$

Idea. Build specific configurations leading to high simplex degrees.

- ▷ $R_k := [0, k] \times [0, (\beta/k)^{1/\gamma}] \subseteq \mathbb{S}$

- ▷ Any $p := (y, v) \in R_k$ and $p' := (z, w) \in [0, k] \times [0, 1]$ are connected.
- ▷ This gives for $\nu(k) := \lambda' k - \log(2)$,

$$\mathbb{P}(\deg(\Delta_m^*) \geqslant \nu(k)) \geqslant \frac{\lambda^{m+1}}{\lambda_m (m+1)!} \int_{R_k^m} \int_0^{(\beta/k)^{1/\gamma}} \mathbb{P}(\deg(\boldsymbol{p}_m(u)) \geqslant \nu(k)) \mathrm{d}u \mathrm{d}\mathbf{p}_m$$

$$\geqslant \frac{\lambda^{m+1} (\beta/k)^{1/\gamma}}{\lambda_m (m+1)!} \int_{R_k^m} \mathbb{P}(\mathcal{P}'([0, k] \times [0, 1]) \geqslant \nu(k)) \mathrm{d}\mathbf{p}_m,$$

Median of $\mathcal{P}'([0, k] \times [0, 1]) \sim \mathrm{Poi}(\lambda' k)$ bounded below by $\nu(k)$

Next, bound $\int_{[0,1]}\int_{\mathbb{S}^m} \mathbb{P}(\mathcal{P}'(B(\boldsymbol{p}_m(u))) \geqslant k)\mathrm{d}\boldsymbol{p}_m\mathrm{d}u$

- $X := \mathcal{P}'(B(\boldsymbol{p}_m(u)))$ is Poisson with parameter $b := |B(\boldsymbol{p}_m(u))|$

$$\mathbb{P}(X \geqslant k) \leqslant \mathbb{1}\{\lambda' b \geqslant k/2\} + \mathbb{P}(X \geqslant k)\mathbb{1}\{\lambda' b < k/2\}$$

- Poisson concentration bound the probability

Idea. Consider ***pairwise*** intersections under the order $u \leqslant v_1 \leqslant \cdots \leqslant v_m$

$$b \leqslant |B(\boldsymbol{p}_1(u))| \wedge \min_{i=1,\ldots,m-1} |B(\{p_i, p_{i+1}\})|.$$

Lemma (Pairwise intersections)

Let $0 \leqslant u \leqslant v \leqslant 1$, *and set* $o := (0, u) \in \mathbb{S}$ *and* $p := (y, v) \in \mathbb{S}$. *Then,*

$$|B(\{o, p\})| \leqslant \frac{2\beta}{1-\gamma'} v^{-\gamma} s_\wedge(u, y)^{1-\gamma'},$$

where $s_\wedge(u, y) := (2\beta u^{-\gamma}|y|^{-1})^{1/\gamma'} \wedge 1$.

Idea.

- Proceed as in (Shirai & Hiraoka, 2018) for Gilbert graph
- Apply ***stabilization technique*** from (Penrose & Yukich, 2003)
- $\beta(\mathcal{P}_n, \mathcal{P}'_n) := \beta_{n,m}(\mathcal{P}_n, \mathcal{P}'_n) = m$th Betti number of the hypergraph G_n
- Introduce the typical points $\boldsymbol{x} := (x, u), \boldsymbol{x}' := (x', w)$.
- ***Key tasks*** Stabilization conditions
  - $\lim_{n\to\infty} \beta(\mathcal{P}_n \cup \{\boldsymbol{x}\}, \mathcal{P}'_n) - \beta(\mathcal{P}_n \cup \{\boldsymbol{x}\}, \mathcal{P}'_n) < \infty$
  - $\lim_{n\to\infty} \beta(\mathcal{P}_n, \mathcal{P}'_n \cup \{\boldsymbol{x}'\}) - \beta(\mathcal{P}_n, \mathcal{P}'_n \cup \{\boldsymbol{x}'\}) < \infty$
- Also need to check ***moment conditions***

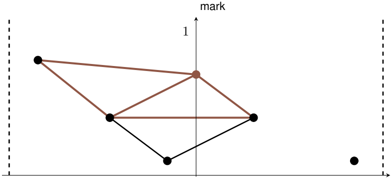
Idea. "holes = cycles - boundaries":

$$\beta_m(G_n) = \mathrm{dim}(Z(G_n)) - \mathrm{dim}(B(G_n))$$

Idea: show weak stabilization for cycles and boundaries separately.

