Modularity in several random graph models

Liudmila Ostroumova Prokhorenkova [1,3]

Advanced Combinatorics and Network Applications Lab
Moscow Institute of Physics and Technology
Moscow, Russia

Paweł Prałat [2,4]

Department of Mathematics
Ryerson University
Toronto, Canada

Andrei Raigorodskii [1,5]

Advanced Combinatorics and Network Applications Lab
Moscow Institute of Physics and Technology
Moscow, Russia

Abstract

Modularity is a graph characteristic which measures the strength of division of a network into clusters (or communities). Networks with high modularity usually have distinguishable clusters with dense connections between the vertices within clusters and sparse connections between vertices of different clusters. In this paper, we investigate the value of modularity in several well-known random graph models.

Keywords: modularity, random d-regular graphs, preferential attachment, spatial preferential attachment.

---

1 Modularity

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 [6]. 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* [14] 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. Many community detection algorithms are based on finding partitions with high modularity [4,8].

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

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

where $e(A) = |\{uv \in E(G) : u, v \in A\}|$ is the number of edges in the graph induced by the set A. Note that $q_{\mathcal{A}}$ is always smaller than one. Also, if $\mathcal{A} = \{V(G)\}$, then $q_{\mathcal{A}} = 0$. The *modularity* of a graph 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 paper, we investigate modularity in random d-regular graphs, the preferential attachment model [2], the average degree graphs, and the spatial preferential attachment model [1]. Due to space constraints we omit the proofs of the theorems here, the complete proofs can be found in [15].

[1] Supported by Russian Science Foundation, grant number 16-11-10014.
[2] Supported in part by NSERC, The Tutte Institute for Mathematics and Computing, and Ryerson University.
[3] Email: ostroumova-la@yandex.ru
[4] Email: pralat@ryerson.ca
[5] Email: mraigor@yandex.ru

Table 1
Upper bounds U_1 and U_2 for $q^*(\mathcal{G}_{n,d})$

d	3	4	5	6	7	8	9	10
U_1	0.9386	0.8900	0.8539	0.8261	0.8038	0.7855	0.7702	0.7570
U_2	0.8038	0.6834	0.6024	0.5435	0.4984	0.4624	0.4330	0.4083

2 Random d-regular graphs

In this section we consider the probability space of *random d-regular graphs* with uniform probability distribution. This space is denoted $\mathcal{G}_{n,d}$, and asymptotics are for $n \to \infty$ with $d \geq 2$ fixed, and n even if d is odd.

2.1 Lower bounds

Let us briefly discuss the following known lower bound for the modularity of $\mathcal{G}_{n,d}$. It is known that a.a.s. for any $d \in \mathbb{N} \setminus \{1,2\}$, $\mathcal{G}_{n,d}$ is Hamiltonian. Authors of [9] propose breaking this cycle into $\lceil \sqrt{n} \rceil$ paths of length at most $\lceil \sqrt{n} \rceil$. This partition gives, a.a.s., $q^*(\mathcal{G}_{n,d}) \geq \frac{2}{d} - O(1/\sqrt{n}) = \frac{2+o(1)}{d}$.

Also, as pointed out in [10], there exists a universal constant $c > 0$ such that a.a.s. $q^*(\mathcal{G}_{n,d}) \geq c/\sqrt{d}$.

2.2 Upper bounds

Known upper bound.

The following upper bound was obtained by McDiarmid and Skerman in [9].

Theorem 2.1 *A.a.s.* $q^*(\mathcal{G}_{n,d}) \leq U_1(d) := \max\{1/2 + \eta/2, 3/4\}$, *where* $0 < \eta < 1$ *is such that* $2^{4/d} < (1-\eta)^{1-\eta}(1+\eta)^{1+\eta}$.

The numerical values for U_1 are presented in Table 1. Investigating random d-regular graphs continues in [10], a very recent paper. In fact, some of our results for this model mentioned below are obtained independently there.

Numerical upper bound.

For a given $d \in \mathbb{N} \setminus \{1,2\}$, let

$$\begin{aligned} f(x,y,d) := {} & x(y/2-1)\log(x) + (1-x)(d-1)\log(1-x) + d\log(d)/2 \\ & - xy\log(y)/2 - x(d-y)\log(d-y) - (d-2xd+xy)\log(d-2xd+xy)/2. \end{aligned}$$

Let $y_3 = y_3(x,d)$ be the largest value of $y \in (0,d)$ such that $f(x,y,d) = 0$ (one can easily show that such a value exists). And let $U_2 = U_2(d) := \sup_{x\in(0,1)} \left(\frac{y_3(x,d)}{d} - x\right)$. The following theorem holds [15].

