Università degli Studi di Padova

CoMuNe lab
COMPLEX MULTILAYER NETWORKS

QUANTIFYING BIASES IN RECONSTRUCTED BRAIN NETWORKS

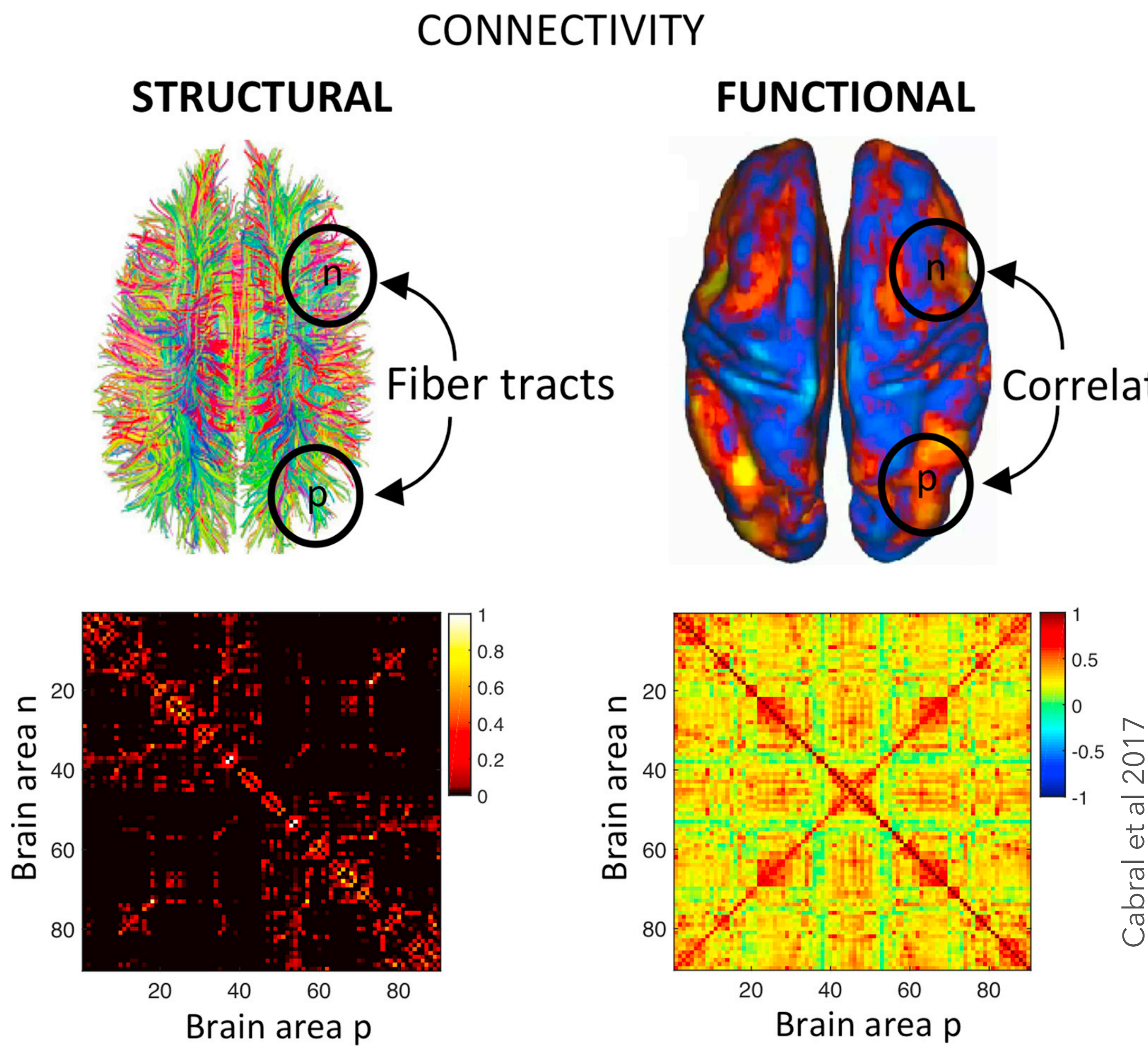
Alessandra Corso, Valeria d'Andrea, Manlio De Domenico

Complex Multilayer Networks Lab
Dept. of Physics & Astronomy "Galileo Galilei", University of Padua

CSS/Italy
Italian Chapter of Complex Systems Society

CCS/Italy 2025
Bari, Italy
15-17 September 2025

BIASES RELATED TO NETWORK RECONSTRUCTION



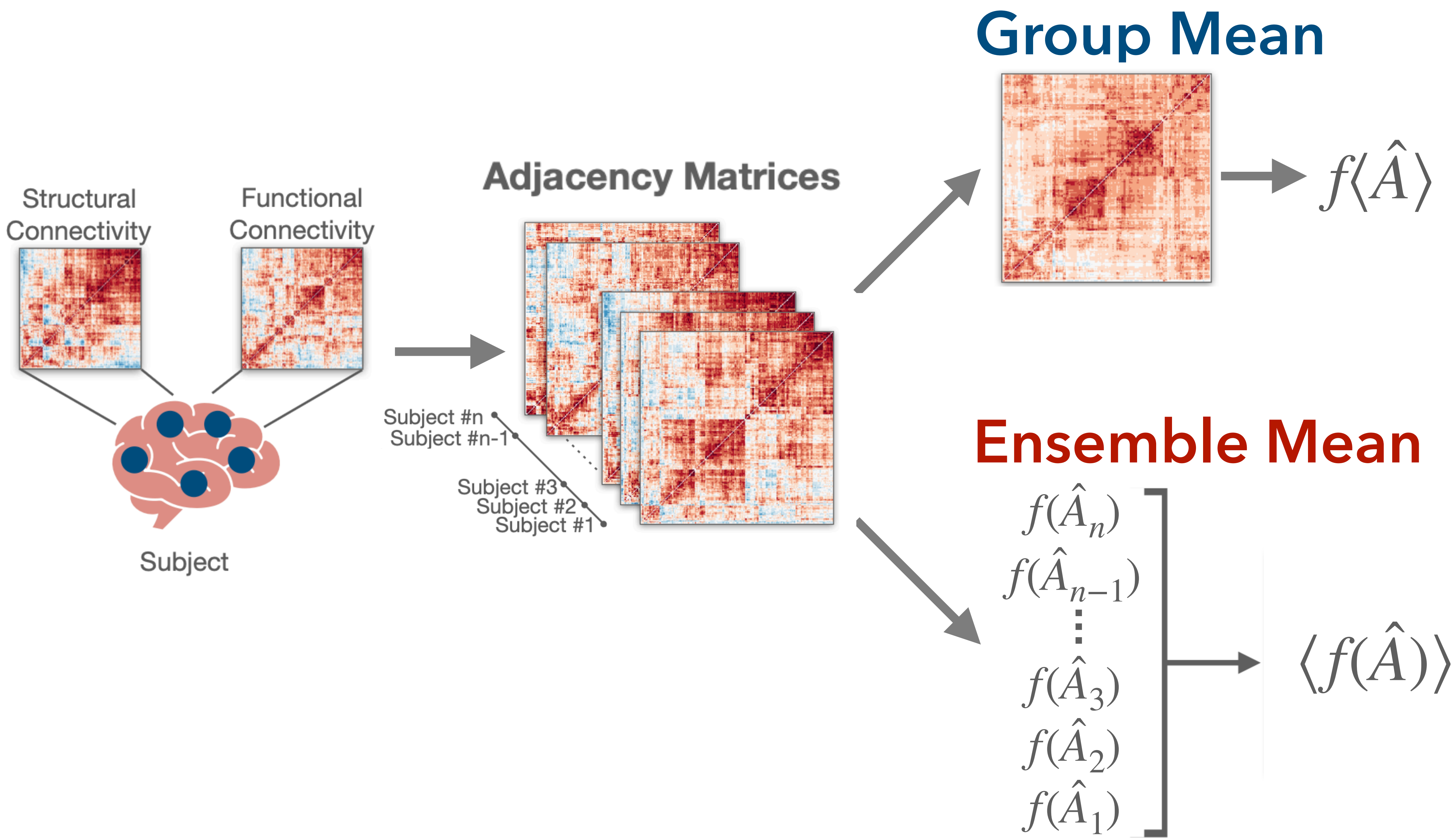
- Functional MRI network reconstruction based on:
  i) a correlation measures
  ii) thresholding procedure

What does the use of thresholds imply?

