

Modularity and random graphs

Colin McDiarmid^{*1} and Fiona Skerman^{†2}

¹Department of Statistics, University of Oxford.

²Department of Mathematics, Uppsala University.

This work will appear as a chapter in a forthcoming volume titled ‘Topics in Probabilistic Graph Theory’. For a given graph G , each partition of the vertices has a modularity score, with higher values indicating that the partition better captures community structure in G . The modularity $q^(G)$ of G is the maximum over all vertex-partitions of the modularity score, and satisfies $0 \leq q^*(G) < 1$. Modularity lies at the heart of the most popular algorithms for community detection. In this chapter we discuss the behaviour of the modularity of various kinds of random graphs, starting with the binomial random graph $G_{n,p}$ with n vertices and edge-probability p .*

1. Introduction
 2. Modularity: definition and some properties
 3. The modularity of binomial random graphs $G_{n,p}$
 4. The modularity of random regular graphs
 5. Modularity in small world models
 6. Modularity and partially observed graphs
 7. Modularity and stochastic block models
 8. The modularity of random graphs embeddable in a surface
 9. Summary of known modularity values
- References

1. Introduction

The ever-increasing amount of network data available in many fields has led to great interest in understanding network structure. In particular we want to be able to identify whether a network can be decomposed into dense clusters or ‘communities’. Modularity was introduced in 2004 by Newman and Girvan [40], to give a measure of how well a graph can be divided into communities, and it now forms the backbone of the most popular algorithms used to cluster real data [25], including the algorithms Louvain [5] and Leiden [48]. Here a ‘community’ is a collection of vertices which are more densely interconnected than one would expect from looking at vertex-degrees.

^{*}Email: cmcd@stats.ox.ac.uk.

[†]Email: fiona.skerman@math.uu.se.

There are many applications, including protein discovery, gene co-expression, and identifying connections between websites; see [17], [24] and [41] on the use of modularity for community detection in networks. For example, the paper [14] studied the choice of words used to discuss climate change online. Its authors collated about 3,000 blogs and, as a preliminary step, produced a graph with a vertex for each blog and an undirected edge for each link between blogs; and then, using the Louvain algorithm [5] (based on modularity), found a partition of these blogs. The authors then analysed how language use varied between these groups.

In the next section we give definitions and notation for modularity, and some basic properties; and we then discuss modularity for several different models of random graphs. In Section 3 we discuss the binomial random graph $G_{n,p}$, where edges appear independently with probability p , and Section 4 concerns random regular graphs. In Section 5 we discuss three ‘small-world’ models of random graph: the preferential attachment model, the spatial preferential attachment model, and random graphs in the hyperbolic plane. In Section 6 we discuss partially observed or ‘percolated’ graphs: mostly we assume that there is a large (unknown) underlying graph G , and the edges of G appear independently with some probability p , forming the observed graph G_p : what can we say about the modularity $q^*(G)$? Section 7 concerns stochastic block models, where there is a ‘planted partition’ which should give a good community structure. In Section 8 we consider random planar graphs, and more generally random graphs embeddable in a given surface. Finally, Section 9 gives a summary of known modularity results for different classes of graphs.

2. Modularity: definition and some properties

Given a graph $G = (V, E)$, modularity gives a score to each partition of the vertex-set V , and the (maximum) modularity $q^*(G)$ of G is the maximum of these scores over all vertex-partitions. Let $\mathbf{1}_{vw \in E}$ be the indicator that vw is an edge. For a set A of vertices, let $e(A)$ (or $e_G(A)$, if needed for clarity) be the number of edges within A , and let the volume $\text{vol}(A)$ be the sum of the degrees d_v over the vertices v in A . Let G be a graph with $m \geq 1$ edges. Then, for a vertex-partition $\mathcal{A}$ of G , the *modularity score* of $\mathcal{A}$ on G is

$$q_{\mathcal{A}}(G) = \frac{1}{2m} \sum_{A \in \mathcal{A}} \sum_{v, w \in A} \left(\mathbf{1}_{vw \in E} - \frac{d_v d_w}{2m} \right) = \frac{1}{m} \sum_{A \in \mathcal{A}} e(A) - \frac{1}{4m^2} \sum_{A \in \mathcal{A}} \text{vol}(A)^2. \quad (2.1)$$

The *modularity* of G is $q^*(G) = \max_{\mathcal{A}} q_{\mathcal{A}}(G)$, where the maximum is taken over all vertex-partitions $\mathcal{A}$. (For a graph G with no edges, where $m = 0$, we set $q_{\mathcal{A}}(G) = 0$ for each vertex-partition $\mathcal{A}$, and $q^*(G) = 0$.) Taking account of the ‘strength’ of connections can be important, and the definition extends naturally to weighted graphs, as we shall see in Section 6 – note in particular equation (6.1).

It follows directly from the definition that $0 \leq q^*(G) < 1$ for all graphs G . Note that isolated vertices are irrelevant – they are not counted in the formula for the modularity score – and so we may assume without further comment that our graphs have no isolated vertices. The second expression for $q_{\mathcal{A}}(G)$ expresses it as the difference between the *edge contribution* or *coverage* $q_{\mathcal{A}}^E(G) = (1/m) \sum_{A \in \mathcal{A}} e(A)$, and the *degree tax* $q_{\mathcal{A}}^D(G) = (1/4m^2) \sum_{A \in \mathcal{A}} \text{vol}(A)^2$. By the Cauchy–Schwarz inequality, if $x_1, x_2, \dots, x_k \geq 0$, then $\sum_i x_i^2 \geq (\sum_i x_i)^2/k$: thus, if the partition $\mathcal{A}$ has k parts, then $q_{\mathcal{A}}^D(G) \geq 1/k$, and so $q_{\mathcal{A}}(G) \leq 1 - 1/k$.

The original rationale for the definition [40] was that while rewarding the partition for capturing edges within the parts, we should penalise by (approximately) the expected number of such edges.

We can see this in two different ways, given a suitable degree-sequence $\mathbf{d} = (d_1, d_2, \dots, d_n)$. By using the configuration model (see for example [20] and Chapter 4), we can generate nearly uniformly at random a multigraph R on $[n] = \{1, 2, \dots, n\}$ with degree-sequence $\mathbf{d}$ (where multiple edges and loops are allowed); and then the expected number of edges in R between distinct vertices v and w is $d_v d_w / (2m - 1)$, where $\sum_v d_v = 2m$. If instead we sample G uniformly at random from all (simple) graphs with degree-sequence $\mathbf{d}$, then the probability that vw is an edge is very close to $d_v d_w / 2m$ – see also Theorem 2.3 at the end of this section. Also, see [36] for a fuller story concerning the degree tax.

We shall discuss the modularity $q^*(G)$ of various sorts of random graph G . This is very different from considering the modularity of a random partition of a fixed graph. For let G be any fixed graph (with at least one edge). Fix $k \geq 2$, and suppose that we generate a random partition $\mathcal{A}$ of the vertices by placing each vertex into one of k parts independently with probability $1/k$. Then, as noted in the concluding remarks of [33],

$$\mathbb{E}[q_{\mathcal{A}}(G)] < 0. \quad (2.2)$$

Thus, comparing the modularity of a partition which we have found to that of a random partition is likely to give a false positive. To see why (2.2) holds, observe that, for an edge e in G , the probability that both endvertices are placed in the same part is $1/k$, so $\mathbb{E}[q_{\mathcal{A}}^E(G)] = 1/k$; and the degree tax is always at least $1/k$, and sometimes is larger (see just before Theorem 2.2 below).

As Reichardt and Bornholdt [43] have remarked: ‘A differentiation between graphs which are truly modular and those which are not can ... only be made if we gain an understanding of the intrinsic modularity of random graphs.’ In the following sections we discuss modularity for several different models of random graphs.

A basic observation is that, given a graph G without isolated vertices, in each optimal partition $\mathcal{A}$, each part has size at least 2 [38] and induces a connected subgraph of G . Also, $V(H)$ is a part of $\mathcal{A}$ for each connected component H of G with modularity $q^*(H) = 0$ (see [29, Corollary 1.5], and see also Theorem 1.4 in that paper for a fuller account). For example, all complete graphs, ‘nearly complete’ graphs (with n vertices and at least $n(n-2)/2$ edges), and complete multipartite graphs have modularity 0 (see [36] and the references therein).

The idea of a resolution limit was introduced by Fortunato and Barthélemy [16] in 2007. Let G be an m -edge graph without isolated vertices, which has a connected component C with $e(C) < \sqrt{2m}$: then every optimal partition for G has one part $V(C)$, so all the vertices of C are clustered together.

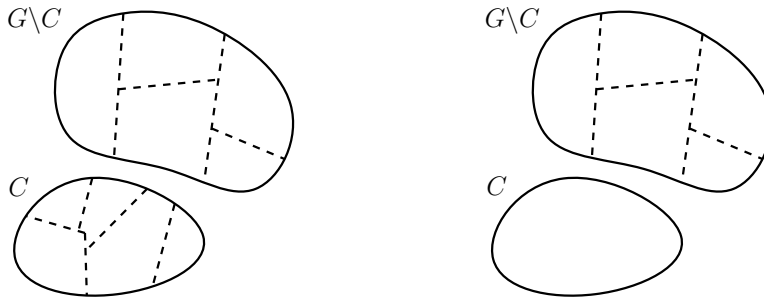


Fig. 1: A partition of G in which C is split into parts (left), and the corresponding partition where C forms a single part (right).