Theorem 2.2 *Let $d \in \mathbb{N} \setminus \{1,2\}$ and $\varepsilon > 0$ be an arbitrarily small constant. Then a.a.s. $q^*(\mathcal{G}_{n,d}) \leq U_2 + \varepsilon/d$, where $U_2 = U_2(d)$ is defined as above.*

The numerical values of U_2 for small values of d can be found in Table 1.

Explicit upper bound.

Theorem 2.2 provides an upper bound that can be easily numerically computed for a given $d \in \mathbb{N} \setminus \{1,2\}$. Now we present a weaker but an explicit bound that can be obtained using the expansion properties of random d-regular graphs that follow from their eigenvalues.

Theorem 2.3 *Let $d \in \mathbb{N} \setminus \{1,2\}$ and $\varepsilon > 0$ be an arbitrarily small constant. Then, a.a.s. $q^*(\mathcal{G}_{n,d}) \leq \frac{2}{\sqrt{d}}$.*

3 Constant average degree graphs

In this section we analyze graphs with constant average degree. We extend the results of [13], where it was proven that trees with maximum degree $\Delta = o(\sqrt[5]{n})$ have asymptotic modularity 1. First, we relax the condition on maximum degree; second, we allow our graphs to be disconnected [15].

Theorem 3.1 *Let $\{F_n\}$ be a sequence of forests, F_n is a forest on n vertices with no isolated ones and $\Delta = \Delta(F_n) = o(n)$. Then $q^*(F_n) \geq 1 - O\left(\sqrt{\frac{\Delta}{n}}\right) = 1 - o(1)$ as $n \to \infty$.*

The assumption $\Delta = o(n)$ cannot be eliminated, since the asymptotic modularity of trees with $\Delta = \Omega(n)$ is strictly less than 1 [13].

For graphs with bounded average degree the following theorem holds [15].

Theorem 3.2 *Let $\{G_n\}$ be a sequence graphs, G_n is a connected graph on n vertices with the average degree $\frac{2|E(G_n)|}{n} \leq D$ for some constant D, and $\Delta = \Delta(G_n) = o(n)$. Then $q^*(G_n) \geq \frac{2}{D} - O\left(\sqrt{\frac{\Delta}{n}}\right) = \frac{2}{D} - o(1)$.*

4 Preferential Attachment model

The *Preferential Attachment* (PA) model, introduced by Barabási and Albert [2], was an early stochastic model of complex networks. The idea is that

Table 2
Lower bounds for $q^*(G_m^n)$

m	7	8	9	10	100	1000
L_1	0.156	0.136	0.136	0.123	0.0397	0.0126
L_2	0.142	0.125	0.111	0.100	0.0100	0.0010

at each step a new vertex is added together with m edges connecting this vertex to m previous vertices, the probability to choose a previous vertex is proportional to its current degree. The precise definition of the model (which is denoted by G_m^n) is given in Bollobás and Riordan in [3].

4.1 Lower bound

The following theorem easily follows from Theorems 3.1 and 3.2 and the fact that a.a.s. $\Delta(G_m^n) = O\left(n^{\frac{1}{2}+2\varepsilon}\right)$ for any $\varepsilon > 0$.

Theorem 4.1 *For any $\varepsilon > 0$ a.a.s. $q^*(G_m^n) \geq \frac{1}{m} - O\left(n^{-1/4+\varepsilon}\right) = \frac{1}{m} - o(1)$.*

As in the case of random d-regular graphs, it is natural to conjecture that the above lower bound is not sharp. Indeed, we can prove the following, stronger, lower bound.

Theorem 4.2 *A.a.s. $q^*(G_m^n) \geq \mathbf{E}\left(\left|\mathrm{Bin}(m, 1/2) - m/2\right|\right)/m + o(1)$. That is, a.a.s.*

$$q^*(G_m^n) \geq \begin{cases} (2^{1-m}/m)\sum_{i=1}^{m/2} i\binom{m}{m/2+i} & \textit{if } m \textit{ is even,} \\ (2^{1-m}/m)\sum_{i=1}^{(m+1)/2}(i-1/2)\binom{m}{(m-1)/2+i} & \textit{if } m \textit{ is odd,} \end{cases}$$

In particular, a.a.s. $q^*(G_m^n) = \Omega\left(1/\sqrt{m}\right)$.

Table 2 presents numerical values for the lower bounds $L_1 = L_1(m) = 1/m$ from Theorem 4.1 and $L_2 = L_2(m)$ from Theorem 4.2 for a few values of m. Note that L_2 is weaker for $m \leq 6$ and stronger for larger values.

4.2 Upper bound

The *edge expansion* of a graph G is defined as $\rho = \min_{S \subset V(G), |S| \leq |V|/2} \frac{e(S, V\setminus S)}{|S|}$. In [12] it was shown that for any $\varepsilon > 0$ we have that a.a.s. $\rho(G_m^n) \geq \frac{m}{2} - \frac{3+\varepsilon}{4}$. Using this observation one can easily obtain the following non-trivial upper bound for $q^*(G_m^n)$.