- Networks with arbitrary and variable edge densities, which can impact the evaluation of network metrics

We aim to quantify how threshold and network edge density affect the estimation of different metrics

BIASES RELATED TO STATISTICAL APPROACH



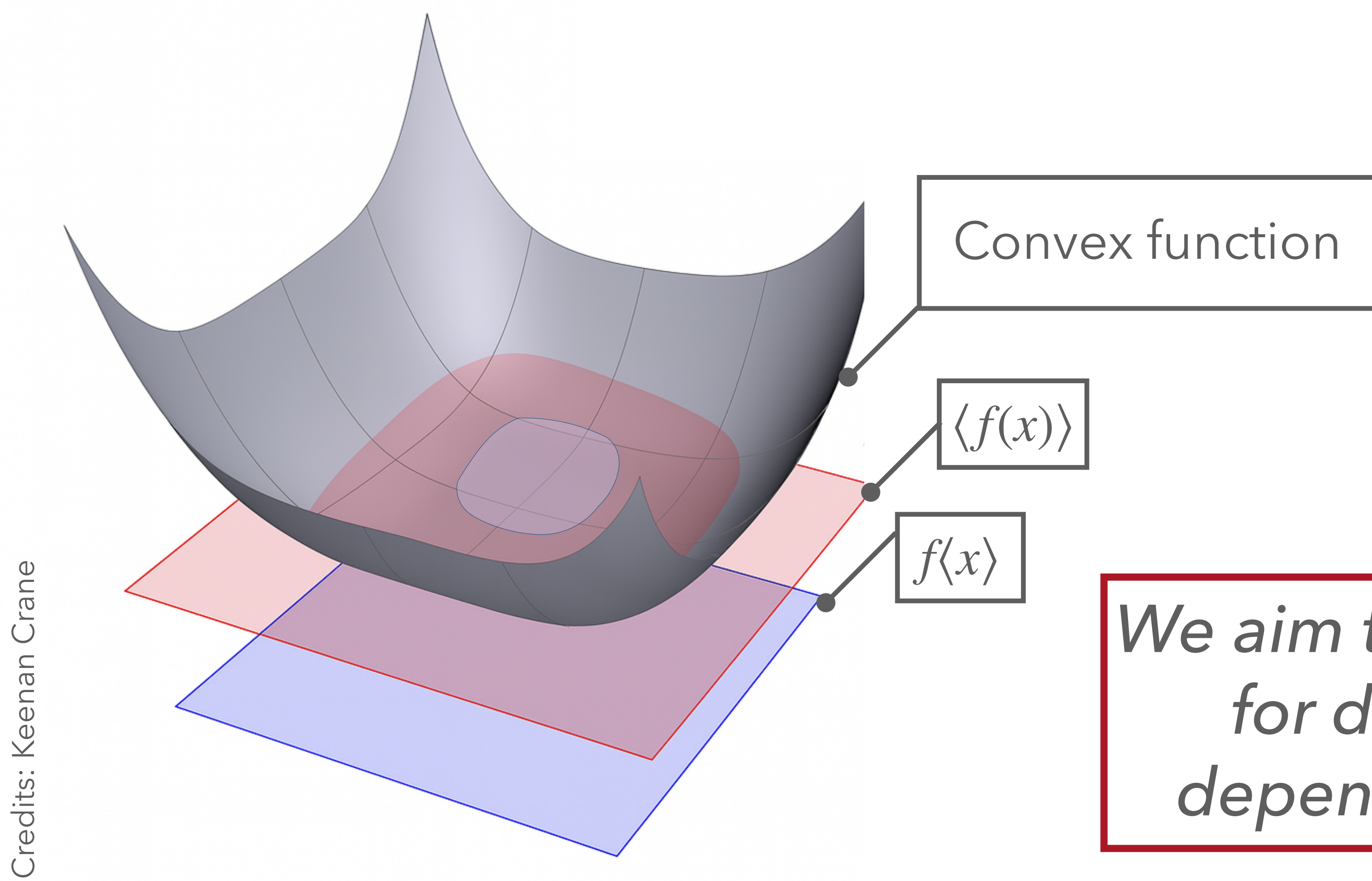
BIASES RELATED TO STATISTICAL APPROACH

Statistical Mechanics Framework:

Evaluation of a metric on the adjacency matrix $\Leftrightarrow$ Application of a non linear function on the adjacency matrix

Jensen's inequality

Expectation of the function of random variable $\boxed{\langle f(x) \rangle} \geq \boxed{f\langle x \rangle}$ Function of the expectation of random variable



We aim to quantify Jensen's inequality for different metrics and how it depends on network edge density

NETWORK METRICS

MACRO
- Average shortest path length
- Global communication efficiency
- Small–world index

MESO
- Modularity
- Global clustering coefficient

MICRO
- Betweenness centrality
- Page rank centrality

NETWORK METRICS

- Average shortest path length
- Global communication efficiency
- Small–world index

MACRO

- Modularity
- **Global clustering coefficient**

MESO

- Betweenness centrality
- Page rank centrality

MICRO

Can be treated analytically, if:

- The network is generated with a configuration model

$$\langle \hat{\mathbf{A}}_{i,j} \rangle = p_{i,j} = \frac{k_i k_j}{2m}$$

$$C(\langle \hat{\mathbf{A}}_{CM} \rangle) = \frac{(\overline{k^2})^2}{N(\overline{k})^3 - \overline{k}\ \overline{k^2}}$$

$$\langle C(\hat{\mathbf{A}}_{CM}) \rangle = \frac{\left[\overline{k^2} - \overline{k}\right]^2}{N(\overline{k})^3}$$

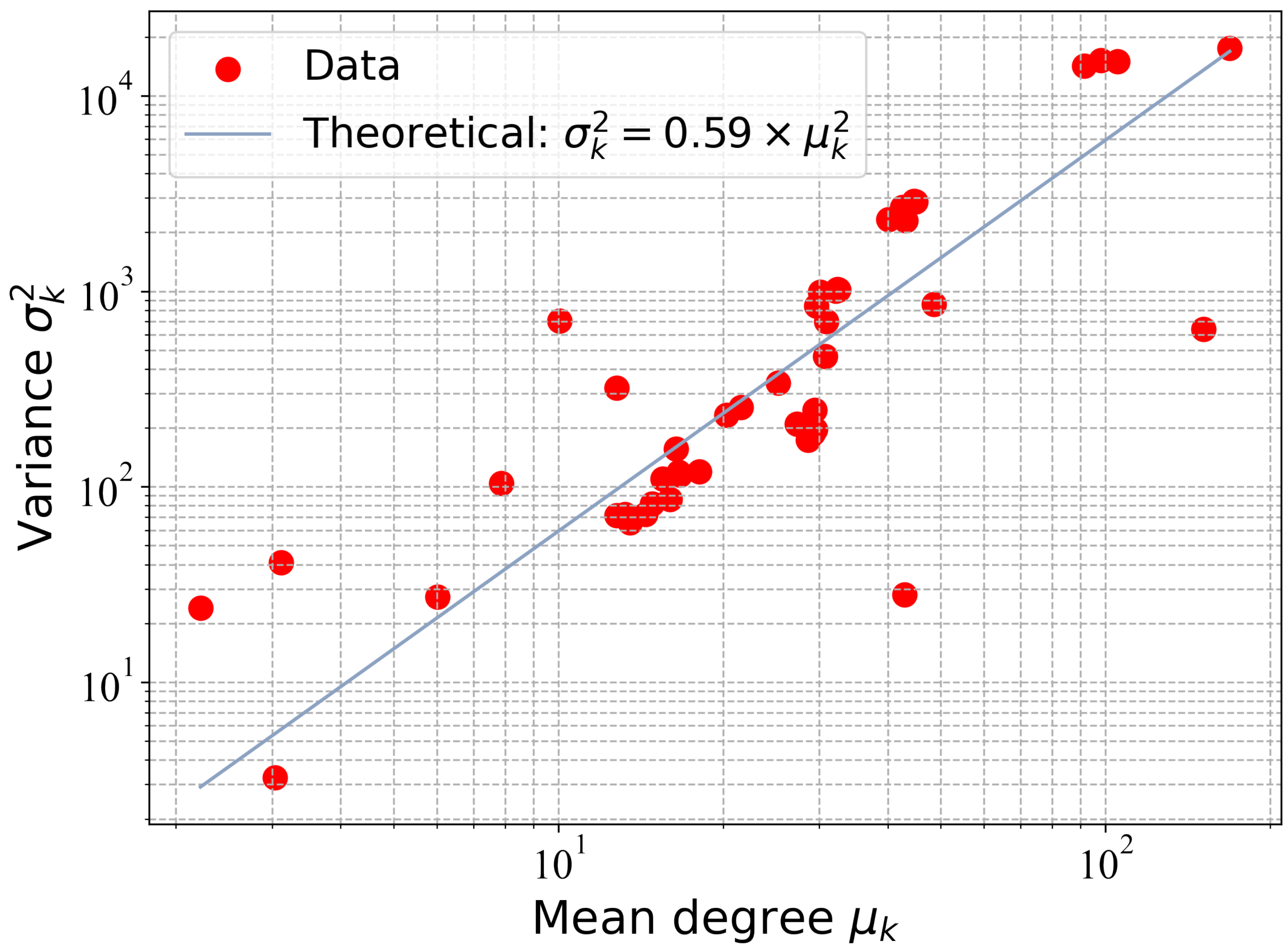
- The degree distribution is either:
  - ➡ a Poisson distribution
  - ➡ a distribution where it holds the Taylor's law

Why Taylor's law?