- Boundedness: $\mathrm{dim}(Z_{n,\boldsymbol{x}}) - \mathrm{dim}(Z_n)$ is bounded by number of new m-simplices containing $\boldsymbol{x}$
- Monotonicity

- Let $n_1 \leqslant n_2$, and consider the canonical map $Z_{n_1,\boldsymbol{x}} \to Z_{n_2,\boldsymbol{x}}/Z_{n_2}$



- Claim: kernel $Z_{n_1,\boldsymbol{x}} \cap Z_{n_2} = Z_{n_1}$
  - $\Delta_m \in \Sigma_{n_1} \implies \Delta_m \in \Sigma_{n_1,o} \cap \Sigma_{n_2} \implies Z_{n_1} \subseteq Z_{n_1,\boldsymbol{x}} \cap Z_{n_2}$
  - $Z_{n_1} \supseteq Z_{n_1,\boldsymbol{x}} \cap Z_{n_2}$
- Induced map: $Z_{n_1,\boldsymbol{x}}/Z_{n_1} \to Z_{n_2,\boldsymbol{x}}/Z_{n_2}$ is injective
- $\dim(Z_{n_1,o}) - \dim(Z_{n_1}) \leqslant \dim(Z_{n_2,o}) - \dim(Z_{n_2})$

Idea. Use CLT for ***associated random variables***

- ▷ $T_1, \ldots, T_k$ ***associated*** if

$$\mathsf{Cov}\big(f_1(T_1, \ldots, T_k), f_2(T_1, \ldots, T_k)\big) \geqslant 0$$

for any increasing $f_1, f_2 \colon \mathbb{R}^k \to \mathbb{R}$.

Express edge count as

$$\sum_{i \leqslant n} T_i := \sum_{i \leqslant n} \sum_{P_j \in [i-1,i] \times [0,1]} \deg(P_j)$$

- ▷ $\{T_i\}_i$ get larger with more Poisson points
- ⇝ $\{T_i\}_i$ are associated by the ***Harris-FKG theorem***
- ▷ Remains to show $\sum_{k \geqslant 1} \mathsf{Cov}(T_1, T_k) < \infty$

Idea. Correlation induced by ***joint in-neighbors*** of T_1 and T_k. Apply Mecke formula

$$\mathrm{Cov}(T_1, T_k) = \lambda' \int_{[0,1]\times[0,1]} \int_{[k-1,k]\times[0,1]} |B(\{p, p'\})| \mathrm{d}p' \mathrm{d}p$$

Lemma (Scaling of $B(\boldsymbol{p}_m(u))$)

Let $u \leqslant 1$, $m \geqslant 1$, $0 < \gamma < 1$ *and* $0 < \gamma' < 1/(m+1)$. *It holds that*

$$\int_{\mathbb{S}^m} |B(\boldsymbol{p}_m(u))| \mathrm{d}\boldsymbol{p}_m = \frac{u^{-\gamma}}{1-(m+1)\gamma'} \frac{(2\beta)^{m+1}}{(1-\gamma)^m}$$

$$\begin{aligned}
\int_{\mathbb{S}^m} |B(\boldsymbol{p}_m(u))| \mathrm{d}\boldsymbol{p}_m &= \iint_{\mathbb{S}\times\mathbb{S}^m} \mathbb{1}\{p' \in B(\boldsymbol{p}_m(u))\} \mathrm{d}\boldsymbol{p}_m \mathrm{d}p' \\
&= \int_{B(o)} \Big(\int_{\mathbb{S}} \mathbb{1}\{|z-y| \leqslant \beta v^{-\gamma} w^{-\gamma'}\} \mathrm{d}(y,v) \Big)^m \mathrm{d}(z,w) \\
&= \int_{B(o)} \Big(\frac{2\beta w^{-\gamma'}}{1-\gamma} \Big)^m \mathrm{d}(z,w) \\
&= u^{-\gamma} \int_0^1 \frac{(2\beta w^{-\gamma'})^{m+1}}{(1-\gamma)^m} \mathrm{d}w = \frac{u^{-\gamma}}{1-(m+1)\gamma'} \frac{(2\beta)^{m+1}}{(1-\gamma)^m}
\end{aligned}$$

Idea. Split into contributions from young and old nodes.

$$S_n^{\geqslant} := \sum_{P_i \in [0,n] \times [u_n, 1]} \deg(P_i), \quad \textbf{\textit{and}} \quad S_n^{\leqslant} := \sum_{P_i \in [0,n] \times [0, u_n]} \deg(P_i),$$

where $u_n := n^{-0.9}$.