To see this, consider an optimal partition $\mathcal{A}$ for G , and note that for each part A in $\mathcal{A}$, the induced subgraph $G[A]$ is connected, and so $A \subseteq V(C)$ or $A \cap V(C) = \emptyset$. Suppose that there are $k \geq 2$ parts $A_1, A_2, \dots, A_k$ with $A_i \subseteq V(C)$. Let $\mathcal{B}$ be the partition formed from $\mathcal{A}$ by replacing these parts A_i by their union $V(C)$ – see Fig. 1. Then $q_{\mathcal{B}}^E(G) - q_{\mathcal{A}}^E(G) \geq (k-1)/m$, since any connected graph with k vertices has at least $k-1$ edges. Also, using again the simple result that if $x_1, x_2, \dots, x_k \geq 0$ then $\sum_i x_i^2 \geq (\sum_i x_i)^2/k$, we have

$$\begin{aligned} q_{\mathcal{B}}^D(G) - q_{\mathcal{A}}^D(G) &= (\text{vol}(C)^2 - \sum_i \text{vol}(A_i)^2) / \text{vol}(G)^2 \\ &\leq (\text{vol}(C)^2 - \text{vol}(C)^2/k) / \text{vol}(G)^2 = (k-1)/k \cdot e(C)^2/m^2. \end{aligned}$$

Hence

$$m(q_{\mathcal{B}}(G) - q_{\mathcal{A}}(G)) \geq (k-1)(1 - e(C)^2/km) > 0 \quad \text{since } e(C) < \sqrt{km}.$$

Thus, no optimal partition splits C , and we have proved the resolution limit result.

Consider an example where the component C consists of any two connected graphs C_1, C_2 with $e(C_1) = e(C_2)$, joined by a single edge. As we have seen, if $\sqrt{2m} > e(C)$ then C is never split in an optimal partition. But if $\sqrt{2m} < e(C)$, then C is always split: to see this, note that if $\mathcal{A}$ splits $V(C)$ into $V(C_1)$ and $V(C_2)$, then in the last displayed equation, the first inequality holds at equality. This example also highlights the sensitivity of optimal partitions to noise in the network: if that edge between C_1 and C_2 (perhaps a mistake in the data) had not been there, then C would be split in every optimal partition. In contrast, although the structure of optimal partitions is not robust to small changes in the set of edges, the modularity value of the graph is robust in this sense, as we see in the following result (see [35], where it is also shown that the constant 2 on the right side of the inequalities is best possible).

Theorem 2.1. *Let the graph $G = (V, E)$ have $m \geq 1$ edges, and let $\mathcal{A}$ be a partition of V . If E_0 is a non-empty subset of E , and $G' = (V, E \setminus E_0)$, then*

$$|q_{\mathcal{A}}(G) - q_{\mathcal{A}}(G')| < \frac{2|E_0|}{m} \quad \text{and} \quad |q^*(G) - q^*(G')| < \frac{2|E_0|}{m}.$$

Next let us consider the number of parts in the partitions. We noted that $q_{\mathcal{A}}(G) \leq 1 - 1/k$ for any k -part partition $\mathcal{A}$. The following result of Dinh and Thai [13] shows that if $q^*(G) > 0$, then there is a 2-part partition (a bipartition) $\mathcal{B}$ for G with $q_{\mathcal{B}}(G) > 0$.

Theorem 2.2. *Let $G = (V, E)$ be a graph, and let $k \geq 2$. Then there is a partition $\mathcal{B}$ with at most k parts such that $q_{\mathcal{B}}(G) \geq (1 - \frac{1}{k}) q^*(G)$.*

Theorem 2.2 is best possible for each $k \geq 2$, in that we cannot replace the factor $(1 - \frac{1}{k})$ by any larger value. For, $q_{\mathcal{B}}(G) \leq 1 - \frac{1}{k}$ since $\mathcal{B}$ has at most k parts; and, given $\varepsilon > 0$, there is a graph G such that $(1 - \frac{1}{k} + \varepsilon) q^*(G) > 1 - \frac{1}{k}$. (For example, we may consider G as consisting of many disjoint edges.)

The deterministic result Theorem 2.2 has a simple and beautiful probabilistic proof. Let $\mathcal{A} = (A_1, A_2, \dots, A_t)$ be a t -part partition of V with $q_{\mathcal{A}}(G) = q^*(G)$, where we may assume that $k < t$. Let the random variables $X_1, X_2, \dots, X_t$ be independent, with each uniformly distributed on $[k]$. We obtain a random partition $\mathcal{B} = (B_1, B_2, \dots, B_k)$ of V (with at most k non-empty parts) by setting $B_i = \bigcup_{\{j: X_j=i\}} A_j$.

Let G have m edges; and for $v, w \in V$ let $C_{v,w} = \mathbf{1}_{vw \in E} - d_v d_w / 2m$, and (in a mild abuse of notation) write $vw \in \mathcal{A}$ if v and w are in the same part of $\mathcal{A}$. (Note that we allow $v = w$ here.) Now

$$q_{\mathcal{A}}(G) = \frac{1}{2m} \sum_{v,w} \mathbf{1}_{vw \in \mathcal{A}} C_{v,w} \quad \text{and} \quad \mathbb{E}[q_{\mathcal{B}}(G)] = \frac{1}{2m} \sum_{v,w} \mathbb{E}[\mathbf{1}_{vw \in \mathcal{B}}] C_{v,w}.$$

Note that $\mathbb{E}[\mathbf{1}_{vw \in \mathcal{B}}] = \mathbf{1}_{vw \in \mathcal{A}} + \frac{1}{k} \mathbf{1}_{vw \notin \mathcal{A}}$. Also, $\sum_{v,w} C_{v,w} = 2m - (2m)^2 / 2m = 0$, and so

$$\sum_{v,w} \mathbf{1}_{vw \notin \mathcal{A}} C_{v,w} = - \sum_{v,w} \mathbf{1}_{vw \in \mathcal{A}} C_{v,w}.$$

Hence

$$\mathbb{E}[q_{\mathcal{B}}(G)] = \frac{1}{2m} \sum_{v,w} (\mathbf{1}_{vw \in \mathcal{A}} + \frac{1}{k} \mathbf{1}_{vw \notin \mathcal{A}}) C_{v,w} = (1 - \frac{1}{k}) q_{\mathcal{A}}(G) = (1 - \frac{1}{k}) q^*(G),$$

and this yields Theorem 2.2.

Again let $G = (V, E)$ be a graph with $m (\geq 1)$ edges. The *relative modularity* of a non-empty set $A \subseteq V$ is

$$q_r(A) = \frac{2e(A)}{\text{vol}(A)} - \frac{\text{vol}(A)}{2m}.$$

Then, for any partition $\mathcal{A}$ of V , $q_{\mathcal{A}}(G)$ is a weighted average of values $q_r(A)$,

$$q_{\mathcal{A}}(G) = \sum_{A \in \mathcal{A}} \frac{\text{vol}(A)}{2m} q_r(A). \quad (2.3)$$

Denote the maximum value of $q_r(A)$ over non-empty $A \subseteq V$ by $q_r^*(G)$. Then, by (2.3), $q_{\mathcal{A}}(G) \leq q_r^*(G)$, and so $q^*(G) \leq q_r^*(G)$. This inequality often yields our best bounds on $q^*(G)$ – see for example the proof of Theorem 4.1. In contrast, however, we can have $q^*(G) = o(1)$ and $q_r^*(G) = 1 - o(1)$. For example, consider an n -vertex graph G consisting of about $\log n$ disjoint edges and a complete graph on the remaining vertices. Then the number of edges ‘missing’ from G is $o(n^2)$ and $q^*(K_n) = 0$ (see [8]), and so $q^*(G) = o(1)$ by Theorem 2.1. However, $q_r(A) = 1 - o(1)$ when A is the set of vertices of the disjoint edges.

The following result concerning simple random graphs on $[n]$ (which we mentioned earlier) is due to Liebenau and Wormald [28, Theorem 1.6] (and more precise results are given in Lemma 7.1 of the same paper).

Theorem 2.3. *Let $m = m(n) \in \mathbb{N}$, and assume that $d = 2m/n$ satisfies $(\log n)^\omega < d < \mu_0 n$, where $\mu_0 > 0$ is sufficiently small and $\omega = \omega(n) \rightarrow \infty$ as $n \rightarrow \infty$. Let $\alpha < 3/5$, and let the n -vector $\mathbf{d} \in \mathbb{N}^n$ with $\sum_i d_i = 2m$ satisfy the ‘near-regular’ condition $\max_i |d_i - d| \leq d^\alpha$. Now suppose that we sample G uniformly from all graphs with degree-sequence $\mathbf{d}$. Then, for distinct vertices v and w ,*

$$\Pr(vw \in E(G)) = \frac{d_v d_w}{d(n-1)} \left(1 - \frac{(d_v - d)(d_w - d)}{d(n-1-d)} \right) + O\left(\frac{\sqrt{d \log n}}{n^2} + \frac{(\sqrt{d \log n})^3}{n^3} \right).$$

Noting that $d(n-1) = 2m(1 - d/2m) = 2m(1 - 1/n)$ this result implies

$$\Pr(vw \in E(G)) = \frac{d_v d_w}{2m} \left(1 - \frac{(d_v - d)(d_w - d)}{d(n-1-d)} \right) \left(1 - \frac{1}{n} \right)^{-1} + O\left(\left(\frac{\log n}{n} \right)^{3/2} \right).$$

Since by assumption $|(d_v - d)(d_w - d)|/d \leq d^{2\alpha-1} = o(n^{1/5})$, we obtain the simpler equation

$$\Pr(vw \in E(G)) = \frac{d_v d_w}{2m} \left(1 + o\left(\frac{1}{n^{4/5}}\right)\right) + O\left(\left(\frac{\log n}{n}\right)^{3/2}\right).$$

3. The modularity of binomial random graphs $G_{n,p}$

Let n be a positive integer. Given $0 \leq p \leq 1$, the *binomial random graph* $G_{n,p}$ has vertex-set $[n]$ and the $\binom{n}{2}$ possible edges appear independently with probability p . Given an integer m with $0 \leq m \leq \binom{n}{2}$, the *Erdős-Rényi random graph* $G_{n,m}$ is sampled uniformly from the graphs on vertex-set $[n]$ with m edges. These two random graphs are closely related when m is about $\binom{n}{2}p$. We shall focus here on $G_{n,p}$, and mention $G_{n,m}$ only briefly towards the end of this section.