Variance V of a non-negative random function is a power function of its mean m:

$$V = am^b$$

- Taylor's law is widely observed in biological systems



Human and Non-human connectomes

ANALYTICAL ESTIMATION OF CLUSTERING COEFFICIENT AS A FUNCTION OF ρ

	Group Mean $f\langle \hat{A} \rangle$	Ensemble Mean $\langle f(\hat{A}) \rangle$	Jensen's difference $f\langle \hat{A} \rangle - \langle f(\hat{A}) \rangle$
POISSON	$C(\langle \hat{A}^{Poiss}_{CM} \rangle \approx \rho + \frac{1}{N\bar{k}} + \frac{2}{N}$	$\langle C(\hat{\mathbf{A}}^{Poiss}_{CM}) \rangle = \frac{\bar{k}}{N} = \rho$	$C(\langle \hat{\mathbf{A}} \rangle^{Poiss}_{CM}) - \langle C(\hat{\mathbf{A}}^{Poiss}_{CM}) \rangle = = \frac{1}{N}\left[\frac{1}{\bar{\bar{k}}} + 2\right]$
TAYLOR'S LAW	$C(\langle \hat{\mathbf{A}}^{TL}_{CM} \rangle) \approx c^2 \rho \left[1 + \frac{c}{N}\right]$	$\langle C(\hat{\mathbf{A}}^{TL}_{CM}) \rangle = = c^2 \rho + \frac{1}{(\bar{k})^2}\rho - \frac{2c}{\bar{\bar{k}}}\rho$	$C(\langle \hat{\mathbf{A}} \rangle^{TL}_{CM}) - \langle C(\hat{\mathbf{A}}^{TL}_{CM}) \rangle = = \rho \left[\frac{c^3}{N} + \frac{2c}{\bar{\bar{k}}} - \frac{1}{(\bar{k})^2}\right]$

- For Poisson degree distribution, Jd does not depend on ρ and goes to zero for $N \rightarrow \infty$
- For Taylor's law degree distribution, all estimates linearly increase with ρ

ANALYTICAL ESTIMATION OF CLUSTERING COEFFICIENT AS A FUNCTION OF ρ

	Group Mean $f\langle\hat{A}\rangle$	Ensemble Mean $\langle f(\hat{A})\rangle$	Jensen's difference $f\langle\hat{A}\rangle - \langle f(\hat{A})\rangle$
POISSON	$C(\langle \hat{A}^{Poiss}_{CM}\rangle \approx \rho + \frac{1}{N\bar{k}} + \frac{2}{N}$	$\langle C(\hat{\mathbf{A}}^{Poiss}_{CM})\rangle = \frac{\bar{k}}{N} = \rho$	$C(\langle\hat{\mathbf{A}}\rangle^{Poiss}_{CM}) - \langle C(\hat{\mathbf{A}}^{Poiss}_{CM})\rangle =$ $= \frac{1}{N}\left[\frac{1}{\bar{\bar{k}}} + 2\right]$
TAYLOR'S LAW	$C(\langle\hat{\mathbf{A}}^{TL}_{CM}\rangle) \approx c^2\rho\left[1 + \frac{c}{N}\right]$	$\langle C(\hat{\mathbf{A}}^{TL}_{CM})\rangle =$ $= c^2\rho + \frac{1}{(\bar{k})^2}\rho - \frac{2c}{\bar{\bar{k}}}\rho$	$C(\langle\hat{\mathbf{A}}\rangle^{TL}_{CM}) - \langle C(\hat{\mathbf{A}}^{TL}_{CM})\rangle =$ $= \rho\left[\frac{c^3}{N} + \frac{2c}{\bar{\bar{k}}} - \frac{1}{(\bar{k})^2}\right]$

- For Poisson degree distribution, Jd does not depend on ρ and goes to zero for $N \to \infty$
- For Taylor's law degree distribution, all estimates linearly increase with ρ

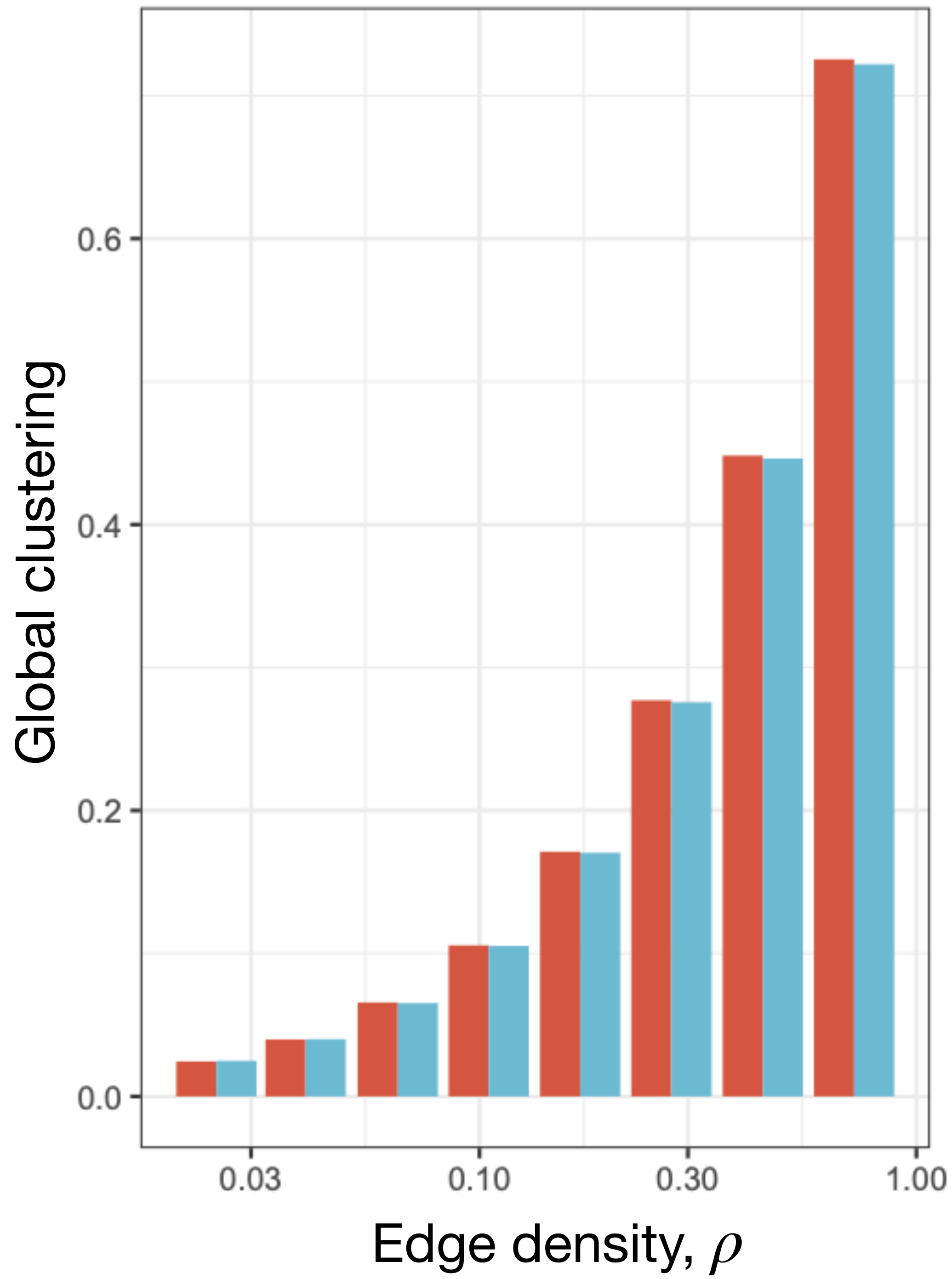
ANALYTICAL ESTIMATION OF CLUSTERING COEFFICIENT AS A FUNCTION OF ρ

	Group Mean $f\langle\hat{A}\rangle$	Ensemble Mean $\langle f(\hat{A})\rangle$	Jensen's difference $f\langle\hat{A}\rangle - \langle f(\hat{A})\rangle$
POISSON	$C(\langle \hat{A}^{Poiss}_{CM}\rangle \approx \rho + \frac{1}{N\bar{k}} + \frac{2}{N}$	$\langle C(\hat{\mathbf{A}}^{Poiss}_{CM})\rangle = \frac{\bar{k}}{N} = \rho$	$C(\langle\hat{\mathbf{A}}\rangle^{Poiss}_{CM}) - \langle C(\hat{\mathbf{A}}^{Poiss}_{CM})\rangle = = \frac{1}{N}\left[\frac{1}{\bar{\bar{k}}} + 2\right]$
TAYLOR'S LAW	$C(\langle\hat{\mathbf{A}}^{TL}_{CM}\rangle) \approx c^2\rho\left[1 + \frac{c}{N}\right]$	$\langle C(\hat{\mathbf{A}}^{TL}_{CM})\rangle = = c^2\rho + \frac{1}{(\bar{k})^2}\rho - \frac{2c}{\bar{\bar{k}}}\rho$	$C(\langle\hat{\mathbf{A}}\rangle^{TL}_{CM}) - \langle C(\hat{\mathbf{A}}^{TL}_{CM})\rangle = = \rho\left[\frac{c^3}{N} + \frac{2c}{\bar{\bar{k}}} - \frac{1}{(\bar{k})^2}\right]$