Young nodes are negligible; old nodes converge to stable distribution

It holds that

$$n^{-\gamma}(S_n^{\geqslant} - \mathbb{E}[S_n^{\geqslant}]) \xrightarrow{d} 0 \quad \textbf{\textit{and}} \quad n^{-\gamma}(S_n^{\leqslant} - \mathbb{E}[S_n^{\leqslant}]) \xrightarrow{d} \mathcal{S}$$

Idea. Apply Mecke formula on domain $B_n = [0, n] \times [u_n, 1]$

$$\mathsf{Var}(S_n^{\geqslant}) = \underbrace{\int_{B_n} \mathbb{E}[\deg(p)^2]\mathrm{d}p}_{I} + \underbrace{\int_{B_n}\int_{B_n} \mathsf{Cov}\big(\deg(p), \deg(p')\big)\mathrm{d}p\mathrm{d}p'}_{II}.$$

Term I.

- ▷ $\deg(x, u) \sim \mathsf{Poi}(\mu(u))$
- ⇝ first term of order $O(nu_n^{1-2\gamma})$; hence in $o(n^{2\gamma})$

Term II.

- ▷ Recall $\mathsf{Cov}\big(\deg(p), \deg(p')\big) = \lambda'|B(\{p, p'\})|$
- ⇝ second term of order $O(n)$; hence in $o(n^{2\gamma})$

Idea. Replace in-degrees by expectation and replace Poisson by binomial process

$$S_n^{(1)} := \sum_{\substack{X_i \in [0,n] \\ U_i \leqslant u_n}} \mu(U_i) \quad \textbf{\textit{and}} \quad S_n^{(2)} := \sum_{\substack{i \leqslant n \\ U_i \leqslant u_n}} \mu(U_i).$$

$\mathbb{E}[|S_n^{\leqslant} - S_n^{(1)}|] \in o(n^\gamma)$.

- Poisson concentration $\rightsquigarrow \mathbb{E}\big[|\deg(x,u) - \mu(u)|\big] \in O(u^{-0.6\gamma})$

$\rightsquigarrow$ $\mathbb{E}[|S_n^{\leqslant} - S_n^{(1)}|] \in O(n u_n^{1-0.6\gamma})$; and hence in $o(n^\gamma)$

$\mathbb{E}[|S_n^{(1)} - S_n^{(2)}|] \in o(n^\gamma)$.

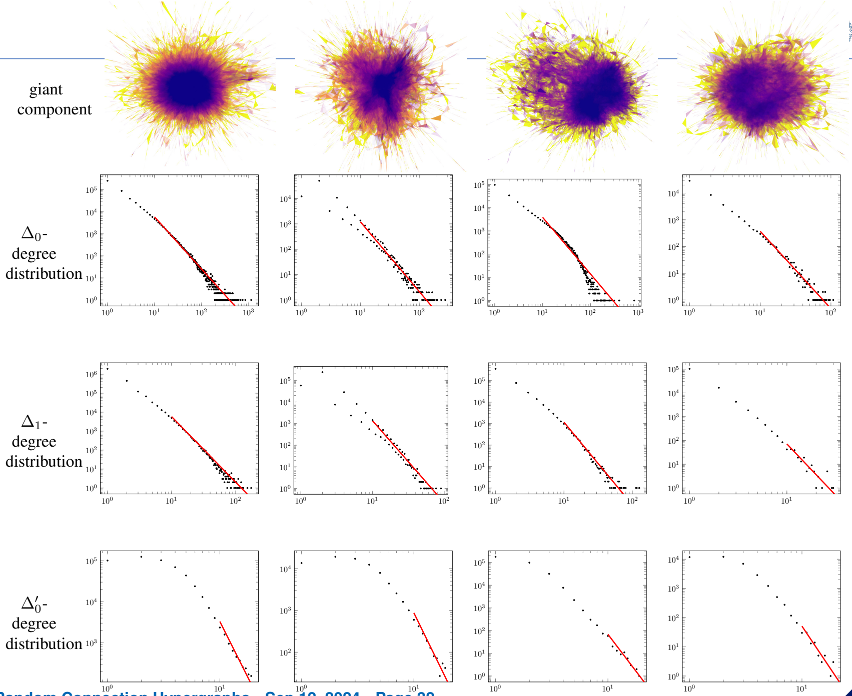
- Let $N :=$ Poi(n). Then,

$$\mathbb{E}[|S_n^{(1)} - S_n^{(2)}|] \leqslant \mathbb{E}\big[|N - n|\big] \int_0^{u_n} \mu(u)\mathrm{d}u.$$