Our first theorem, the ‘three phases theorem’, is from [34] and gives the big picture. The three phases correspond to when the expected vertex-degree (essentially np) is (a) at most about 1, (b) larger than 1 but bounded, or (c) tending to infinity.

Theorem 3.1. *Let $p = p(n)$ satisfy $0 < p \leq 1$.*

(a) *If $n^2 p \rightarrow \infty$ and $np \leq 1 + o(1)$, then whp $q^*(G_{n,p}) = 1 - o(1)$.*

(b) *Given constants $1 < c_0 \leq c_1$, there exists $\delta = \delta(c_0, c_1) > 0$ such that, if $c_0 \leq np \leq c_1$ for n sufficiently large, then whp $\delta < q^*(G_{n,p}) < 1 - \delta$.*

(c) *If $np \rightarrow \infty$ then whp $q^*(G_{n,p}) = o(1)$.*

For any graph G , let $\mathcal{C}$ be the partition of the vertex-set into the vertex-sets of its connected components; then $q_{\mathcal{C}}^E(G) = 1$. If np is a little less than 1, then whp $\mathcal{C}$ is the unique optimal partition for $G_{n,p}$ (see [47]). Also, if $np \leq 1 + o(1)$, then $q_{\mathcal{C}}(G_{n,p}) \xrightarrow{P} 1$, and so $q^*(G_{n,p}) \xrightarrow{P} 1$. To show this, it suffices to check that whp the fraction of edges in any component is $o(1)$, and so the degree tax is $o(1)$; this yields part (a) of Theorem 3.1. Parts (b) and (c) largely follow from more detailed results as in Theorem 3.2 below, see [34] for the upper bound in part (b). (We shall see later that part (c) also follows from Theorem 6.2(a) on percolated random graphs.)

In each of the parts (a), (b), (c) of Theorem 3.1 we can be more precise. For the rest of this section, we shall focus on the denser ranges (b) and (c). The following result is from [34].

Theorem 3.2. *There is a constant b such that, for all $0 < p = p(n) \leq 1$, we have $q^*(G_{n,p}) < b/\sqrt{np}$ whp. Also, given $0 < \varepsilon < 1$, there exists $a = a(\varepsilon) > 0$ such that, if $p = p(n)$ satisfies $np \geq 1$ and $p \leq 1 - \varepsilon$ for n sufficiently large, then $q^*(G_{n,p}) > a/\sqrt{np}$ whp.*

This theorem shows that, if $p = p(n)$ satisfies $1/n \leq p \leq 0.99$, then $q^*(G_{n,p}) = \Theta((np)^{-1/2})$ whp, confirming a conjecture in the physics literature [43] (see also the comment following Theorem 7.4 below).

The lower bound in Theorem 3.2 comes from analysing a simple algorithm ‘Swap’, which works as follows. Given a graph G with vertex-set $[n]$, we start with the bipartition (A, B) into the odd-numbered and even-numbered vertices: this partition has modularity score near 0 whp. By interchanging some pairs a_i, b_i between A and B , whp we can significantly increase the edge contribution, without changing the distribution of the degree tax. In a little more detail, assume

for simplicity that n is divisible by 6, let $k = n/6$, and relabel A as $a_1, a_2, \dots, a_{3k}$ and B as $b_1, b_2, \dots, b_{3k}$. We keep $A_1 = \{a_{2k+1}, a_{2k+2}, \dots, a_{3k}\}$ and $B_1 = \{b_{2k+1}, b_{2k+2}, \dots, b_{3k}\}$ fixed. For each $i = 1, 2, \dots, 2k$ we consider the edges between a_i, b_i and A_1, B_1 , and interchange (swap) a_i and b_i if that would increase the number of these edges which lie within the parts. See also Section 7 below where we discuss the stochastic block model.

Now consider the upper bound in Theorem 3.2, which will follow from spectral results. Define the *normalised Laplacian* of a graph H to be $\mathcal{L}_H = \mathbf{I} - \mathbf{D}^{-1/2} \mathbf{A} \mathbf{D}^{-1/2}$, where $\mathbf{A}$ is the adjacency matrix of H , d_i is the degree of vertex i , and $\mathbf{D}^{-1/2}$ is the diagonal matrix with i th diagonal entry $d_i^{-1/2}$. The *spectral gap* of an n -vertex graph H is $\bar{\lambda}_H = \max_{i \neq 0} |1 - \lambda_i|$, where $0 = \lambda_0 \leq \lambda_1 \leq \dots \leq \lambda_{n-1}$ are the eigenvalues of $\mathcal{L}_H$.

The modularity of a graph G is bounded above by its spectral gap: we have $q^*(G) \leq \bar{\lambda}_G$ (see [34]). Moreover, $q^*(G)$ is bounded above by a simple function of the relative size and the spectral gap of any subgraph H , as in the following result [29, Corollary 1.9].

Theorem 3.3. *Let the graph G have a subgraph H with spectral gap $\bar{\lambda}_H$, and let $\alpha = e(H)/e(G)$. Then*

$$q^*(G) \leq 1 - \alpha \min\{\alpha, 1 - \bar{\lambda}_H\}.$$

To complete the proof of the upper bound on $q^*(G_{n,p})$, we apply the bound in Theorem 3.3 to a graph H obtained from $G_{n,p}$ by deleting a few vertices, such that whp H has small spectral gap.

In 2025, Rybarczyk and Sulkowska [46] improved the upper bound in Theorem 3.2, by replacing the constant b by the explicit value 3.06, showing that

$$\text{for all } 0 < p = p(n) = o(1), \quad q^*(G_{n,p}) < 3.06/\sqrt{np} \text{ whp.} \quad (3.1)$$

In their proof they first used Chernoff's inequality to show bounds on $e(S)$, $e(\bar{S})$ and $e(S, \bar{S})$ which whp hold simultaneously for all vertex subsets S .

If we weaken the lower bound in Theorem 3.2 by introducing an extra factor $\sqrt{1-p}$, then we can extend the bound for p up to when the complementary graph has expected average degree just above 1 (see [34]). Observe that if p is bounded below 1 (as in the lower bound in Theorem 3.2), then this factor is just a constant. The next two results are from [36].

Theorem 3.4. *For each $\varepsilon > 0$, there exists $\alpha > 0$ such that, if $p = p(n)$ satisfies $1/n \leq p \leq 1 - (1 + \varepsilon)/n$, then $q^*(G_{n,p}) \geq \alpha \sqrt{(1-p)/np}$ whp.*

We saw in Theorem 3.2 that $q^*(G_{n,p}) < b/\sqrt{np}$ whp, and so if $np \rightarrow \infty$ then $q^*(G_{n,p}) \rightarrow 0$ whp (as in Theorem 3.1(c)). When do we have the stronger result that whp $q^*(G_{n,p}) = 0$? Consider the very dense case, where $p = 1 - \Theta(1/n)$. In this case, $\sqrt{(1-p)/np} = \Theta(1/n)$, and so the lower bound in part (c) below follows directly from Theorem 3.4.

Theorem 3.5. *Let $0 \leq p = p(n) \leq 1$.*

- (a) *If p satisfies $p \geq 1 - 1/n + \omega(n) n^{-3/2}$, where $\omega(n) \rightarrow \infty$ as $n \rightarrow \infty$, then $q^*(G_{n,p}) = 0$ whp.*
- (b) *If $p = 1 - 1/n$, then $\mathbb{P}(q^*(G_{n,p}) = 0) \rightarrow 1/2$ as $n \rightarrow \infty$.*
- (c) *Given $1 < c_1 < c_2$, there exist α, β with $0 < \alpha < \beta$ such that, if p satisfies $1 - c_2/n \leq p \leq 1 - c_1/n$, then $\alpha/n \leq q^*(G_{n,p}) \leq \beta/n$ whp.*

It is known that (for $n \geq 4$) the least number of edges missing from an n -vertex graph G with $q^*(G) > 0$ is $\lfloor n/2 \rfloor + 1$ (see [36]). But when $1 - p$ is very small (below $1/n$), then whp at most $n/2$ edges are missing from $G_{n,p}$ and so $q^*(G_{n,p}) = 0$; this yields part (a) of Theorem 3.5. Part (c) says that when $1 - p$ is just a little larger (above $1/n$), then whp $q^*(G_{n,p}) = \Theta(1/n)$.

The Erdős–Rényi random graph $G_{n,m}$

Recall that, given $m = m(n)$, the random graph $G_{n,m}$ is sampled uniformly from the graphs on $[n]$ with m edges. It is similar to $G_{n,p}$ with $p = m/\binom{n}{2}$, and so it is natural that there are results on $q^*(G_{n,m})$ which correspond to Theorems 3.1 to 3.5 above (see [34], Section 1.3). For example, if $m \rightarrow \infty$ and $m \leq (\frac{1}{2} + o(1))n$, then $q^*(G_{n,m}) \rightarrow 1$ in probability, and if $m/n \rightarrow \infty$, then $q^*(G_{n,m}) \rightarrow 0$ in probability.

Friendly bisections

Given a fixed p with $0 < p < 1$, whp $q^*(G_{n,p}) = \Theta(n^{-1/2})$ by Theorem 3.2, and so whp the maximum score over bipartitions is $\max_{|\mathcal{A}|=2} q_{\mathcal{A}}(G_{n,p}) = \Theta(n^{-1/2})$ by Theorem 2.2. Thus, whp there is a bipartition with small but positive modularity score.

A different take on bipartitions may reflect the community structure of the graph. A partition is *balanced* if the sizes of the parts differ by at most 1, and a balanced bipartition is also called a *bisection*. A *friendly bisection* of an n -vertex graph is a bisection in which all but $o(n)$ vertices have more neighbours in their own part than in the other part. It is shown in [15] that whp $G_{n,1/2}$ has a friendly bisection.