Theorem 4.3 *For any $\varepsilon > 0$ a.a.s. $q^*(G_2^n) \leq \frac{15+\varepsilon}{16}$. Moreover, for any $m \geq 3$ a.a.s. $q^*(G_m^n) \leq \frac{15}{16}$.*

5 Spatial Preferential Attachment model

5.1 Definition

Let $S = [0,1]^m$ be the unit hypercube in $\mathbb{R}^m$, equipped with the torus metric derived from any of the L_p norms. The parameters of the model consist of the *link probability* $p \in [0,1]$, and two positive constants A_1 and A_2, which, in order to avoid the resulting graph becoming too dense, must be chosen so that $pA_1 < 1$. The SPA model generates stochastic sequences of directed graphs $(G_t : t \geq 0)$, where $G_t = (V_t, E_t)$, and $V_t \subseteq S$. Let $\deg^-(v,t)$ be the in-degree of the vertex v in G_t, and $\deg^+(v,t)$ its out-degree. The *sphere of influence* $S(v,t)$ of the vertex v at time $t \geq 1$ is the ball centered at v with volume $|S(v,t)| = \min\left\{\frac{A_1 \deg^-(v,t)+A_2}{t}, 1\right\}$.

The process begins at $t = 0$, with G_0 being the null graph. At each step t, a new vertex v_t is chosen *uniformly at random* from S, and added to V_{t-1} to create V_t. Next, independently, for each vertex $u \in V_{t-1}$ such that $v_t \in S(u, t-1)$, a directed link (v_t, u) is created with probability p.

The SPA model produces scale-free networks, which exhibit many of the characteristics of real-life networks [1,5]. In [7], it was shown that the SPA model gave the best fit, in terms of graph structure, for a series of social networks derived from Facebook.

5.2 Results

As the modularity is defined for undirected graphs, we consider $\hat{G}_n$ that is a graph obtained from G_n by replacing each directed edge (u,v) by undirected edge uv (note that $\hat{G}_n$ is always a simple graph).

We use the following properties of the SPA model to estimate the modularity. A.a.s. for every pair i,t such that $1 \leq i \leq t \leq n$ we have that $\deg^-(v_i,t) = O\Big((t/i)^{pA_1}\log^2 n\Big)$, $\deg^+(v_i,t) = O\Big(\log^2 n\Big)$, and also $|E(G_n)| = \Theta(n)$.

The following theorem shows that modularity of the SPA model is asymptotically one, unlike d-regular and preferential attachment graphs.

Theorem 5.1 *Let $p \in (0,1]$, $A_1, A_2 > 0$, and suppose that $pA_1 < 1$. Then, a.a.s., $q^*(\hat{G}_n) = 1 - O\left(n^{\max\{-1/m,-1+pA_1\}/2}\log^{9/2} n\right) = 1 - o(1)$.*

References

[1] Aiello, W., A. Bonato, C. Cooper, J. Janssen, and P. Prałat, A spatial web graph model with local influence regions, Internet Mathematics 5 (2009) 175–196.

[2] Barabási, A.L., R. Albert, Emergence of scaling in random networks, *Science* **286** (1999) 509–512.

[3] Bollobás, B., O. Riordan, The diameter of a scale-free random graph, *Combinatorica* **24** (2004) 5–34.

[4] Clauset, A., M.E.J. Newman, C. Moore, Finding community structure in very large networks, Phys. Rev. E 70 (2004) 066111.

[5] Cooper, C., A. Frieze, and P. Prałat, Some typical properties of the Spatial Preferred Attachment model, Internet Mathematics 10 (2014) 27–47.

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

[7] Janssen, J., M. Hurshman, and N. Kalyaniwalla. Model selection for social networks using graphlets. Internet Mathematics, 8(4) (2013) 338–363.

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

[9] McDiarmid, C. and F. Skerman, Modularity in random regular graphs and lattices, Electronic Notes in Discrete Mathematics 43 (2013) 431–437.

[10] McDiarmid, C. and F. Skerman, Modularity of tree-like and random regular graphs, preprint.

[11] Dorodnykh, A., L. Ostroumova Prokhorenkova, E. Samosvat, Preferential placement for community structure formation, Proc. WAW, LNCS (2017).

[12] Mihail, M., C. Papadimitriou and A. Saberi, On certain connectivity properties of the Internet topology, Proc. Conf. on Foundations of Computer Science (2003) 28–35.

[13] Montgolfier, F. De, M. Soto, and L. Viennot, Asymptotic modularity of some graph classes, ISAAC (2011) 435–444.

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

[15] Ostroumova Prokhorenkova, L., P. Prałat, and A. Raigorodskii, Modularity of complex networks models, Proceedings WAW, LNCS 10088 (2016) 115–126.