Reinventing the Triangles:

Rule of Thumb for Assessing Detectability

Jeremi K. Ochab, jeremi.ochab@uj.edu.pl
M. Smoluchowski Institute of Physics
Jagiellonian University in Kraków

Aims

Based on the global clustering coefficient (GCC) construct a criterion that decides whether
1) a given network has a significant community structure,
2) a given partition of the network is statistically significant,
3) a given community of the partition is statistically significant

Background

After a phase transition between undetectable and detectable cluster structures had been discovered [1,2], the connection between spectra of adjacency matrices and detectability limits was shown [3]. Both were calculated for networks with arbitrary degree distribution and modular structure. In practice, the full eigenspectrum is not known, and whether a given network has any communities within detectability regime cannot be easily established.

The relation between GCC

$$C = \frac{3 \times \text{number of triangles}}{\text{number of pairs of adjacent edges}}$$

and the 3rd moment of eigspectrum

$$C = \frac{\sum_{l=1}^{N} \lambda_l^3 / N}{\langle k^2 \rangle - \langle k \rangle} \qquad (1)$$

allows for fast [4] checking, whether a network is within detectability regime or not.

Results

The **criteria for detectability** are:

- if condition

$$C \leq C_{\rm uc} \qquad \textbf{(A)}$$

is true, there is no detectable community structure,

- if condition **(A) is false**, but

$$C < \frac{1}{N} \frac{\langle k^2 \rangle^3}{\langle k \rangle^3 (\langle k^2 \rangle - \langle k \rangle)} \qquad \textbf{(B)}$$

is true, either there is no detectable community structure or there are some detectable and some undetectable communities,

- if **(A) and (B) are false**, there is some detectable community structure.