A bisection such that all but one vertex in each part have more neighbours in their own part may have modularity score 0. For example, let the graph G consist of two disjoint wheels on $k+1$ vertices, so that G has $n = 2k+2$ vertices and $m = 4k$ edges. Let A contain the rim of one wheel and the centre of the other, let B be the complementary set of vertices, and consider the bisection $\mathcal{A} = \{A, B\}$. Each of the $2k$ vertices in the rims has two neighbours in the same part and one in the other part, and the two centres have all k neighbours in the other part. Thus, $\mathcal{A}$ is a friendly bisection of G ; but $q_{\mathcal{A}}^E(G) = 2k/m = \frac{1}{2}$, and $q_{\mathcal{A}}^D(G) = \frac{1}{2}$ by symmetry, and so $q_{\mathcal{A}}(G) = 0$.

However, suppose that, given a bipartition $\mathcal{A}$ for G (not necessarily balanced), *all* vertices have at least as many neighbours in their own part as in the other part, and at least one vertex has strictly more. In this case, $q_{\mathcal{A}}(G) > 0$. To see this, let there be x edges between A and B , y edges within A , and z edges within B . Then $x \leq 2y$ and $x \leq 2z$, with at least one of these inequalities strict, and so $x^2 < 4yz$. But $q_{\mathcal{A}}^E(G) = (y+z)/(x+y+z)$ and $q_{\mathcal{A}}^D(G) = ((x+2y)^2 + (x+2z)^2)/(2x+2y+2z)^2$. Hence

$$4(x+y+z)^2 q_{\mathcal{A}}(G) = 4(x+y+z)(y+z) - ((x+2y)^2 + (x+2z)^2) = 2(4yz - x^2) > 0,$$

and so indeed $q_{\mathcal{A}}(G) > 0$.

4. The modularity of random regular graphs

For a regular graph G , modularity scores have a simplified form (see (2.1) and (2.3)):

$$q_{\mathcal{A}}(G) = \sum_{A \in \mathcal{A}} \left(\frac{e(A)}{m} - \frac{|A|^2}{n^2} \right) = \sum_{A \in \mathcal{A}} \frac{|A|}{n} q_r(A), \quad \text{where} \quad q_r(A) = \left(\frac{2e(A)}{r|A|} - \frac{|A|}{n} \right). \quad (4.1)$$

Random 2-regular graphs $G_{n,2}$ are easy to handle. They have only a few cycles whp, and so, by the robustness result in Theorem 2.1, whp their modularity is close to that of the n -cycle C_n . It follows (see [33]) that

$$q^*(G_{n,2}) = 1 - \frac{2}{\sqrt{n}} + o\left(\frac{\log^2 n}{n}\right) \quad \text{whp.} \quad (4.2)$$

In [33] it is also shown that, for random 3-regular graphs $G_{n,3}$ (with n even),

$$2/3 - o(1) \leq q^*(G_{n,3}) < 0.804 \quad \text{whp.}$$

Here, the lower bound follows easily from the result that whp $G_{n,3}$ has a Hamiltonian cycle (see [44]). We ignore the rest of the graph (a perfect matching) and cut the circuit into $\omega(n)$ paths of about equal length to give a vertex-partition, with edge contribution $\frac{2}{3} - o(1)$ and degree tax $o(1)$. Indeed, for each $r \geq 2$, $q^*(G) \geq 2/r - 2\sqrt{6/n}$ for *every* n -vertex r -regular graph G (see [33, Proposition 1.4].) It was conjectured in [33] that whp $q^*(G_{n,3}) > 2/3 + \delta$ for some $\delta > 0$. This was shown in [27] together with an improved upper bound.

Theorem 4.1. $0.667026 \leq q^*(G_{n,3}) < 0.789998$ whp.

To prove that $q^* > \frac{2}{3} + \delta$ it suffices to show that there exists a set C of $\Theta(n)$ vertices such that $q_r(C) \geq \frac{2}{3} + \varepsilon$ and the remaining graph may still be partitioned so that the weighted sum contributes $(\frac{2}{3} - o(1))(1 - |C|/n)$ to the modularity score (see (4.1)). We noted earlier that to prove an upper bound, it suffices to show that the relative modularity of all vertex-sets has at most some value.

To construct a vertex-set with high relative modularity and linear size, and thus to prove the lower bound in Theorem 4.1, we consider the configuration model. Choose a vertex v uniformly at random, and start exploring its connected component one edge at a time in a certain order, until a set C_0 of εn vertices has been explored. Next, explore one by one the open half-edges sticking out of C_0 , to grow it little by little. Then start adding to C_0 short paths containing exactly two explored vertices – the first and the last vertex of each path. In this way, the relative modularity of C_0 increases at each step. By analysing the first few steps of this procedure, using the differential equation method [12] carefully, and optimizing over ε , we can complete the proof.

To prove the upper bound in Theorem 4.1, we note that, for the modularity to be at least the given upper bound, there must be a set A of vertices having relative modularity at least that value. First we show that there is an $\varepsilon_0 > 0$ such that whp this does not hold for any ‘small’ set A with $|A| < \varepsilon_0 n$. Then we show that it suffices to prove that there is no ‘large’ set B , containing many edges, and such that the induced subgraphs on B and on its complement contain only vertices of degree 2 or 3. Finally, we use the first moment method to show that whp there is no such set B .

The next theorem (from [33]) shows that $q^*(G_{n,r})$ is $\Theta(1/\sqrt{r})$ whp for large (fixed) values r . For simplicity, we restrict our attention to even values of n .

Theorem 4.2. *For each $r \geq 2$, $q^*(G_{n,r}) < 2/\sqrt{r}$ whp, and there is a constant r_0 such that, for each $r \geq r_0$, $q^*(G_{n,r}) > 0.7632/\sqrt{r}$ whp.*

Here, the lower bound follows from a result on the bisection width of random regular graphs. Dembo, Montanari and Sen [11] showed that whp there is a vertex-set U with $|U| = n/2$ such that $e(U, \bar{U}) \leq rn/4 - c\sqrt{rn}/2 + o(\sqrt{rn})$, where $c = 0.76321 \pm 0.00003$. Thus, in this case, the bipartition $\mathcal{A} = \{U, \bar{U}\}$ has edge contribution at least $q_{\mathcal{A}}^E = \frac{1}{2} + (c + o(1))/\sqrt{r}$ and degree tax $q_{\mathcal{A}}^D = \frac{1}{2}$, and this establishes the whp lower bound. The upper bound follows from a spectral bound on random regular graphs by Friedman [19], and from the inequality in Theorem 3.3 (with $H = G$, $\alpha = 1$) which provides an upper bound on the modularity of a graph in terms of its spectral gap.

5. Modularity in small-world models

In this section we discuss modularity for some ‘small-world’ models, where typical distances are small.

The preferential attachment model

The preferential attachment random graph model G_n^h is of particular interest because it exhibits a ‘rich-get-richer’ behaviour: vertices which already have many incident edges are more likely to be joined to the incoming vertices; also, vertex-degrees follow a power-law distribution. To construct G_n^h on the vertex-set $[n]$, we first construct a graph (with loops) on the ‘mini-vertices’ $1, 2, \dots, hn$. Let T_1 be the graph consisting of the single mini-vertex 1 with a loop e_1 (so that v_1 has degree 2). For $1 \leq t \leq hn - 1$, the graph T_{t+1} is constructed from T_t by adding the mini-vertex $t + 1$ and joining it by an edge e_{t+1} to a mini-vertex $s \in \{1, 2, \dots, t\}$ with probability $d_s^t / (2t + 1)$, and adding a loop at the mini-vertex $t + 1$ with probability $1 / (2t + 1)$, where d_s^t is the degree of the mini-vertex s in T_t . Finally, for $h \geq 2$, the random multigraph G_n^h is obtained from T_{hn} by merging each set of h consecutive mini-vertices together into a single vertex, keeping loops and multiple edges. Observe that G_n^h has hn edges, and so has average degree $2h$. We consider whp bounds on $q^*(G_n^h)$ for large values of h .

Theorem 5.1. *If $h \geq 2$, then whp $\Omega(1/\sqrt{h}) = q^*(G_n^h) = O(\sqrt{\log h} / \sqrt{h})$.*

Here, the lower bound was proved by Prokhorenkova, Prałat and Raigorodskii [42], by analysing an algorithm which, as the random graph was constructed, iteratively placed the new vertex into one of two parts, always choosing the part to which that vertex had more incident vertices. The same lower bound was also shown to hold for a generalised model of preferential attachment in [1] – using only properties of the degree-sequence of that random graph. Note that the whp lower bounds for both the preferential attachment model and the binomial (or Erdős–Rényi) model are $\Theta(1/\sqrt{d})$, where d is the expected average degree.

The upper bound in Theorem 5.1 was proved by Rybarczyk and Sulkowska [45], and we describe their proof. The main step is to show that whp, for all vertex-sets $S \subseteq V(G)$, both $e(S)$ and $\text{vol}(S)$ are bounded by functions of a deterministic value $\mu(\tilde{S})$, where $\tilde{S}$ is the set of mini-vertices in T_{hn} corresponding to S . Note that the index of a mini-vertex gives the time at which it was added in the construction process. If we define

$$\mu(\tilde{S}) = \frac{1}{2} \sqrt{\pi} \sum_{j \in \tilde{S}} (2j - 3)!! / (2j - 2)!! ,$$

we may establish that whp, for all non-empty $S \subseteq V(G)$,

$$e(S) \leq \mu(\tilde{S})^2 + O(n\sqrt{h \log h}) \quad \text{and} \quad \text{vol}(S) \geq 2\sqrt{hn} \mu(\tilde{S}) + O(n\sqrt{h \log h}).$$

We also use the nice observation that

$$q^*(G) \leq 4 \max_{S \subseteq V(G)} (e(S)/e(G) - \text{vol}(S)^2/\text{vol}(G)^2) ,$$

which holds since by Theorem 2.2 there is a bipartition $\mathcal{A}$ with $q^*(G) \leq 2q_{\mathcal{A}}(G)$. From these inequalities, and bounds on the deterministic function $\mu(\tilde{S})$, the upper bound on $q^*(G_n^h)$ quickly follows.