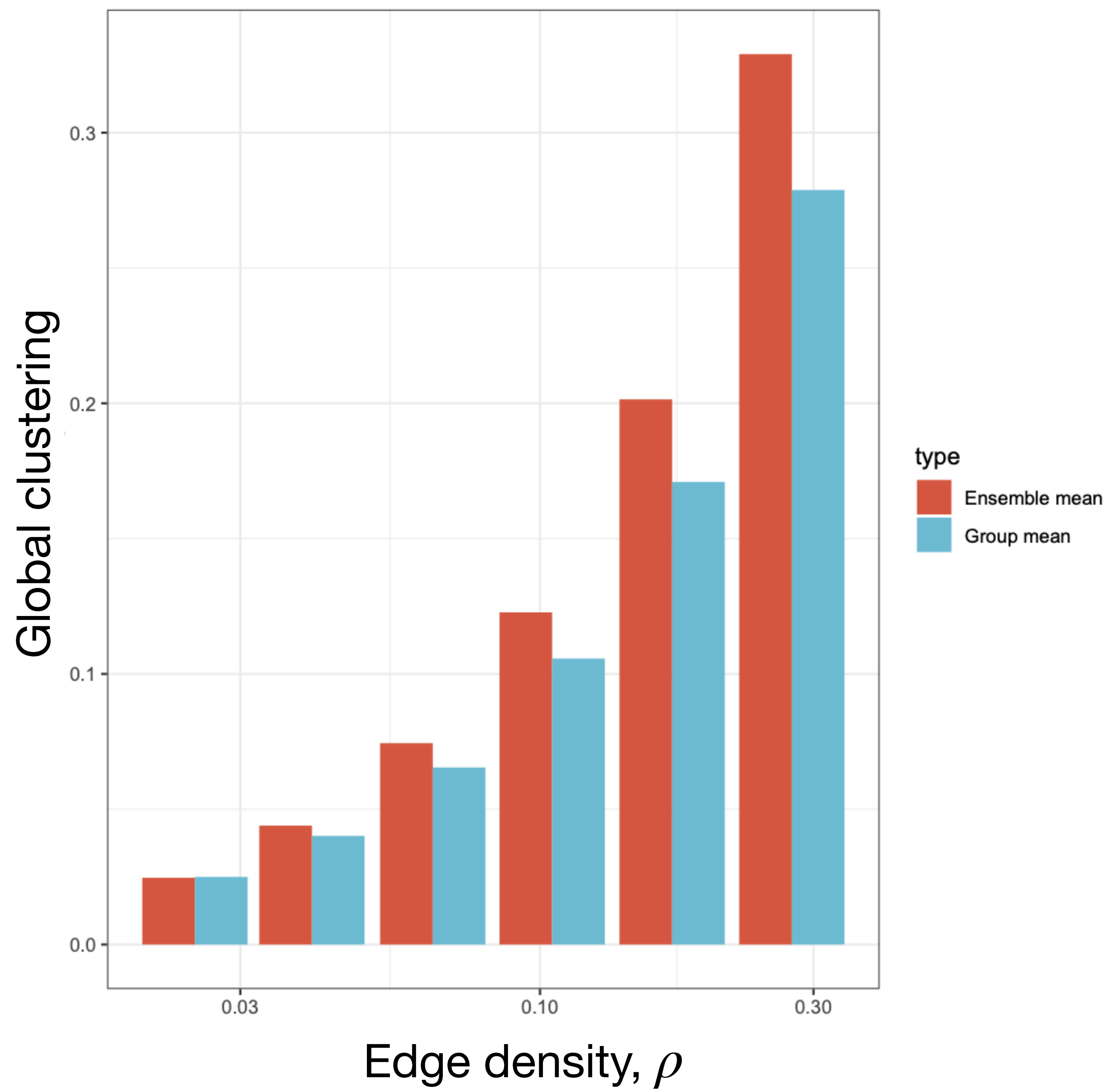
- For Poisson degree distribution, Jd does not depend on ρ and goes to zero for $N \to \infty$
- For Taylor's law degree distribution, all estimates linearly increase with ρ

ANALYTICAL RESULTS ARE VALIDATED IN SIMULATIONS ON SYNTHETIC NETWORKS

N=1000, configuration model with Poisson degree distribution



N=1000, configuration model with Taylor law degree distribution



OTHER METRICS: ANALYSIS OF MULTIMODAL NEUROIMAGING DATA

Real data

- Nathan S. Kline Institute - Rockland Sample (NKI-RS):
  - structural and functional connectomes
  - 196 healthy subjects
  - N=188

frontiers in NEUROSCIENCE
REVIEW ARTICLE
published: 16 October 2012
doi: 10.3389/fnins.2012.00152

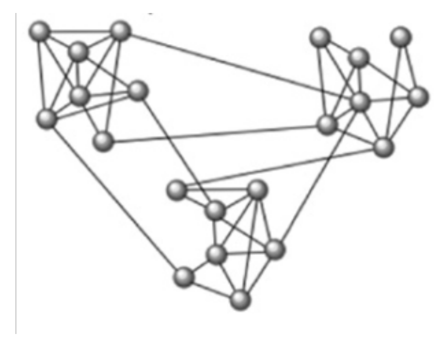
The NKI-Rockland sample: a model for accelerating the pace of discovery science in psychiatry

*Kate Brody Nooner[1,2], Stanley J. Colcombe[1], Russell H. Tobe[1], Maarten Mennes[3,4], Melissa M. Benedict[1], Alexis L. Moreno[1], Laura J. Panek[1], Shaquanna Brown[1], Stephen T. Zavitz[1], Qingyang Li[5], Sharad Sikka[5], David Gutman[6], Saroja Bangaru[3], Rochelle Tziona Schlachter[1], Stephanie M. Kamiel[1], Ayesha R. Anwar[5], Caitlin M. Hinz[5], Michelle S. Kaplan[5], Anna B. Rachlin[5], Samantha Adelsberg[3], Brian Cheung[5], Ranjit Khanuja[5], Chaogan Yan[5], Cameron C. Craddock[7], Vincent Calhoun[8,9], William Courtney[9], Margaret King[9], Dylan Wood[9], Christine L. Cox[3], A. M. Clare Kelly[3], Adriana Di Martino[3], Eva Petkova[1,3], Philip T. Reiss[1,3], Nancy Duan[5], Dawn Thomsen[5], Bharat Biswal[10], Barbara Coffey[1,6], Matthew J. Hoptman[1,6], Daniel C. Javitt[1,11], Nunzio Pomara[1,6], John J. Sidtis[1,6], Harold S. Koplewicz[5], Francisco Xavier Castellanos[1,3], Bennett L. Leventhal[1] and Michael P. Milham[1,5]**

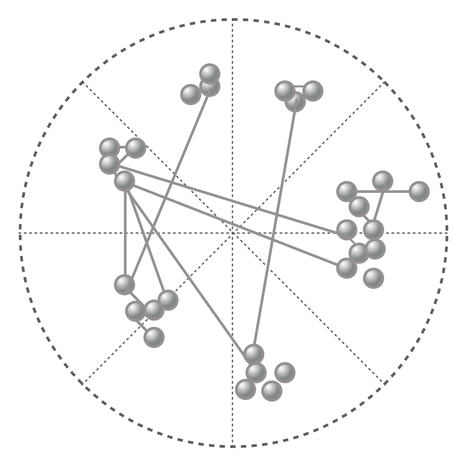
Synthetic models

Which network feature do explain results in real networks?

- Configuration model (degree distribution)

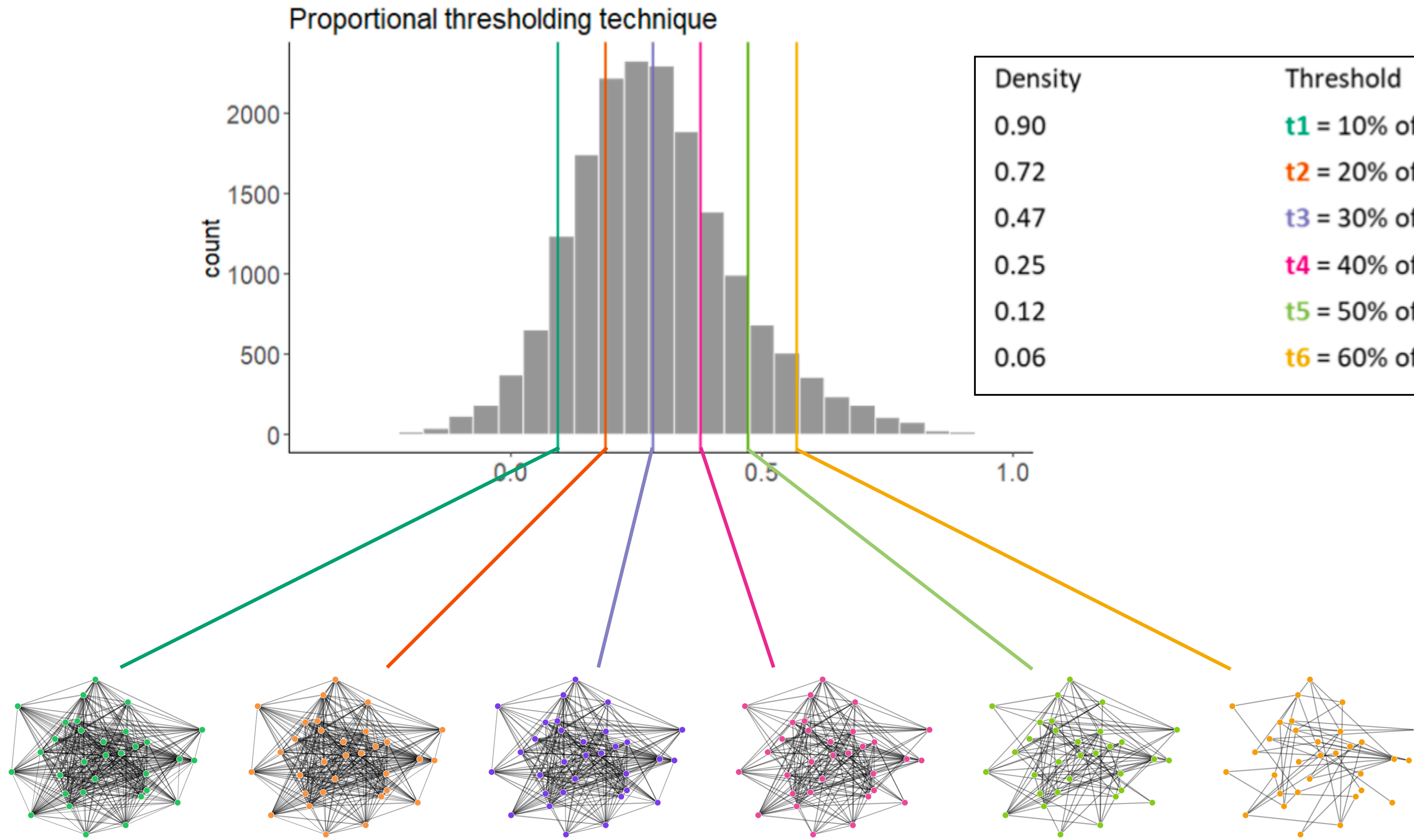


- Stochastic block model (modular structure/ mesoscale organization)



- Hyperbolic model (degree-degree correlations)