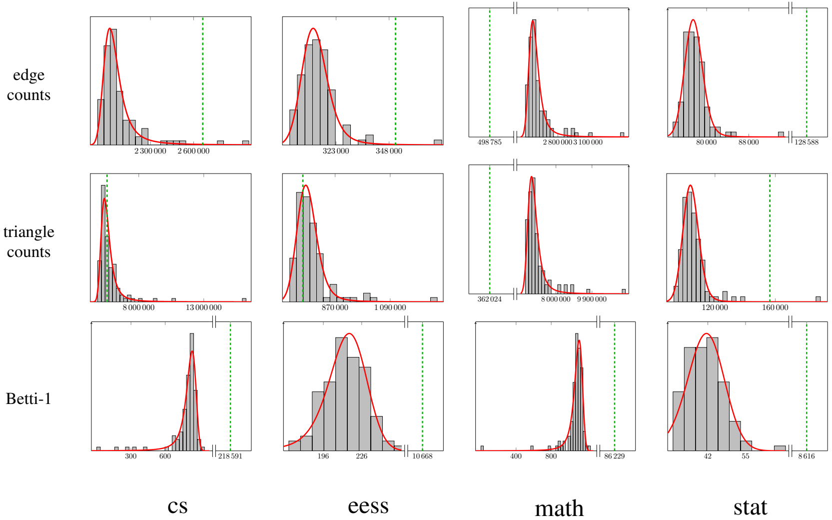
$\rightsquigarrow$ $\mathbb{E}\big[\big|S_n^{(1)} - S_n^{(2)}\big|\big] \in O(\sqrt{n} u_n^{1-\gamma})$; and hence in $o(n^\gamma)$

dataset	authors	documents	components	size of largest component
cs	433 244	452 881	22 576	370 494
eess	77 686	69 594	5 533	54 147
math	198 601	466 428	26 197	152 441
stat	44 380	36 689	4 049	32 373

dataset	Δ_0-degree		$\mathcal{P}'$-vertex degree		β	λ	λ'
	exponent	γ	exponent	γ'			
cs	2.37	0.73	5.46	0.22	$8.19E-07$	$579\,719$	$532\,491$
eess	2.75	0.57	5.50	0.22	$7.89E-06$	$98\,528$	$83\,985$
math	2.44	0.69	6.51	0.18	$1.03E-06$	$231\,606$	$588\,628$
stat	2.89	0.53	5.74	0.21	$1.13E-05$	$57\,488$	$41\,655$



giant component
Δ_0-degree distribution
Δ_1-degree distribution
Δ_0'-degree distribution


edge counts
triangle counts
Betti-1
2 300 000
2 600 000
323 000
348 000
498 785
2 800 000
3 100 000
80 000
88 000
128 588
8 000 000
13 000 000
870 000
1 090 000
362 024
8 000 000
9 900 000
120 000
160 000
300
600
218 591
196
226
10 668
400
800
86 229
42
55
8 616
cs
eess
math
stat

- ***Random connection hypergraph*** as new model ✓
- ***Scale-free*** extension of AB random geometric graph ✓
- ***Higher-order degree distribution*** for ADRCM ✓
- ***CLT*** for Betti number and simplex count ✓
- ***Alpha-stable limit*** for simplex count ✓
- ***Speed of convergence*** in normal approximation**!**
- Time-varying extensions **!**
- Alpha-stable convergence of Betti number**!**

Thank you!

Literature

[1] M. Benaim, I. Benjamini, J. Chen, and Y. Lima.
A generalized Pólya's urn with graph based interactions.
Random Structures Algorithms, 46(4):614–634, 2015.

[2] M. Heydenreich and C. Hirsch.
Extremal linkage networks.
Extremes, 2021, forthcoming.

[3] C. Hirsch, M. Holmes, and V. Kleptsyn.
Absence of WARM percolation in the very strong reinforcement regime.
Ann. Appl. Probab., 31(1):199–217, 2021.

[4] R. v. d. Hofstad, M. Holmes, A. Kuznetsov, and W. Ruszel.
Strongly reinforced Pólya urns with graph-based competition.
Ann. Appl. Probab., 26(4):2494–2539, 2016.

[5] M. Holmes and V. Kleptsyn.
Proof of the WARM whisker conjecture for neuronal connections.
Chaos, 27(4):043104, 2017.

[6] D. Kious, C. Mailler, and B. Schapira.
Finding geodesics on graphs using reinforcement learning.
Ann. Appl. Probab., 2022+.

[7] V. B. Tadić.
Convergence and convergence rate of stochastic gradient search in the case of multiple and non-isolated extrema.
Stochastic Process. Appl., 125(5):1715–1755, 2015.