The spatial preferential attachment model

The spatial preferential attachment model (SPA) yields a sequence of t -vertex directed graphs $(G_t)_{0 \leq t \leq n}$, where each vertex has an embedding in $[0, 1]^m$. (The edges will be undirected later.) The vertices are added one by one. For a vertex v already present in the graph, an edge vw may be added from v to the newly added vertex w if w falls within a certain distance of v , such that existing vertices of higher degree may have an edge to w at greater distances; here, distances are taken as on a torus. We follow the notation of [42]. Let $a_1, a_2 > 0$ and let $0 < p \leq 1$. The ‘sphere of influence’ $S(v, t)$ of vertex v at time t is defined to be the ball centred at v with volume

$$|S(v, t)| = \min\{(a_1 d_{G_t}^-(v) + a_2)/t, 1\},$$

where $d_{G_t}^-(v)$ is the in-degree of v in G_t . Let G_0 be the empty graph with no vertices. We construct the graph G_t with vertex-set V_t from G_{t-1} by adding a new vertex w_t randomly embedded uniformly in $[0, 1]^m$; then, independently for each vertex $v \in V_{t-1}$, if $w_t \in S(v, t-1)$ we add the directed edge (w_t, v) with probability p . Finally, let $\tilde{G}_n$ be the undirected graph constructed from G_n by replacing each directed edge by an edge. Prokhorenkova, Prałat and Raigorodskii [42] proved that $\tilde{G}_n$ has asymptotically maximal modularity.

Theorem 5.2. *Let $m \in \mathbb{N}$, let $a_1, a_2 > 0$, let $p \in (0, 1]$, and suppose that $pa_1 < 1$. Then for $\tilde{G}_n$ generated as above, whp $q^*(\tilde{G}_n) = 1 - o(1)$.*

The proof of Theorem 5.2 proceeds by giving an explicit construction for partitioning $[0, 1]^m$ into a slowly growing number of slices. Interestingly, this partition depends on parameters of the model, but is independent of the random graph which has been sampled. The partition of $[0, 1]^m$ induces a vertex-partition of $\tilde{G}_n$, which is then shown to yield modularity score near 1. In the proofs that the edge contribution is near 1 and that the degree tax is near 0, the vertices are considered separately based on their arrival time to the graph.

The modularity of random graphs in the hyperbolic plane

We next consider the KPKBV model introduced by Krioukov *et al.* [23], and denote the n -vertex graph by $G_{n,\alpha,\nu}$ for parameters $\alpha, \nu > 0$. This graph is of interest because it has many properties associated with observed networks: power-law degree distribution with exponent $2\alpha + 1$, non-vanishing clustering coefficient and short typical distances between vertices – see, for example, [18] and the references therein. Note that an exponent between 2 and 3 for the degree distribution corresponds to $\alpha \in (\frac{1}{2}, 1)$, and this is the range most usually considered.

For $\alpha, \nu > 0$, we may construct the n -vertex hyperbolic random graph $G'_{n,\alpha,\nu}$ as follows. We generate the position (r_i, Θ_i) (in polar co-ordinates) of vertex v_i , independently for each i . Set $R = R(n, \nu)$ to satisfy $n = \nu e^{R/2}$, and sample r_i with probability density function $\rho_n(r) = \alpha \sinh(\alpha r) / (\cosh(\alpha R) - 1)$ for $0 < r < R$. Also, sample Θ_i uniformly in $(0, 2\pi]$, independently of r_i . Include the edge $v_i v_j$ if the vertices are within (hyperbolic) distance R of each other – that is, if

$$\cosh(r_i) \cosh(r_j) - \sinh(r_i) \sinh(r_j) \cos(|\Theta_i - \Theta_j|_{2\pi}) \leq \cosh(R),$$

where $|a|_b = \min(|a|, b - |a|)$. Define also the ‘Poissonisation’ $G_{n,\alpha,\nu}$ of the above model – which has $\text{Po}(n)$ points, rather than n . The vertex-set is the point-set of a Poisson point process on the disc of radius R with intensity $n(2\pi)^{-1} \rho_n(r) dr d\Theta$. The edges are included as for the original model.

The Poissonisation of the hyperbolic graph has asymptotically maximal modularity value (see [9]).

Theorem 5.3. *For any $\alpha > \frac{1}{2}$ and $\nu > 0$, whp $q^*(G_{n,\alpha,\nu}) = 1 - o(1)$.*

Loosely, this result follows by cutting up the space in a natural way, regardless of the graph, and then showing that whp the induced partition of the graph achieves a modularity score near to 1. Let k be a large constant, let $A_c = \{v_i : \Theta_i \in ((c-1)/k, c/k]\}$ and let $\mathcal{A} = \{A_1, A_2, \dots, A_k\}$ (so that we are radially partitioning the space into k equal sectors).

6. Modularity and partially observed graphs

In this section, we first consider the percolated random subgraph G_p of a graph G , where the edges are retained independently with probability p . We then briefly consider the more general case where the edge-probabilities may differ. Finally, we discuss the ‘vertex-sampling’ model, where we sample a random k -set of vertices and observe the corresponding induced subgraph.

The percolated random graph model

Given a graph G and $0 < p \leq 1$, let the *percolated* random graph G_p be the random subgraph of G on the same vertex-set, obtained by considering each edge of G independently, keeping it in the graph with probability p , and otherwise deleting it. We assume that we can see the graph G_p but not the underlying graph G , so that G is ‘partially observed’. What can we infer about G ? The binomial random graph $G_{n,p}$ (discussed in Section 3) may be written as $(K_n)_p$, where K_n is the n -vertex complete graph. Let $e(H)$ be the number of edges in a graph H , so that the expected number of edges in G_p is $e(G)p$.

In this section we present two theorems from [35] that relate $q^*(G_p)$ and $q^*(G)$, as well the modularity score on G of partitions constructed looking only at G_p ; see Fig. 2 and Fig. 3. The first theorem concerns when we want to have $q^*(G_p) > q^*(G) - \varepsilon$ with probability near 1, and it shows that this happens if the expected number of edges seen is large enough.

Theorem 6.1. *Given $\varepsilon > 0$, there exists $c = c(\varepsilon)$ such that the following holds. For each graph G and probability p such that $e(G)p \geq c$, the random graph G_p satisfies $q^*(G_p) > q^*(G) - \varepsilon$ with probability $\geq 1 - \varepsilon$.*

The proof of Theorem 6.1 in [35] shows that we may take $c(\varepsilon) = \Theta(\varepsilon^{-3} \log \varepsilon^{-1})$, and an example graph involving a star and some disjoint edges shows that $c(\varepsilon)$ must be $\Omega(\varepsilon^{-1} \log \varepsilon^{-1})$.

The second theorem concerns when we want also to have $q^*(G_p) < q^*(G) + \varepsilon$ with probability near 1, and it shows that this will happen if the expected average degree is sufficiently large. We see also that, in this case, finding a good partition $\mathcal{A}$ for G_p helps us to find a good partition $\mathcal{A}'$ for G . Observe that the expected average degree in G_p is $2e(G)p/v(G)$, where $v(G)$ is the number of vertices in G .

Theorem 6.2. *For each $\varepsilon > 0$, there exists $c = c(\varepsilon)$ such that the following holds. Let the graph G and probability p satisfy $e(G)p/v(G) \geq c$. Then, with probability $\geq 1 - \varepsilon$,*

- (a) *the random graph G_p satisfies $|q^*(G_p) - q^*(G)| < \varepsilon$; and*
- (b) *given any partition $\mathcal{A}$ of the vertex-set, and seeing only G_p , we can find in a linear number of operations a partition $\mathcal{A}'$ with $q_{\mathcal{A}'}(G) \geq q_{\mathcal{A}}(G_p) - \varepsilon$.*

Here, part (b) says roughly that, given a good partition $\mathcal{A}$ for G_p , we can quickly construct a good partition $\mathcal{A}'$ for G . Using also part (a), we may see that, if the partition $\mathcal{A}$ for G_p satisfies

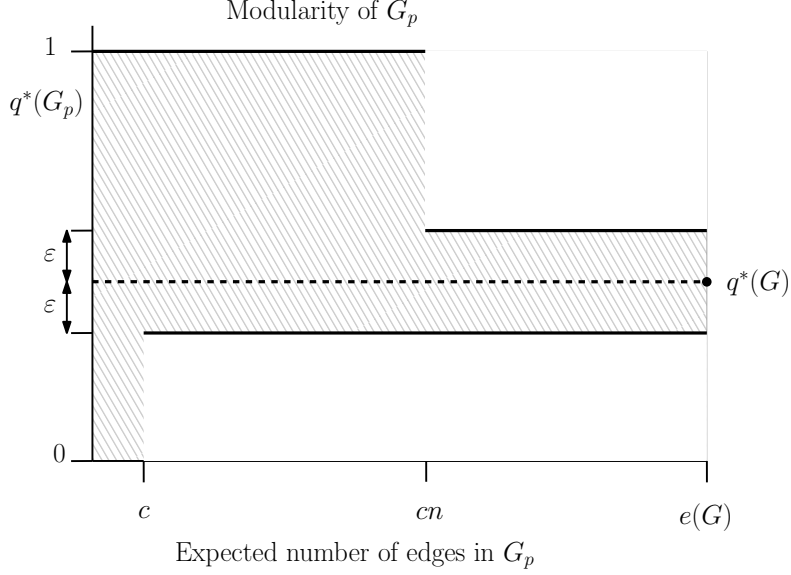


Fig. 2: For any $\varepsilon > 0$, there exists c such that the modularity of the percolated graph G_p lies in the region shown with probability at least $1 - \varepsilon$. See Theorems 6.1 and 6.2(a).

$q_{\mathcal{A}}(G_p) \geq q^*(G_p) - \varepsilon$ with probability at least $1 - \varepsilon$, then the partition $\mathcal{A}'$ satisfies $q_{\mathcal{A}'}(G) > q^*(G) - 2\varepsilon$ with probability at least $1 - 2\varepsilon$. (See Section 3 of [35] for the greedy amalgamating algorithm used to construct $\mathcal{A}'$ from $\mathcal{A}$, where every part of $\mathcal{A}'$ is a union of parts of $\mathcal{A}$.)

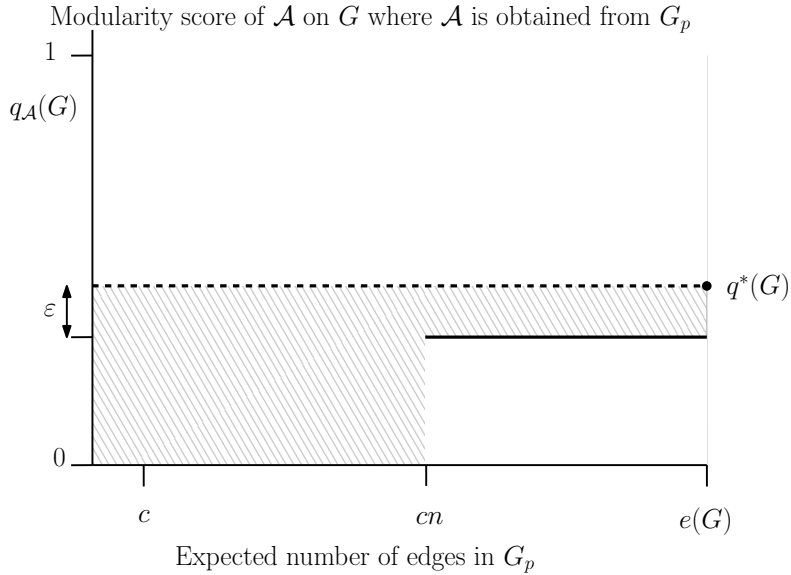


Fig. 3: For any $\varepsilon > 0$, there exists c such that the partition $\mathcal{A}$ constructed looking only at G_p has score $q_{\mathcal{A}}(G)$ in the region shown, with probability at least $1 - \varepsilon$. See Theorem 6.2(b).

The proof of Theorem 6.2 in [35] shows that we may take $c(\varepsilon) = \Theta(\varepsilon^{-3} \log \varepsilon^{-1})$. An example with $G = K_n$ (so that G_p is $G_{n,p}$ and $q^*(G) = 0$) shows that $c(\varepsilon)$ must be at least $\Omega(\varepsilon^{-2})$. Taking $G = K_n$ again and applying part (a) of Theorem 6.2 shows that $q^*(G_{n,p}) \rightarrow 0$ if $np \rightarrow \infty$, and so

we have another proof of Theorem 3.1 (c).

The assumption in Theorem 6.2 that the expected average degree is large is much stronger than the assumption in Theorem 6.1, but still $e(G_p)$ may whp be much smaller than $e(G)$. If we go further, and assume that at most an ε -proportion of edges are missed (that is, $|E \setminus E'| \leq \varepsilon|E|$), then deterministically we have $|q^*(G') - q^*(G)| \leq 2\varepsilon$, by Theorem 2.1.

Suppose that we take as the underlying graph G the dolphin social network (see [31]), with 62 vertices and 159 edges. It is known that $q^*(G) = 0.529$ to three decimal places (see [8]). In the upper part of Fig. 4, each point corresponds to the estimated modularity $\tilde{q}(G_p)$ of an instance of the sampled graph G_p . For each edge-probability $p = 0.1, 0.2, \dots, 0.9$, the graph G_p was sampled 50 times. For each sampled graph G_p , the plot shows the maximum modularity score of the partitions output by 200 runs of both the Louvain [5] and Leiden [48] algorithms. The lower part of Fig. 4 examines, for each random instance of G_p , how well the modularity maximising partition found for G_p yields a good partition for the underlying graph G . In detail, for each sampled graph G_p , the plot shows the score $q_{\tilde{\mathcal{A}}'}(G)$, where $\tilde{\mathcal{A}}$ is the highest scoring partition on G_p found in the runs of the Louvain and Leiden algorithms, and $\tilde{\mathcal{A}}'$ is the partition modified from $\tilde{\mathcal{A}}$ as in Theorem 6.2(b). For more details and simulations on a larger network, see [35].

Let us discuss briefly how to prove Theorems 6.1 and 6.2. A key preliminary result for both proofs is the ‘fattening lemma’ below. Given a graph G and a number η with $0 < \eta \leq 1$, we say that a vertex-partition $\mathcal{A}$ is η -fat if each part has volume at least $\eta \text{vol}(G)$. There is a natural greedy algorithm (see [35]) which, given a graph G , a number η with $0 < \eta \leq 1$, and a vertex-partition $\mathcal{B}$, amalgamates some parts in $\mathcal{B}$ to yield an η -fat partition $\mathcal{A} = \mathcal{A}(G, \eta, \mathcal{B})$ with modularity score at least about that of $\mathcal{B}$. Note that $|\mathcal{A}| \leq |\mathcal{B}|$.

Theorem 6.3 (The fattening lemma). *For each graph G and each number η with $0 < \eta \leq 1$, there is an η -fat partition $\mathcal{A}$ of $V(G)$ such that $q_{\mathcal{A}}(G) > q^*(G) - 2\eta$. Indeed, given any partition $\mathcal{B}$ of $V(G)$, the greedy amalgamating algorithm uses a linear number of operations and constructs an η -fat partition $\mathcal{A} = \mathcal{A}(G, \eta, \mathcal{B})$ such that $q_{\mathcal{A}}(G) > q_{\mathcal{B}}(G) - 2\eta$.*

The main step in the above algorithm is a number-bipartitioning problem: we have a set of parts B_i from $\mathcal{B}$ which we want to split into two sets, each with its sum of volumes at least $\eta \text{vol}(G)$. The constant 2 in Theorem 6.3 is best possible (see [35]). For comparison, recall Theorem 2.2 concerning the maximum modularity score $q_{\leq k}^*(G)$ for a partition of G with at most k parts. Observe that if $\mathcal{A}$ is an η -fat partition for G and $\bar{k} = \lfloor 1/\eta \rfloor$, then $|\mathcal{A}| \leq \bar{k}$, and so $q_{\mathcal{A}}(G) \leq q_{\leq \bar{k}}^*(G)$.

To prove Theorem 6.1, we note first that we may assume that $q^*(G) \geq \varepsilon$, and let $\eta = \varepsilon/4$. By Theorem 6.3, there is an η -fat partition $\mathcal{A}$ for G with $q_{\mathcal{A}}(G) \geq q^*(G) - 2\eta$. Since $|\mathcal{A}| \leq \eta^{-1}$, we have $q_{\mathcal{A}}^D(G) \geq \eta$, and also $q_{\mathcal{A}}^E(G) = q_{\mathcal{A}}(G) + q_{\mathcal{A}}^D(G) \geq 3\eta$. Since $\mathcal{A}$ is η -fat, and both $q_{\mathcal{A}}^E(G)$ and $q_{\mathcal{A}}^D(G)$ are bounded away from 0, we can use Chernoff probability bounds to show that $q_{\mathcal{A}}^E(G_p)$ is unlikely to be much smaller than $q_{\mathcal{A}}^E(G)$, and similarly that $q_{\mathcal{A}}^D(G_p)$ is unlikely to be much larger than $q_{\mathcal{A}}^D(G)$. The theorem follows.

The proof of Theorem 6.2 is more involved. The rough idea is that we can use the fattening lemma to bound the probability that a vertex-partition behaves badly by the probability that a fat vertex-partition behaves similarly badly. We can then use probabilistic methods (including Chernoff bounds, as above) to handle fat partitions.

