# Modularity in several random graph models

Liudmila Ostroumova Prokhorenkova,
Paweł Prałat, Andrei Raigorodskii

One important property of many complex networks is their community structure, that is, the organization of vertices in clusters, with many edges joining vertices of the same cluster and comparatively few edges joining vertices of different clusters [2, 3]. In social networks communities may represent groups by interest, in citation networks they correspond to related papers, in the Web communities are formed by pages on related topics, etc.

*Modularity* is at the same time a global criterion to define communities, a quality function of community detection algorithms, and a way to measure the presence of community structure in a network. Modularity was first introduced by Newman and Girvan in [6]. Since then, many popular and applied algorithms used to find clusters in large data-sets are based on finding partitions with high modularity [1, 4, 5].

The main idea behind modularity is to compare the actual density of edges inside communities with the density one would expect to have if the vertices of the graph were attached at random, regardless of community structure. Formally, for a given partition $\mathcal{A} = \{A_1, \ldots, A_k\}$ of the vertex set $V(G)$, let

$$q_{\mathcal{A}} = \sum_{A \in \mathcal{A}} \left( \frac{e(A)}{|E(G)|} - \frac{(\sum_{v \in A} \deg(v))^2}{4|E(G)|^2} \right), \tag{1}$$

where $e(A) = |\{uv \in E(G) : u, v \in A\}|$ is the number of edges in the graph induced by the set $A$. The first term, $\sum_{A \in \mathcal{A}} \frac{e(A)}{|E(G)|}$, is called the *edge contribution*, whereas the second one, $\sum_{A \in \mathcal{A}} \frac{(\sum_{v \in A} \deg(v))^2}{4|E(G)|^2}$, is called the *degree tax*. It is easy to see that $q_{\mathcal{A}}$ is always smaller than one. Also, if $\mathcal{A} = \{V(G)\}$, then $q_{\mathcal{A}} = 0$. The *modularity* $q^*(G)$ is

$$q^*(G) = \max_{\mathcal{A}} q_{\mathcal{A}}(G).$$

If $q^*(G)$ approaches 1 (which is the maximum), we observe a strong community structure; conversely, if $q^*(G)$ is close to zero, we are given a graph with no community structure.

Unfortunately, modularity is not a well studied parameter for the existing random graph models, at least from a rigorous, theoretical point of view. In this work, we investigate modularity in random $d$-regular graphs, the preferential attachment model, and the spatial preferential attachment model. The detailed results can be found in [7].

## References


[1] A. Clauset, M.E.J. Newman, C. Moore, Finding community structure in very large networks, Phys. Rev. E 70 (2004)

[2] S. Fortunato, Community detection in graphs, Physics Reports, vol. 486(3-5) (2010), 75–174.

[3] M. Girvan and M. E. Newman, Community structure in social and biological networks, Proceedings of the National Academy of Sciences, 99(12) (2002), 7821–7826.

[4] A. Lancichinetti and S. Fortunato, Limits of modularity maximization in community detection, Phys. Rev. E 84 (2011) 066122.

[5] M.E.J. Newman, Fast algorithm for detecting community structure in networks, Phys. Rev. E 69 (2004) 066133.

[6] M.E.J. Newman and M. Girvan, Finding and evaluating community structure in networks, Phys. Rev. E 69 (2004) 026–113.

[7] L. Ostroumova Prokhorenkova, P. Prałat, and A. Raigorodskii, Modularity of complex networks models, Proceedings of the 13th Workshop on Algorithms and Models for the Web Graph (WAW 2016), Lecture Notes in Computer Science 10088, Springer, 2016, 115–126.