

# Systematic assessment of the quality of fit of the stochastic block model for empirical networks

(*PhysRevE.105.054311*)

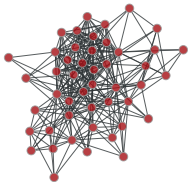
Felipe Vaca-Ramírez & Tiago P. Peixoto

*Central European University  
Vienna, Austria*

NetSci, July 2022

# Assessing the Quality of Fit

Data:  $\mathbf{A}$  →

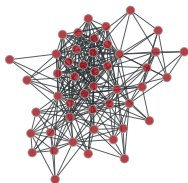


Fit:  $P(\mathbf{b}, \boldsymbol{\theta} | \mathbf{A})$



# Assessing the Quality of Fit

Data:  $\mathbf{A}$  →



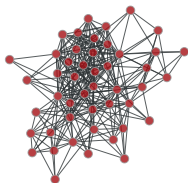
Fit:  $P(\mathbf{b}, \boldsymbol{\theta} | \mathbf{A})$



*What happens if we generate networks and compute descriptors?*

# Assessing the Quality of Fit

Data:  $\mathbf{A}$  →

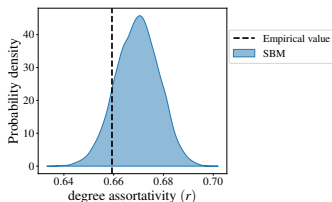


Fit:  $P(\mathbf{b}, \theta | \mathbf{A})$

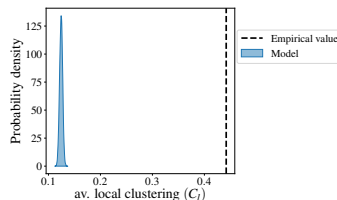


*What happens if we generate networks and compute descriptors?*

Good Agreement

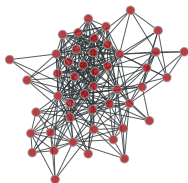


Bad Agreement



# Assessing the Quality of Fit

Data:  $\mathbf{A}$  →

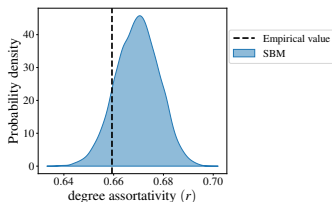


Fit:  $P(\mathbf{b}, \theta | \mathbf{A})$

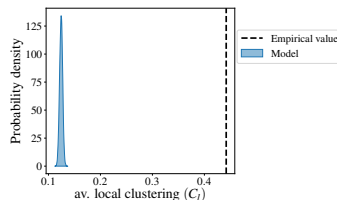


*What happens if we generate networks and compute descriptors?*

Good Agreement



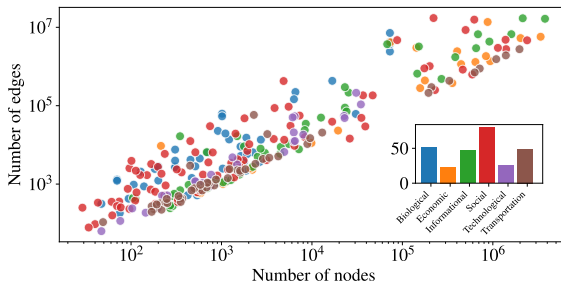
Bad Agreement



*How capable is the SBM in capturing relevant features of empirical networks?*

# Network Corpus

275 real-world networks, structurally diverse



Netzschleuder

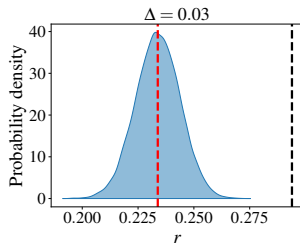
network catalogue, repository and centrifuge

<https://networks.skewed.de/>

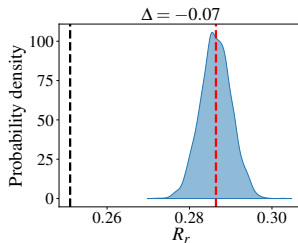
# Which Network descriptors?