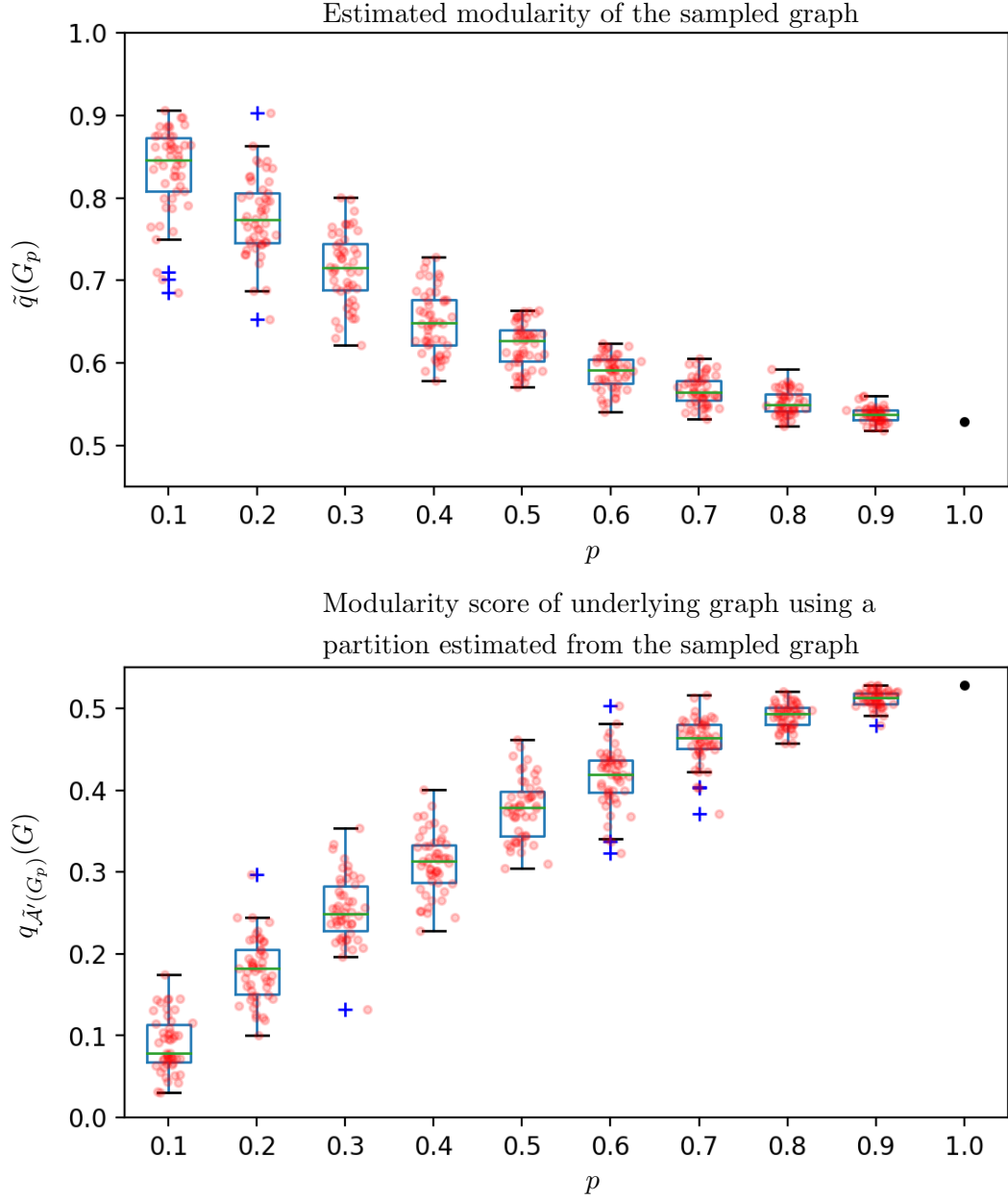


Fig. 4: Simulation results, based on the dolphin social network G , with $q^*(G) \approx 0.529$. In the upper part, each point corresponds to the estimated modularity $\tilde{q}(G_p)$ of an instance of the sampled graph G_p . (The noise in the x -axis is to allow one to see the points.) In the lower part, for each sampled graph G_p , the corresponding point shows the score $q_{\tilde{\mathcal{A}}'}(G)$, where $\tilde{\mathcal{A}}$ is the highest scoring partition found for G_p , and $\tilde{\mathcal{A}}'$ is the partition for G modified from $\tilde{\mathcal{A}}$ as in Theorem 6.2(b).

Percolation allowing distinct edge probabilities

Up to now we have considered the model G_p , where each edge from an underlying graph G appears independently with probability p , but we may also consider a model in which each edge uv in G appears in the observed graph with probability p_{uv} . Again, under certain conditions, the modularity of this graph converges whp to a deterministic quantity, and this is the modularity of a weighted graph.

Let V be a non-empty vertex-set, and let $w : V \times V \rightarrow \mathbb{R}$ satisfy $w_{uv} = w_{vu} \geq 0$ for all vertices u and v . We call w a *weight function* on V^2 . Define the (weighted) degree of a vertex u by setting $d_u^w = \sum_v w_{uv}$. Similarly, define the (weighted) volume of a vertex-set X by $\text{vol}_w(X) = \sum_{u \in X} d_u^w$, and (corresponding to $e(X)$) let $e_w(X) = \frac{1}{2} \sum_{u,v \in X} w_{uv}$, with $m_w = e_w(V)$ (corresponding to m).

Assume that w is not identically zero, so that $m_w > 0$. For a given partition $\mathcal{A}$ of V , the *modularity score* of $\mathcal{A}$ on w is

$$q_{\mathcal{A}}(w) = \frac{1}{2m_w} \sum_{A \in \mathcal{A}} \sum_{u,v \in A} \left(w_{uv} - \frac{d_u^w d_v^w}{2m_w} \right), \quad (6.1)$$

and the *modularity* of w is $q^*(w) = \max_{\mathcal{A}} q_{\mathcal{A}}(w)$. As in the unweighted case, $0 \leq q^*(w) < 1$ and we may ignore vertices with degree 0. Also, if w is identically 0 we set $q_{\mathcal{A}}(w) = 0$ and $q^*(w) = 0$. If w is $\{0,1\}$ -valued and $w_{vv} = 0$ for each vertex v , we recover the usual modularity score and modularity value for the corresponding graph. The values $q_{\mathcal{A}}(w)$ and $q^*(w)$ are unchanged under a re-scaling of w , and so we may assume that $0 \leq w_{uv} \leq 1$ for each $u, v \in V$, and in this case we call w a *probability weight function*.

Given a probability weight function w , let G_w be the random (unweighted) graph obtained by considering each edge independently, and including each edge uv with probability w_{uv} .

Theorem 6.4. *Given $\varepsilon > 0$, there exists $c = c(\varepsilon)$ such that the following holds. If the probability weight function w on V^2 satisfies $e_w(V) \geq c$, then with probability at least $1 - \varepsilon$, the random graph G_w satisfies $q^*(G_w) > q^*(w) - \varepsilon$.*

Theorem 6.5. *Given $\varepsilon > 0$, there exists $c = c(\varepsilon)$ such that the following holds. If the probability weight function w on V^2 satisfies $e_w(V) \geq c|V|$, then with probability at least $1 - \varepsilon$,*

- (a) *the random graph G_w satisfies $|q^*(G_w) - q^*(w)| < \varepsilon$; and*
- (b) *given any partition $\mathcal{A}$ of the vertex-set, in a linear number of operations (seeing only G_w) the greedy amalgamating algorithm finds a partition $\mathcal{A}'$ with $q_{\mathcal{A}'}(w) \geq q_{\mathcal{A}}(G_w) - \varepsilon$.*

In the next section, Theorem 6.5 is used to establish the modularity score of the stochastic block model.

Estimating modularity by vertex sampling ('parameter estimation')

Another common model of partially observing an underlying graph G is to sample a vertex-subset U of constant size k , and observe the corresponding induced subgraph $G[U]$. A graph parameter f is *estimable* (or *testable*) for a class $\mathcal{G}$ of graphs (see [7] and [30]) if, for every $\varepsilon > 0$, there exists $k = k(\varepsilon)$ such that, if G is a graph in $\mathcal{G}$ with at least k vertices, then $\mathbb{P}(|f(G) - f(G[X])| > \varepsilon) < \varepsilon$ when X is a uniformly random k -subset of $V(G)$. The following result ([35, Theorem 1.5]) shows that the modularity value $q^*(G)$ is estimable for dense graphs G , but not more generally.

Theorem 6.6. (a) For any fixed ρ with $0 < \rho < 1$, modularity is estimable for graphs with density at least ρ .

(b) For any given function $\rho(n) = o(1)$, modularity is not estimable for n -vertex graphs G with density at least $\rho(n)$ (even if we assume that G is connected).

To understand the proof of Theorem 6.6 we need some definitions. If G is a graph and $S, T \subseteq V(G)$, then $e_G(S, T)$ is the number of ordered pairs $(s, t) \in S \times T$ for which there is an edge in G between s and t ; note that S and T may overlap. If G and G' are graphs with the same vertex-set V , then the *cut distance* is

$$d_{\square}(G, G') = |V|^{-2} \max_{S, T} |e_G(S, T) - e_{G'}(S, T)|.$$

Given a graph G and a positive integer b , we let $G(b)$ denote the b -blow-up of G , where each vertex of G is replaced by b independent copies of itself; thus, $v(G(b)) = bv(G)$ and $e(G(b)) = b^2 e(G)$.

By [30, Theorem 15.1], $q^*(G)$ is estimable for $\mathcal{G}$ if and only if the following two conditions hold:

(i) If graphs G_n, G'_n in $\mathcal{G}$ have the same vertex-set and $d_{\square}(G_n, G'_n) \rightarrow 0$ as $n \rightarrow \infty$, then $q^*(G_n) - q^*(G'_n) \rightarrow 0$ as $n \rightarrow \infty$.

(ii) For every graph G in $\mathcal{G}$, $q^*(G(b))$ has a limit as $b \rightarrow \infty$.

It is straightforward to show that (ii) holds, and in fact $q^*(G(b)) \rightarrow \sup_b q^*(G(b))$ as $b \rightarrow \infty$. By considering separately the edge contribution and degree tax, much as in the proof of Theorem 2.1, we can show that (i) holds for graphs with density at least ρ .

Finally, let us see that (i) fails (badly) in the non-dense case, even for connected graphs. Let $0 \leq \rho(n) < 1$, and let $\rho(n) \rightarrow 0$ arbitrarily slowly as $n \rightarrow \infty$; we may assume that $\rho(n)n/\log n \rightarrow \infty$. Then there are connected graphs G_n, G'_n on the vertex-set $[n]$ with density at least $\rho(n)$, such that each has density $o(1)$, so $d_{\square}(G_n, G'_n) \rightarrow 0$, but $q^*(G_n) \rightarrow 1$ and $q^*(G'_n) \rightarrow 0$. We may take G_n to be a collection of disjoint cliques of size about $2\rho(n)n$, joined by edges to form a path. For G'_n , we may consider a binomial random graph $G_{n,\rho}$, which whp is connected, has density between $\rho/2$ and 2ρ , and by Theorem 3.1(c) has modularity $o(1)$.

7. Modularity and stochastic block models

We focus mostly on balanced stochastic block models. Let $n, k \geq 2$, and let $0 \leq q \leq p \leq 1$: we define the corresponding stochastic block model $G_{n,k,p,q}$ as follows. The vertex-set V is $[n]$, and the random variables σ_v for $v \in V$ are independent, with each uniformly distributed on $[k]$. This yields k random blocks or communities $V_i = \{v \in V : \sigma_v = i\}$, each with expected size n/k , which form the ‘planted partition’ of V . Edges appear independently, with probability p within blocks and with probability q between blocks. Typically we are not told the planted partition.

For each fixed $k \geq 2$, as long as the average degree tends to ∞ , the planted partition whp yields a modularity score which is close to optimal.

Theorem 7.1. Let $k \geq 2$, and let $p = p(n)$ and $q = q(n)$ satisfy $0 \leq q \leq p \leq 1$ and $np \rightarrow \infty$ as $n \rightarrow \infty$. Then

$$q^*(G_{n,k,p,q}) = \frac{(p-q)(1-1/k)}{p + (k-1)q} + o(1) \quad \text{whp,}$$

and whp the planted partition has this modularity value.