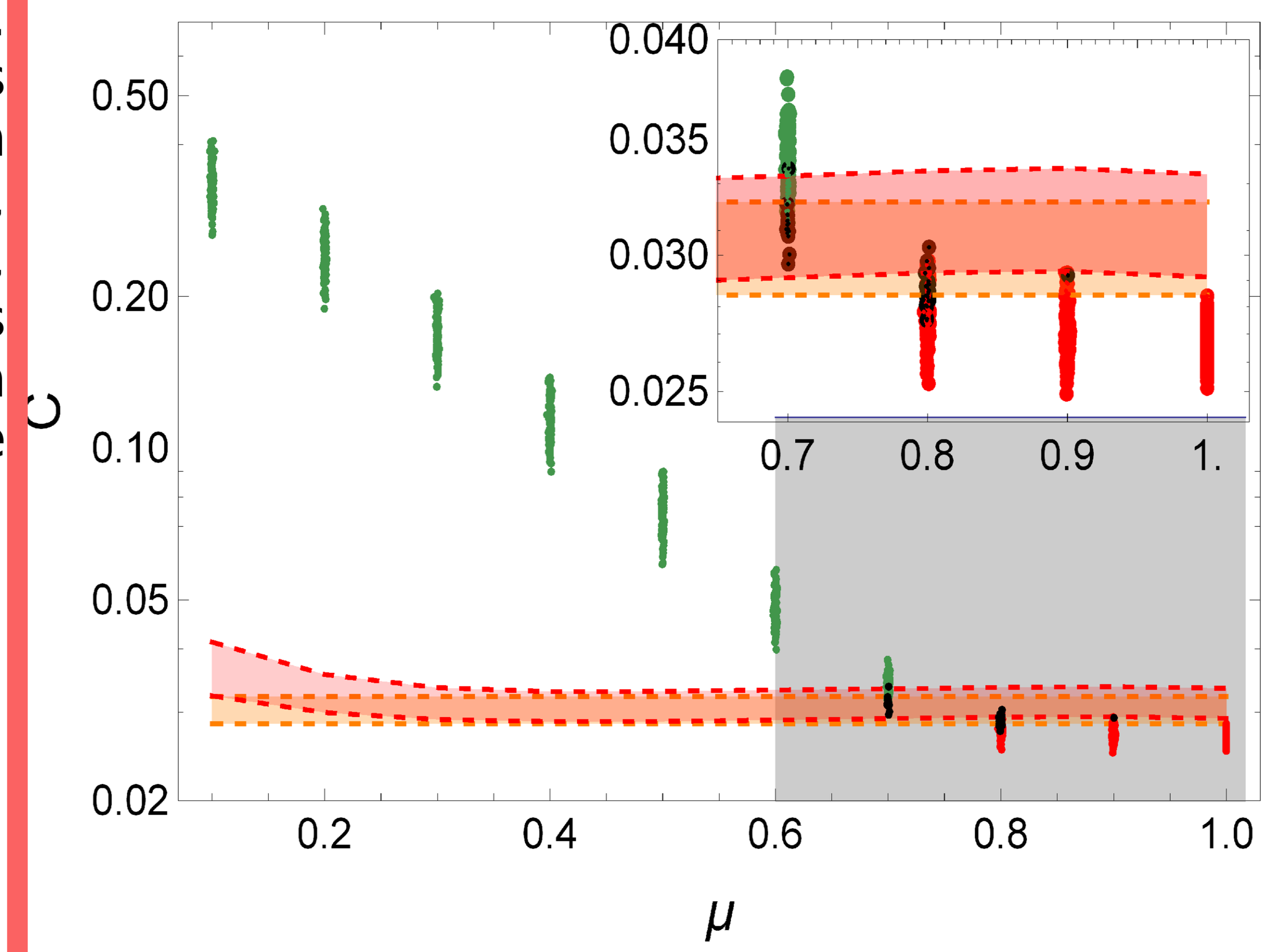

Fig. 1: LFR [7] benchmark. Green = detectable, red = undetectable, black = transient. Grey background shows where community detection algorithms fail.

Assumptions

i. the network is not bipartite,
ii. the network is uncorrelated,

$$C_{\rm uc} \approx \frac{1}{N} \frac{(\langle k^2 \rangle - \langle k \rangle)^2}{\langle k \rangle^3}$$

iii. $\lambda_1 \geq \langle k^2 \rangle / \langle k \rangle - 1$
iv. N is large.

Ad *i.* In graphs which are bipartite, or even partly so—e.g., Watts-Strogatz graphs with small *p*—negative isolated eigenvalues can appear. We assume only positive isolated eigenvalues.
Ad *iii.* Equality and (3) yields *ii.* Hence, breaking *iii.* reverses the order of (A) and (B). Condition (B) comes from (3) and (4).

Facts

- GCC is proportional to the third moment of the eigenspectrum, see (1),
- (empirical) for graphs with a community structure GCC is greater than for random graphs (RG), but it reaches the RG value before its connectivity is fully random, see Fig. 1,
- (empirical) GCC is strictly lower than what can be predicted by the largest eigenvalue alone, see Fig. 2,

$$C < \frac{1}{N} \frac{\lambda_1^3}{\langle k^2 \rangle - \langle k \rangle} \qquad (3)$$

- GCC for the uncorrelated graphs can be approximated by assum. *ii.*, see e.g. [6],
- principal eigenvalue can be approximates by [5]

$$\lambda_1 = (1 + o(1)) \max \left\{ \frac{\langle k^2 \rangle}{\langle k \rangle}, \sqrt{k_{\rm MAX}} \right\} \qquad (4).$$