Easier...

deg. assortativity

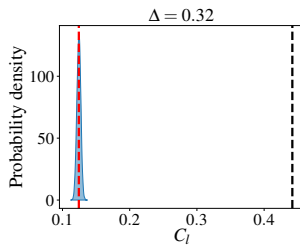


percolation profile

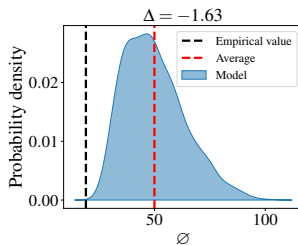


Harder...

av. local clustering



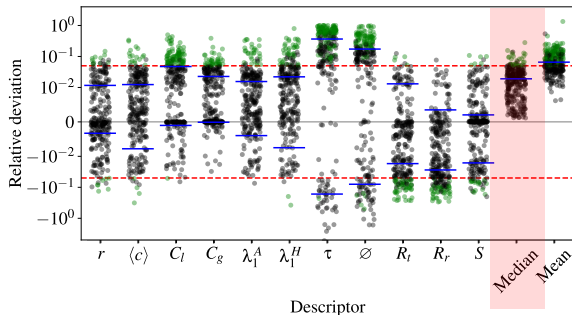
diameter



( $\Delta$ : relative deviation)

# Reproducing descriptors

Wide variety of deviations, but overall good!



— Median values

- -  $|\Delta| = 0.05$

● non-reproduced networks

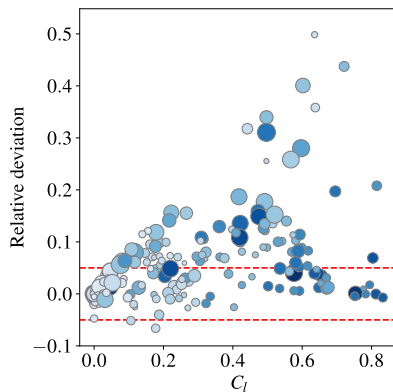
$\tau$ : Characteristic time of a random walk

$\emptyset$ : diameter



# Average local clustering $C_l$

SBM can be a reasonable approximation



discrepancy vs empirical values

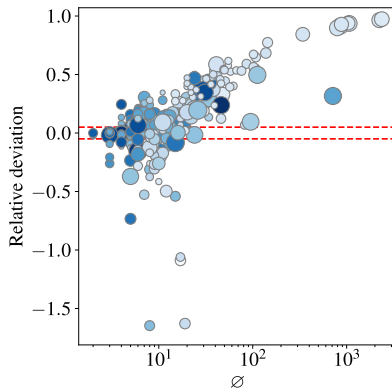
--  $|\Delta| = 0.05$

Marker size:  $\log(\# \text{ edges})$

Marker darkness: mean degree

# Diameter $\varnothing$

SBM is inadequate



discrepancy vs empirical values

--  $|\Delta| = 0.05$

Marker size:  $\log(\# \text{ edges})$

Marker darkness: mean degree

In which directions can we extend the model?

## In which directions can we extend the model?

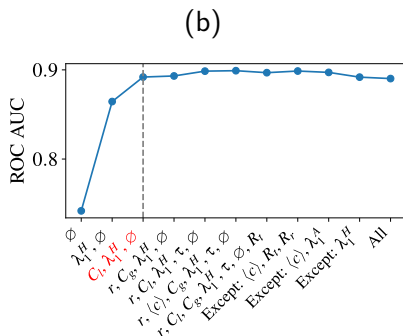
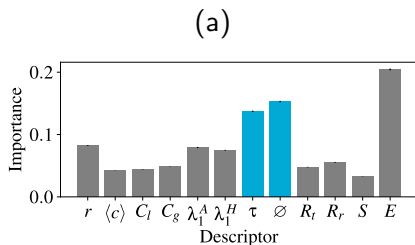
*Predicting the quality of fit of SBM:*

- **Target:**  $y_i = 1$  if network  $i$  is well-described, otherwise  $y_i = 0$ .
- **Features:** empirical values of descriptors.

# In which directions can we extend the model?

*Predicting the quality of fit of SBM:*

- **Target:**  $y_i = 1$  if network  $i$  is well-described, otherwise  $y_i = 0$ .
- **Features:** empirical values of descriptors.



$\tau$ : Characteristic time of a random walk

$\phi$ : diameter

## Concluding remarks

- Overall, SBM offers good description.
- Worst agreements on networks having large diameter and slow-mixing random walks.
- Modelling improvements: diameter, characteristic time of a random walk, and clustering.
- Open-ended analysis.
- More details on:

PhysRevE.105.054311

- Powered by:

