Surprise maximization reveals

the community structure of complex networks

Rodrigo Aldecoa and Ignacio Marín

SUPPLEMENTARY INFORMATION

Supplementary Table S1. Cases where $S_{max} > S_{orig}$ in the RC benchmarks with R = 10 %

Small differences between S_{max} and S_{orig} are due to the rapid degradation of small communities. In most cases, several algorithms find the S_{max} partition instead of the original one, strongly supporting the idea that the community structure has actually changed.

No. of algorithms	Algorithms	S_{max}	S_{orig}	Differences
4	CPM, RN, RNSC, SCluster	18616.55	18615.55	A community of 2 nodes is split into two
4	CPM, RN, RNSC, SCluster	18881.92	18879.94	A community of 3 nodes is split into two: 2 nodes + 1 node
4	CPM, RN, RNSC, SCluster	18442.72	18440.74	Two communities of 2 nodes are split into two
4	CPM, RN, RNSC, SCluster	19089.77	19088.78	A community of 2 nodes is split into two
4	CPM, RN, RNSC, SCluster	19187.13	19186.13	A community of 2 nodes is split into two
3	CPM, RN, SCluster	18312.46	18312.11	A community of 4 nodes is divided into two: 3 + 1 (displayed in Figure 4)
3	CPM, RN, SCluster	19897.8	19896.81	A community of 2 nodes is split into two
3	CPM, RN, SCluster	17980.46	17979.13	A community of 5 nodes is split into two: 4 + 1
3	CPM, RN, Scluster	17992.87	17991.54	A community of 5 nodes is split into two: 4 + 1
3	CPM, RN, SCluster	19579.76	19578.77	A community of 2 nodes is split into two
3	CPM, RN, SCluster	18008.52	18005.89	A community of 2 nodes is split into two and a community of 3 nodes is split into two: 2 + 1
3	CPM, RN, SCluster	18835.32	18834.33	A community of 2 nodes is split into two
2	CPM, SCluster	17803.45	17803.14	A community of 2 nodes is split into two and a community of 4 nodes is split into two: 3 + 1
2	CPM, SCluster	19928.95	19927.51	A community of 4 nodes is split into two: 2 + 2
2	CPM, SCluster	17749.38	17748.06	A community of 4 nodes is split into two: 2 + 2
2	RN, SCluster	18295.54	18295.19	A community of 4 nodes is split into two: 3 + 1
1	SCluster	18685.24	18684.48	A community of 5 nodes is split into two: 3 + 2
1	SCluster	19122.29	19121.88	A community of 4 nodes is split into two: 3 + 1
1	SCluster	18837.36	18836.58	A community of 7 nodes is split into two: 6 + 1
1	SCluster	18853.16	18852.39	A community of 5 nodes is split into two: 3 + 2
1	SCluster	19285.36	19284.95	A community of 4 nodes is split into two: 3 + 1
1	SCluster	18875.51	18875.13	A community of 4 nodes is split into two: 3 + 1
1	CPM	18141.17	18139.54	Two communities of 3 nodes are restructured in two communities of 4 and 2 nodes

Supplementary Table S2. Algorithms not included in our study

Name	Strategy used by the algorithm	Reference	Reasons for not including the algorithm
AFG	Multiresolution Potts Model	Arenas, A., Fernandez, A. & Gomez, S. *New Journal of Physics* **10**, 23 (2008).	Ambiguous choice of the best partition. Too slow for good modularity optimization heuristics in our benchmarks
EM	Maximum Likelihood	Ball, B., Karrer, B. & Newman, M.E.J. *Phys. Rev. E* **84**, 036103 (2011)	Needs initialization. Not every nodes are assigned to a single cluster
HQcut	Multilevel modularity maximization	Ruan, J. & Zhang, W. *Phys. Rev. E* **77**, 016104 (2007).	Unable to complete all the analyses
iMod	Modularity maximization	Xu, G., Bennett, L., Papageorgiou, L.G. & Tsoka, S. *Algorithms Mol. Biol.* **5**, 36 (2010).	According to the authors, only they can run the algorithm, given its particular platform and software dependencies
Infomod	Information compression	Rosvall, M. & Bergstrom, C.T. *Proc. Natl. Acad. Sci. USA* **104**, 7327 (2007).	Unable to complete all the analyses
QMC	Qualified Min-Cut	Zhang, X.-S., Li, Z., Wang, R.-S. & Wang, Y. *J. Comb. Optim.* **23**, 425-442 (2010).	Authors did not answer our request
Random walks	Consensus hierarchical clustering	Steinhaeuser, K. & Chawla, N.V. *Pattern Recogn Lett* **31**: 413–421 (2010)	The number of communities must be specified *a priori*