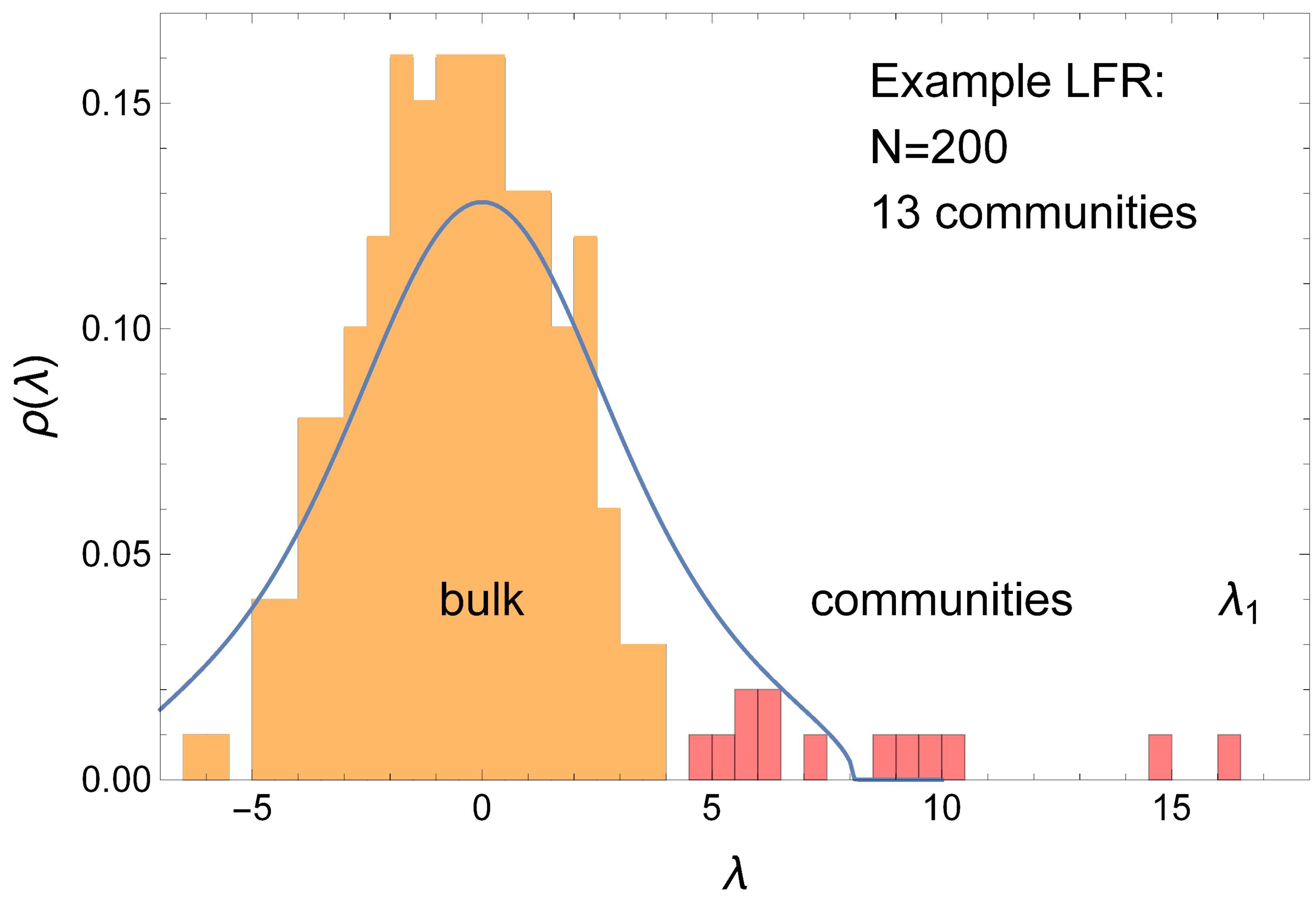

Fig. 2: Histogram of eigenvalues (EVs) of the adjacency matrix. Isolated EVs (right) are associated with community structure. Continuous line predicts [3] bulk spectrum for an uncorrelated graph with identical degree sequence.

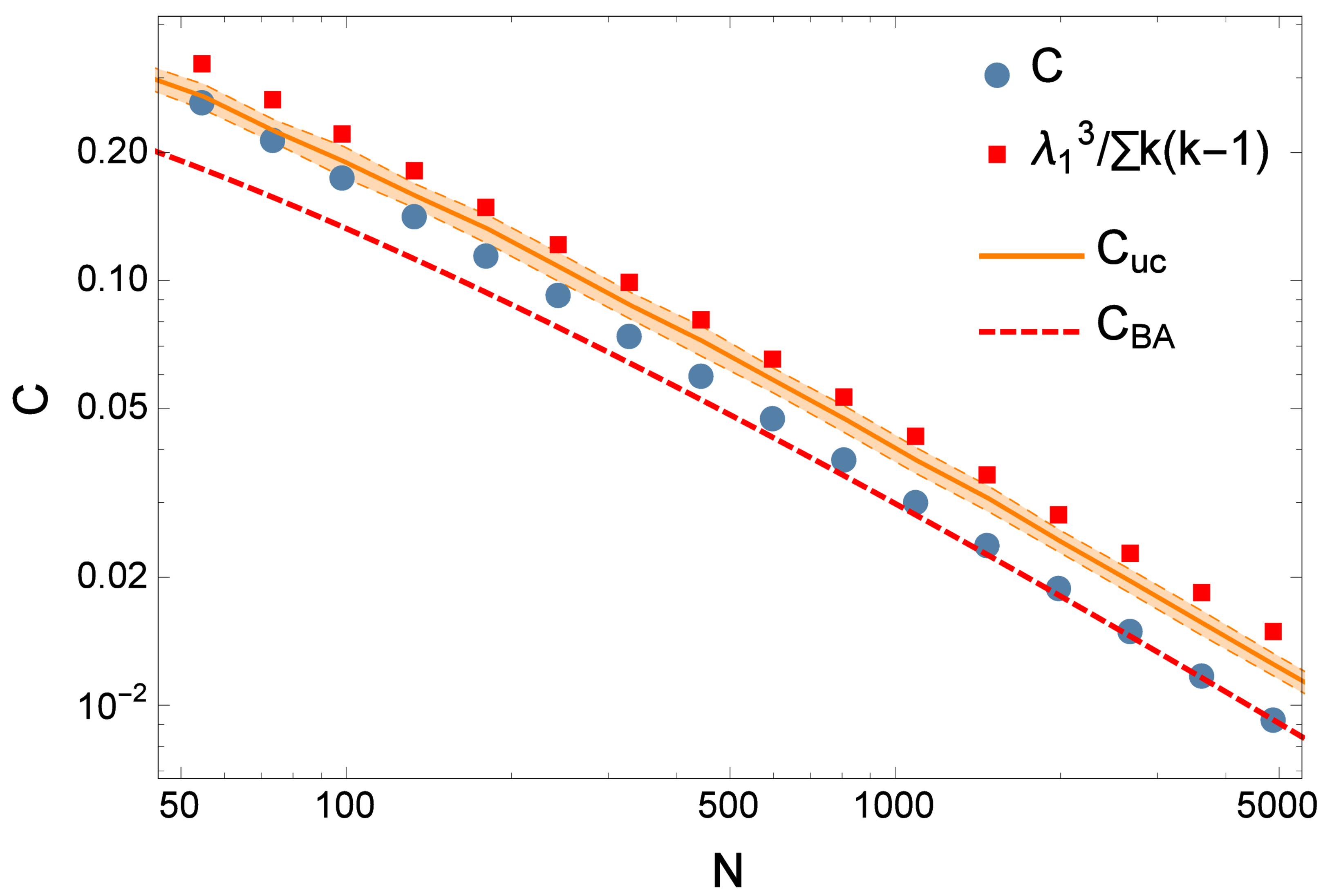

Fig. 3: Clustering coefficients of Barabasi-Albert graphs match the order of conditions (A) and (B): circles are true values; orange line is the uncorrelated network approximation; squares are the contribution of the principal eigenvalue to *C*.

Acknowledgments

My thanks go to Santo Fortunato for the discussions at the early stage of the study and to Zdzisław Burda for consultations throughout. The work has been funded by Grant No. DEC-2013/09/N/ST6/01419 of the National Science Centre of Poland.

National Science Centre Poland

References

[1] J. Reichardt and M. Leone, Phys. Rev. Lett. **101** (2008) 078701.
[2] A. Decelle et al., Phys. Rev. Lett. **107** (2011) 065701.
[3] X. Zhang, R. R. Nadakuditi, and M. E. J. Newman, Phys. Rev. E **89** (2014) 042816.
[4] C. E. Tsourakakis, M. N. Kolountzakis, and G. L. Miller, arXiv:0904.3761 [cs.DS] (2009).
[5] F. Chung, L. Lu, and V. Vu, PNAS **100** (2003) 6313.
[6] Z. Burda, J. Jurkiewicz, and A. Krzywicki, Phys. Rev. E **69** (2004) 026106.
[7] A. Lancichinetti, S. Fortunato, and F. Radicchi, Phys. Rev. E **78** (2008) 046110.