NETWORK MEASURES ARE SENSITIVE TO RECONSTRUCTION CHOICES

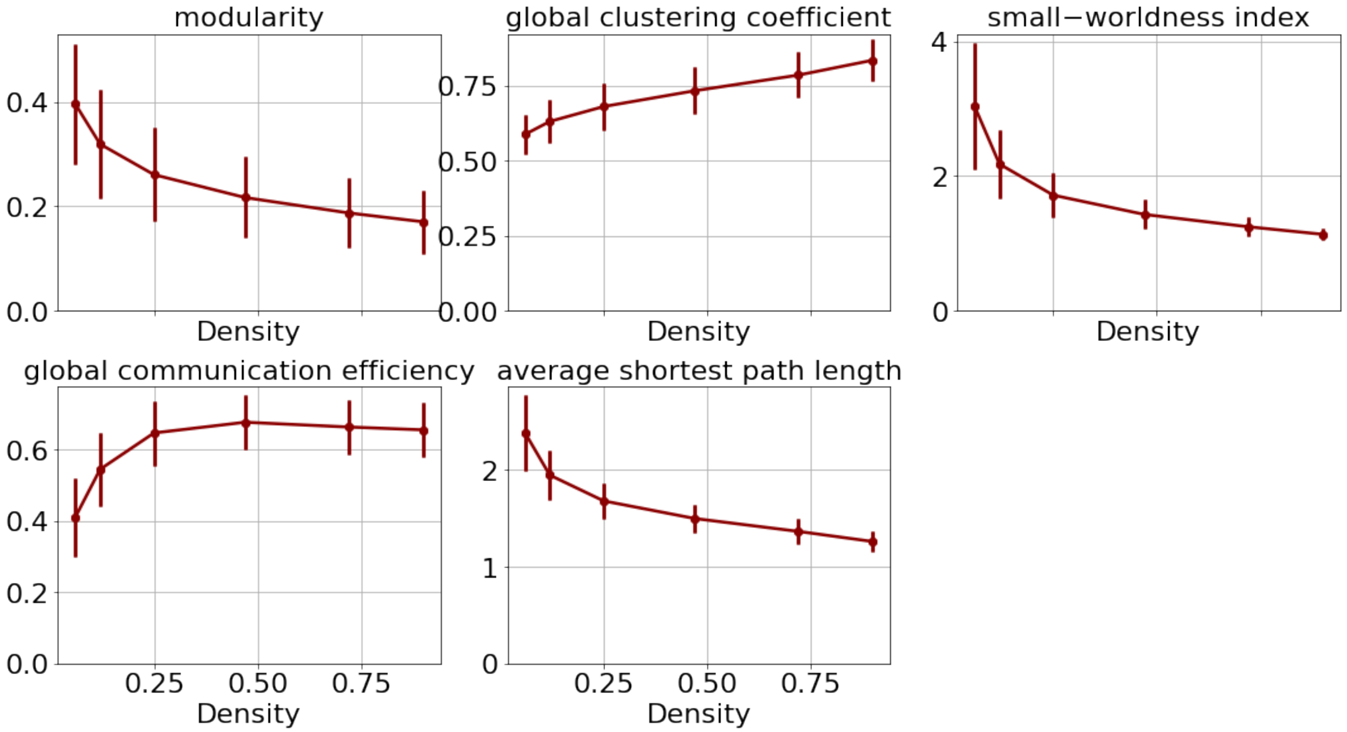


Density	Threshold
0.90	t1 = 10% of $\text{weight}_{\max}$
0.72	t2 = 20% of $\text{weight}_{\max}$
0.47	t3 = 30% of $\text{weight}_{\max}$
0.25	t4 = 40% of $\text{weight}_{\max}$
0.12	t5 = 50% of $\text{weight}_{\max}$
0.06	t6 = 60% of $\text{weight}_{\max}$

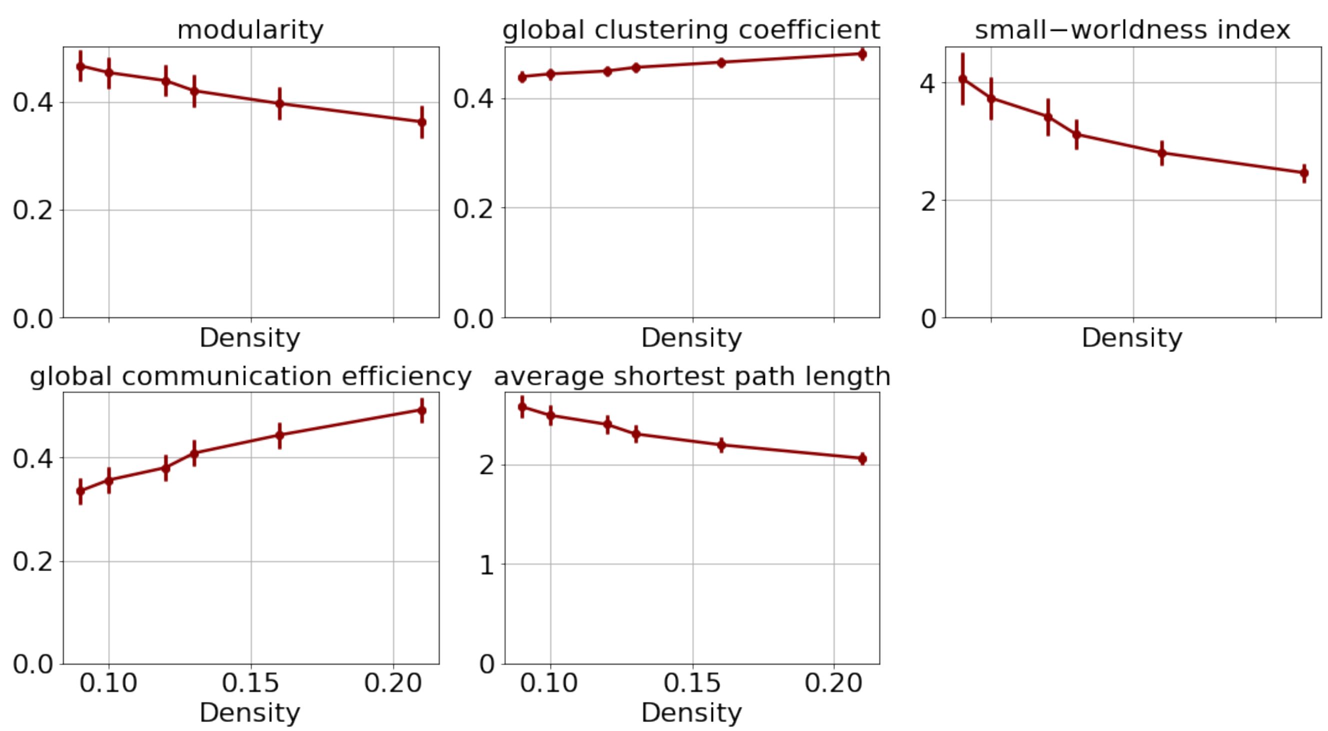
NETWORK MEASURES ARE SENSITIVE TO RECONSTRUCTION CHOICES

- Real data - Ensemble mean estimates

Functional

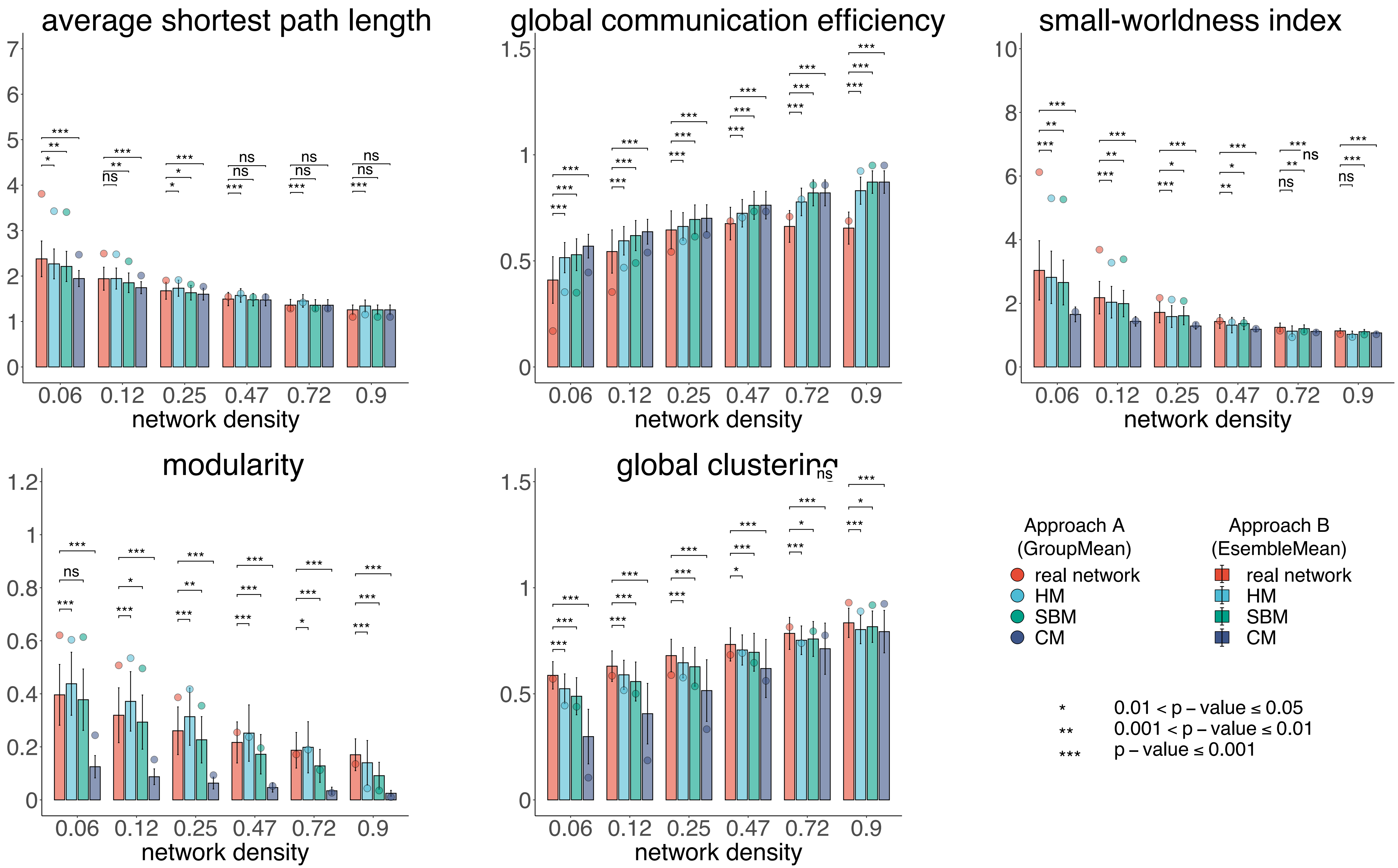


Structural



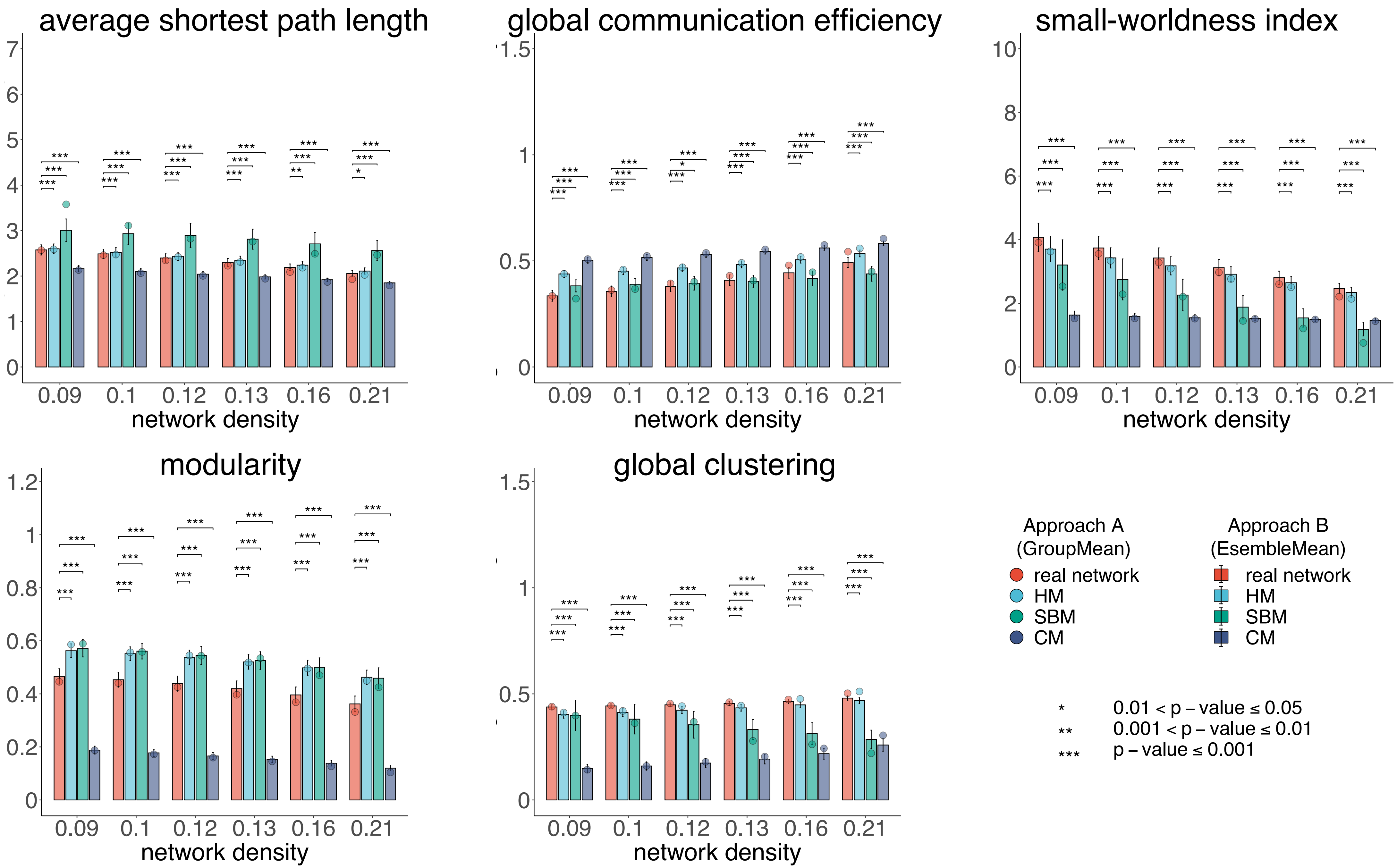
- Strong dependency on ρ in functional networks
- Weaker dependency on ρ in structural networks
- Different metrics have different dependency trends

FUNCTIONAL NETWORKS: COMPARISON WITH GENERATIVE MODELS

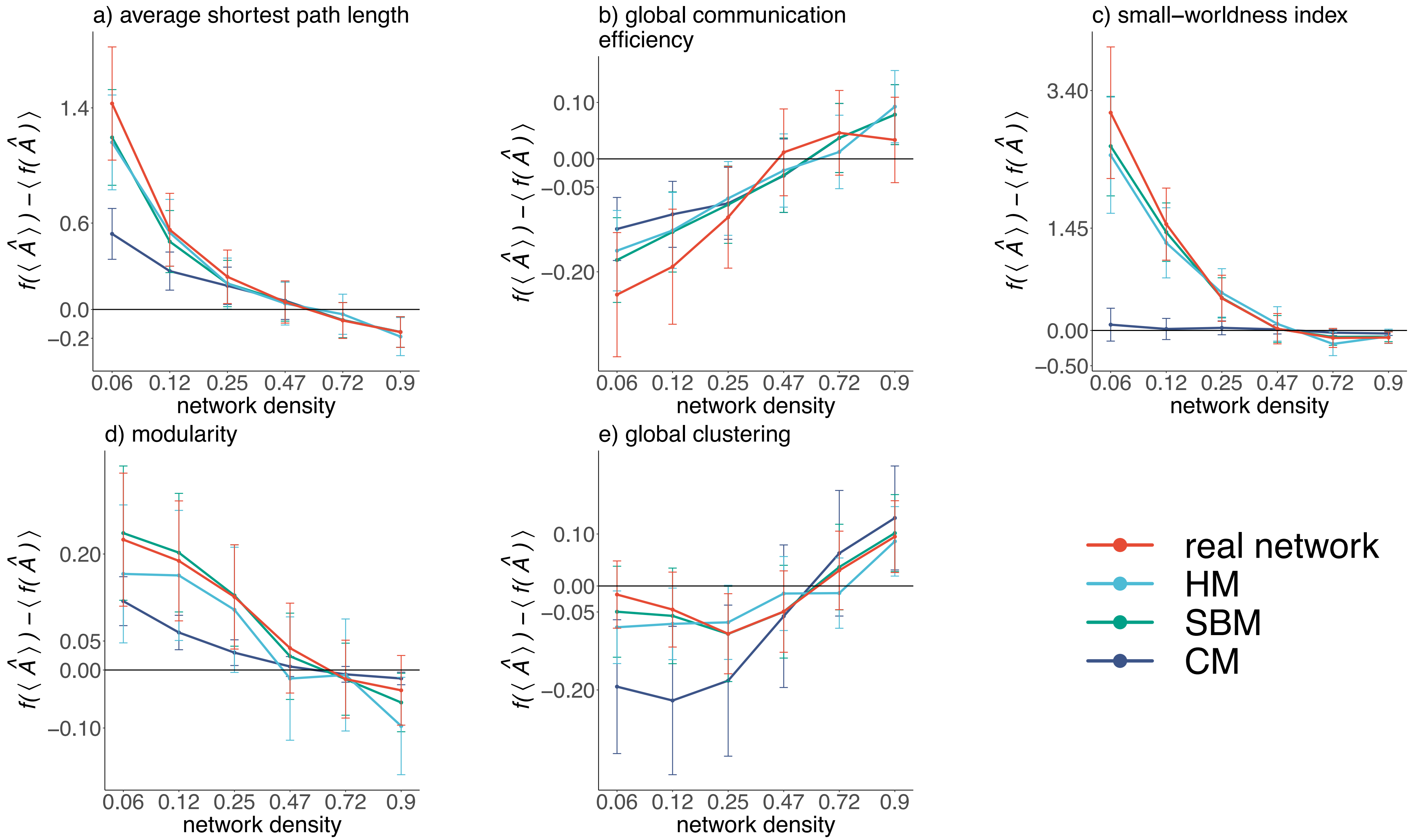


- Hyperbolic model best approximates real data
- At high densities real network estimates do not differ from estimate in random ones: effect of uncertain link

STRUCTURAL NETWORKS: COMPARISON WITH GENERATIVE MODELS



JENSEN'S DIFFERENCE IN FUNCTIONAL NETWORKS



- Jensen's difference itself depends on edge density
- Jensen's difference is zero for $\rho > 0.25$ (select of more than top $60\,\%$ weights of links)

➡For high value of network density, results are reconciled but real and random networks measures are not statistically different

KEY TAKEAWAY: UNCERTAINTY MUST BE EMBRACED

- We presented 2 types of biases (variability across subject and uncertainty about a link) that can affect brain network analysis
- At high densities, measures lose discriminative power, making the reconstructed networks indistinguishable from random ones.
- The results of our study highlight the limitations of threshold-based reconstruction and emphasize the need to account for the inherent variability in the reconstruction of a system.

KEY TAKEAWAY: UNCERTAINTY MUST BE EMBRACED

- We presented 2 types of biases (variability across subject and uncertainty about a link) that can affect brain network analysis

- At high densities, measures lose discriminative power, making the reconstructed networks indistinguishable from random ones.

- The results of our study highlight the limitations of threshold-based reconstruction and emphasize the need to account for the inherent variability in the reconstruction of a system.

How?

Statistical Inference Methods considering generative models

nature communications

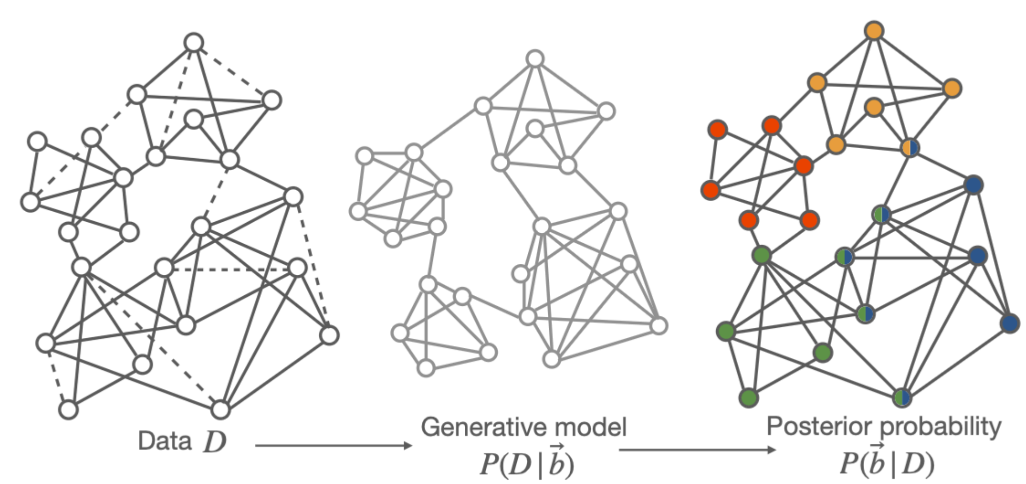
Perspective https://doi.org/10.1038/s41467-022-34267-9

Statistical inference links data and theory in network science

Received: 1 June 2021

Accepted: 18 October 2022

Leto Peel[1], Tiago P. Peixoto[2] & Manlio De Domenico[3]



OUR PAPER IS CURRENTLY UNDER (FINAL!) REVIEW

Quantifying biases in reconstructed brain networks

Alessandra Corso[*]
Department of Electrical Electronic and Computer Science Engineering, University of Catania, Catania, Italy

Valeria d'Andrea[*]
Department of Physics and Astronomy "Galileo Galilei",
University of Padua, Via F. Marzolo 8, 315126 Padova, Italy

Manlio De Domenico[†]
Department of Physics and Astronomy "Galileo Galilei",
University of Padua, Via F. Marzolo 8, 315126 Padova, Italy
Padua Center for Network Medicine, University of Padua, Via F. Marzolo 8, 315126 Padova, Italy
Istituto Nazionale di Fisica Nucleare, Sez. Padova, Italy

(Dated: May 27, 2025)

THANK YOU! ANY QUESTIONS?

Valeria d'Andrea, Ph.D.
University of Padua
E-mail : valeria.dandrea@unipd.it

Finanziato dall'Unione europea
NextGenerationEU

Ministero delle Infrastrutture e dei Trasporti

Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA

CLUSTERING COEFFICIENT

$$C(\langle \hat{\mathbf{A}} \rangle) = \frac{3 \times \text{number of triangles in the graph}}{\text{number of connected triplets of nodes}} = \frac{\mathrm{Tr}(\langle \hat{\mathbf{A}} \rangle^3)}{\mathrm{Tr}(\langle \hat{\mathbf{A}} \rangle^2 (\hat{U} - \hat{I}))}$$

$$\hat{U} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \qquad \hat{I} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \equiv \delta_{i,j}$$

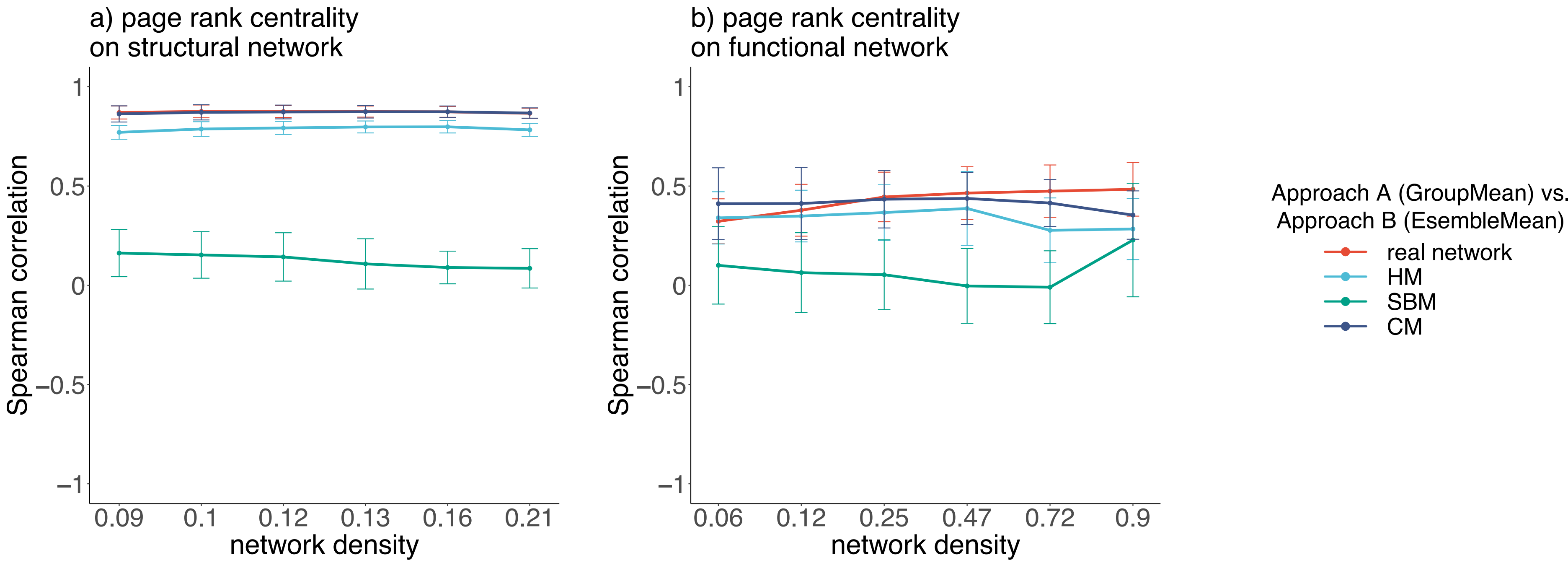
In a configuration model $\langle \hat{\mathbf{A}}_{i,j} \rangle = p_{i,j} = \frac{k_i k_j}{2m}$

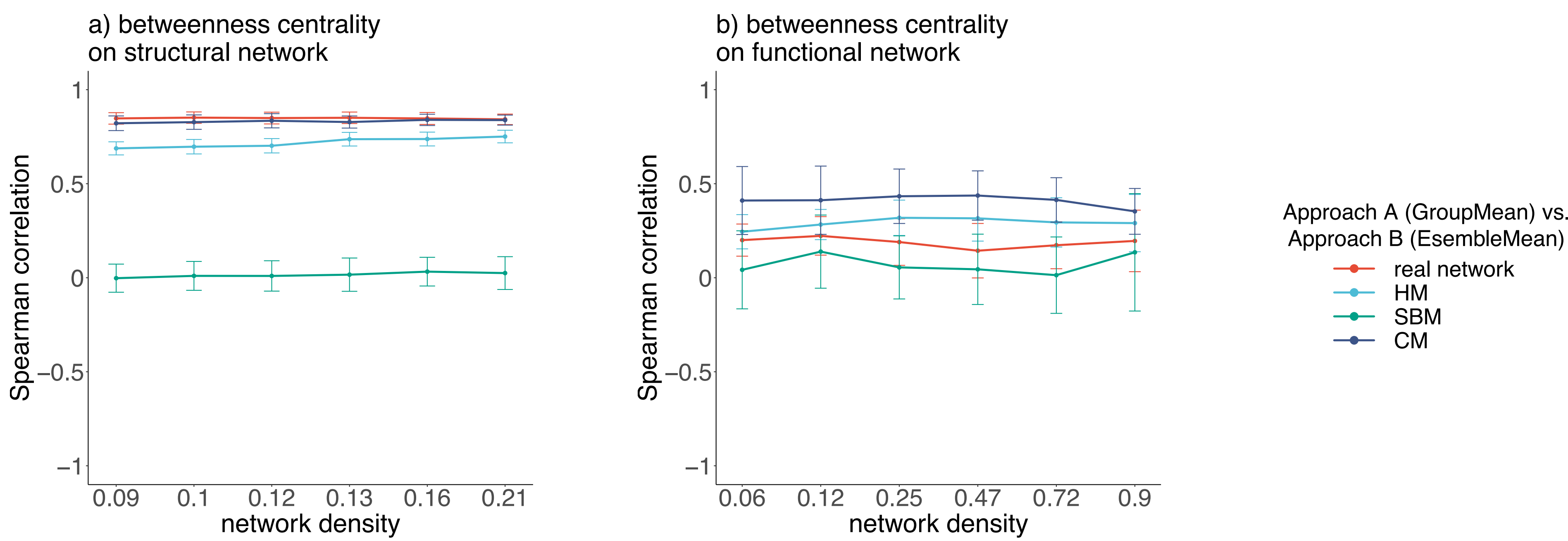
$$C(\langle \hat{\mathbf{A}}_{CM} \rangle) = \frac{\frac{1}{2m} \sum_{i,j,l} (k_i k_j)(k_j k_l)(k_l k_i)}{\sum_{i,j,l} (k_i k_j)(1 - \delta_{j,k})(k_l k_i)} = \frac{\frac{1}{2m} \left[\sum_i k_i^2 \right]^3}{\sum_i k_i^2 (\sum_i k_i)^2 - (\sum_i k_i^2)^2} = \frac{(\overline{k^2})^2}{N(\overline{k})^3 - \overline{k}\, \overline{k^2}}$$

$$\langle C(\hat{\mathbf{A}}_{CM}) \rangle = \frac{\left[\overline{k^2} - \overline{k} \right]^2}{N(\overline{k})^3}$$

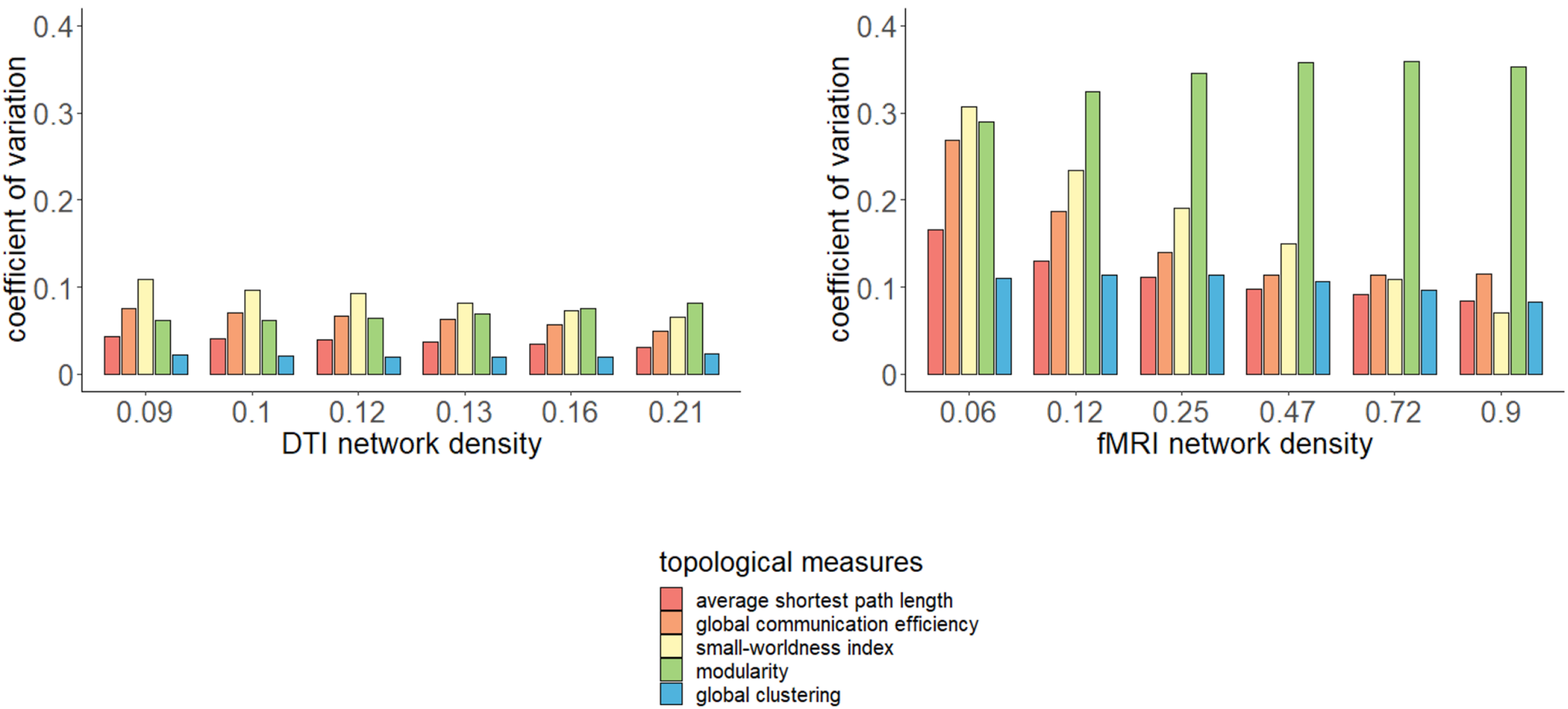
Then we need to remove the dependency on $\overline{k^2}$ to write expressions as a function of network density $\rho = \frac{\overline{k}}{N}$

MICROSCALE MEASURES





COEFFICIENT OF VARIATION



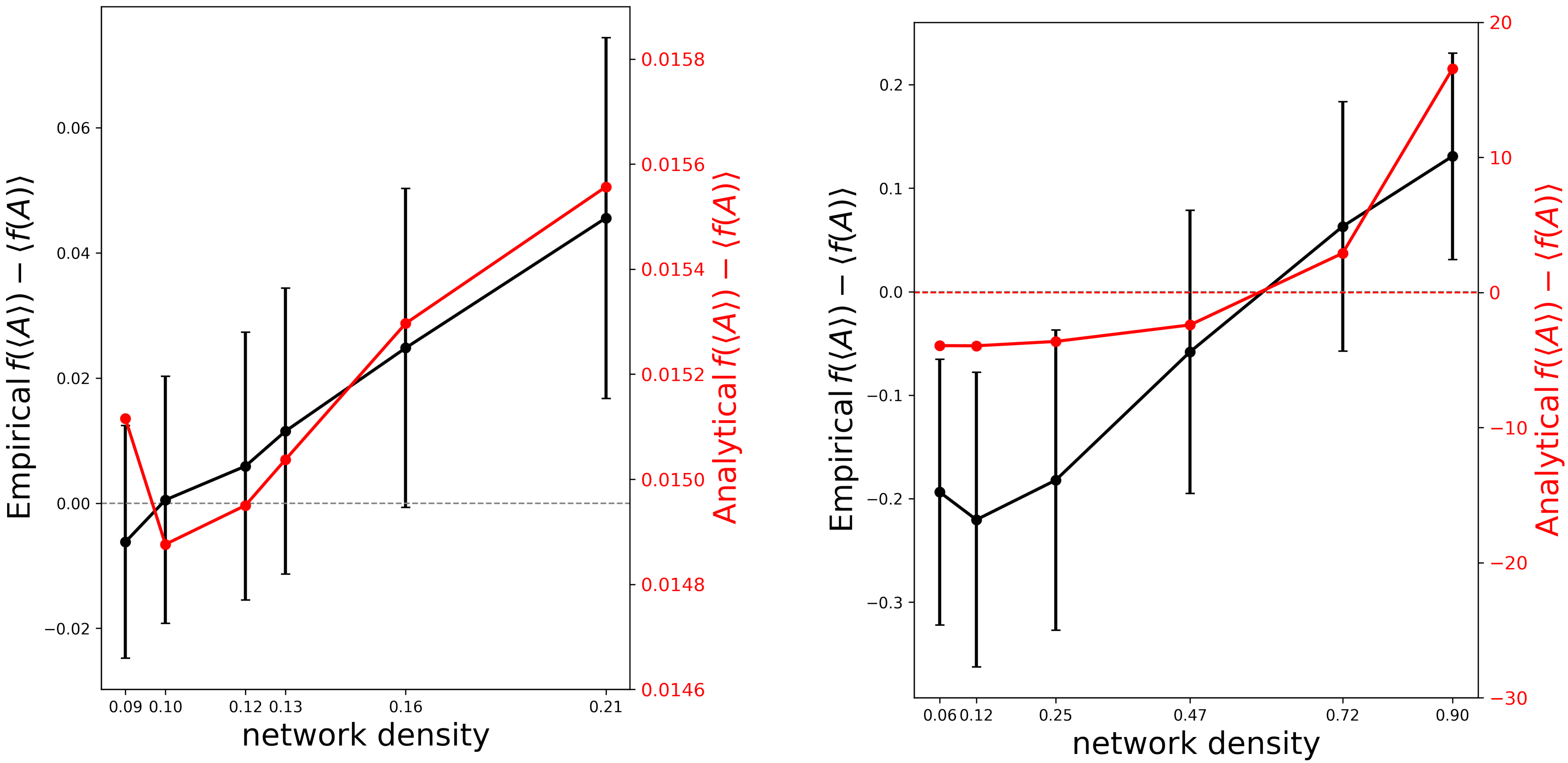
LOG NORMAL DISTRIBUTION AND TAYLOR'S LAW

$$X \sim \mathrm{LogNormal}(\mu, \sigma^2)$$

$$\mathrm{E}[X] = e^{\mu + \frac{\sigma^2}{2}}$$

$$\mathrm{Var}[X] = (e^{\sigma^2} - 1) \cdot e^{2\mu + \sigma^2}$$

$$\mathrm{Var}[X] \propto (\mathrm{E}[X])^b$$



Empirical $f(\langle A \rangle) - \langle f(A) \rangle$
Analytical $f(\langle A \rangle) - \langle f(A) \rangle$
network density
0.09 0.10 0.12 0.13 0.16 0.21
−0.02 0.00 0.02 0.04 0.06
0.0146 0.0148 0.0150 0.0152 0.0154 0.0156 0.0158
0.06 0.12 0.25 0.47 0.72 0.90
−0.3 −0.2 −0.1 0.0 0.1 0.2
−30 −20 −10 0 10 20