A direct proof of this result is given in [35], using ideas from [22]. Given the planted partition, we construct a weighted graph with edge-weights a within communities and b between communities; as we saw in Section 6, there is a natural way to define the modularity of such a weighted network. Using Theorem 6.5, we can show that the stochastic block model has modularity value very close to that of this underlying weighted graph. Thus it remains only to determine the modularity value of a simple deterministic weighted graph, and this is quite straightforward.

Let us now consider a different way to prove a result like Theorem 7.1, at least in the case when the ratio p/q is fixed. We saw above that the planted partition whp yields a close-to-optimal modularity value, but is it close to an optimal partition? The following result, Theorem 7.2 (see [4] and the references therein), answers this question. We can thus give another proof of a result like Theorem 7.1, by using Theorem 7.2 and the robustness result Theorem 2.1 and calculating the likely modularity score of the planted partition.

Theorem 7.2. *Let $a > b > 0$, and let $\omega = \omega(n) \rightarrow \infty$ as $n \rightarrow \infty$. Let $\mathcal{A}^*$ be a modularity-optimal partition of $G_{n,k,p,q}$.*

If $p = a\omega/n$ and $q = b\omega/n$, then whp $\mathcal{A}^$ and the planted partition differ on $o(n)$ vertices.*

If $p = a(\omega \log n)/n$ and $q = b(\omega \log n)/n$, then whp $\mathcal{A}^$ is exactly the planted partition.*

We next define a generalised k -block stochastic block model. Let $k \geq 2$, let $\pi = (\pi_1, \pi_2, \dots, \pi_k)$ with $\pi_i \geq 0$ and $\sum_i \pi_i = 1$, let $\mathbf{P} = [p_{ij}]$ be a $k \times k$ matrix with $0 \leq p_{ij} \leq 1$, let $n \geq 2$, and let $\rho = \rho(n)$ with $0 < \rho \leq 1$. We define $G_{n,\pi,\mathbf{P},\rho}$ as follows. The vertex-set V is $[n]$, and for $v \in V$, we let the random variables σ_v be independent and identically distributed, taking values in $[k]$ with probability π_i of taking value i . This yields the k -block ‘planted partition’, where $V_i = \{v \in V : \sigma_v = i\}$ with expected size $\pi_i n$. The edges within each block V_i appear with probability ρp_{ii} (following [49], we allow loops), and the edges between blocks V_i and V_j appear with probability ρp_{ij} .

When the probabilities within blocks and between blocks can depend on the labels of the blocks, the problem of recovering communities is subtle. A balanced 3-block example is given in [3], in which a partition that puts two blocks together has a higher modularity score than the planted partition. In this example, $p_{11} = 0.06$, $p_{22} = 0.12$, $p_{33} = 0.66$, $p_{12} = p_{23} = 0.04$, $p_{13} = 0$, and $\rho = 1$. Let $\mathcal{A}$ be the planted partition (with three parts), and let $\mathcal{A}'$ be the bipartition which merges blocks 1 and 2. Then whp $q_{\mathcal{A}}(G) \sim 0.84 - 0.54 = 0.3$ and $q_{\mathcal{A}'}(G) \sim 0.92 - 0.58 = 0.34$, where we recall that $\sim$ indicates that equality holds with a $(1 + o(1))$ factor, and where the two terms are the edge contribution and degree tax. In this example, the minimum within-block probability is higher than the maximum between-block probability. For the following theorem, see [3], [4] and [49].

Theorem 7.3. *Let $G_{n,\pi,\mathbf{P},\rho}$ be a generalised k -block stochastic block model as above, and let $\mathcal{A}^*$ be a modularity-optimal partition of $G_{n,\pi,\mathbf{P},\rho}$. Let $\bar{p} = \sum_{i,j} \pi_i \pi_j p_{ij}$, and suppose that, for all $i \neq j$,*

$$p_{ii} > \left(\sum_h \pi_h p_{ih} \right)^2 / \bar{p} \quad \text{and} \quad p_{ij} < \left(\sum_h \pi_h p_{ih} \right) \left(\sum_h \pi_h p_{jh} \right) / \bar{p}. \quad (7.1)$$

Let $\omega = \omega(n) \rightarrow \infty$ as $n \rightarrow \infty$. If $\rho = \omega/n$ then whp $\mathcal{A}^$ and the planted partition differ on $o(n)$ vertices, and if $\rho = (\omega \log n)/n$ then whp $\mathcal{A}^*$ is exactly the planted partition.*

Condition (7.1) on π and $\mathbf{P}$ is equivalent to the following inequalities for the random graph G constructed from $G_{n,\pi,\mathbf{P},\rho}$ by removing loops:

$$\frac{\mathbb{E}[e(V_i)]}{\mathbb{E}[e(G)]} > \frac{\mathbb{E}[\text{vol}(V_i)]^2}{\mathbb{E}[\text{vol}(G)]^2} \quad \text{and} \quad \frac{\mathbb{E}[e(V_i, V_j)]}{2 \mathbb{E}[e(G)]} < \frac{\mathbb{E}[\text{vol}(V_i)] \mathbb{E}[\text{vol}(V_j)]}{\mathbb{E}[\text{vol}(G)]^2}. \quad (7.2)$$

(Here $e(V_i)$ means $e_G(V_i)$, etc.)

Returning to the balanced model $G_{n,k,p,q}$, we note that [4] and [49] show that Theorem 7.2 holds also for modified versions of the modularity function, including the Erdős–Rényi modularity, where the $d_u d_v / 2m$ degree tax term is replaced by $2m/n$. The same papers extend Theorem 7.2 to a ‘degree-corrected’ version of the stochastic block model, using a quality function similar to modularity.

The lower bound on the modularity $q^*(G_{n,p})$ in Theorem 3.2 covers a wide range of probabilities p , and has a stand-alone algorithmic proof. Recall that when $1/n \leq p \leq 1 - c_0/n$, the algorithm *Swap* whp finds a balanced bipartition achieving a modularity score of at least $\alpha \sqrt{(1-p)/np}$ (and the constant α may be taken to be $\frac{1}{5}$ in part of that range). For the special case when $p = c/n$, attractive results on contiguity between binomial random graphs and stochastic block models (see [2] and [39]) yield another proof giving a better constant.

Theorem 7.4. *For each constant $c > 1$, $q^*(G_{n,c/n}) > 0.668/\sqrt{c}$ whp.*

This result may be compared to the value $q^*(G_{n,c/n}) \sim 0.97/\sqrt{c}$ predicted using spin-glass models (see [43]). We recall also the bound in [46] that, for c sufficiently large, whp $q^*(G_{n,c/n}) \geq 0.7632/\sqrt{c}$, which is stronger for large c – see (3.1) for a corresponding lower bound. To prove Theorem 7.4, we can show that, for each $k \geq 2$, whp there is a balanced k -part partition with a modularity score of at least about $f(k)/\sqrt{c}$, for an explicit function $f(k) > 0$. In detail, let $f(2) = \frac{1}{2}$, and for $k \geq 3$ let $f(k) = \sqrt{2(k-1) \log(k-1)}/k$ (recall that $\log$ denotes natural logarithm). Then, given a constant $c > 1$, for each $k \geq 2$ whp there is a balanced k -part partition $\mathcal{A}_k$ such that

$$q_{\mathcal{A}_k}(G_{n,c/n}) \geq f(k)/\sqrt{c} + o(1), \quad (7.3)$$

and choosing $k = 6$ parts here yields the constant given in Theorem 7.4.

We next sketch the proof of (7.3) when $k = 2$, and then briefly discuss the case $k \geq 3$. Let $0 < b \leq a \leq n$, let $p = a/n$ and $q = b/n$, and consider the stochastic block model $G_{n,k,p,q}$.

Balanced bipartitions, $k = 2$

When a and b are close together, the planted bisection model $G_{n,2,p,q}$ is *contiguous* with $G_{n,c/n}$ where $c = (a+b)/2$ – that is, events A_n hold whp in $G_{n,2,p,q}$ if and only if they hold whp in $G_{n,c/n}$. A result of Mossel, Neeman and Sly [39] says precisely that the models are contiguous if and only if $(a-b)^2 \leq 2(a+b)$. It follows that, if we fix $c > 1$ and let $a = c + \sqrt{c}$ and $b = c - \sqrt{c}$, then the models $G_{n,c/n}$ and $G_{n,2,p,q}$ are (just) contiguous. It is thus sufficient to show that whp we have the desired bipartition in $G_{n,2,p,q}$.

As usual, let $\omega = \omega(n) \rightarrow \infty$ as $n \rightarrow \infty$, with $\omega = o(n)$. Also, let $V_1 = \{v : \sigma_v = 1\}$ and $V_2 = \{v : \sigma_v = 2\}$, and let $\mathcal{A}$ be the partition into V_1 and V_2 , the planted partition. We can use Chebyshev’s inequality repeatedly to show that, for $i = 1$ and 2 , whp we have $|V_i| = \frac{1}{2}n + o(\sqrt{\omega n})$, $e(V_i) = \frac{1}{8}an + o(\sqrt{\omega n})$, and $\text{vol}(V_i) = \frac{1}{4}(a+b)n + o(\sqrt{\omega n})$. Since $e(G_{n,2,p,q}) = \frac{1}{4}(a+b)n + o(\sqrt{\omega n})$ whp, we have

$$q_{\mathcal{A}}(G_{n,2,p,q}) = \frac{a}{a+b} - \frac{1}{2} + o(\sqrt{\omega/n}) = \frac{a-b}{2(a+b)} + o(\sqrt{\omega/n}) \quad \text{whp}, \quad (7.4)$$

and so, by our choice of a and b , whp $q_{\mathcal{A}}(G_{n,2,p,q}) = 1/(2\sqrt{c}) + o(\sqrt{\omega/n})$. Further, whp the number of isolated vertices in V_i is $\frac{1}{2}ne^{-c} + o(\sqrt{\omega n})$, for $i = 1, 2$. We may shuffle isolated vertices in a

partition without changing the modularity, so whp we may modify $\mathcal{A}$ to give a balanced partition, as required.

Balanced k -part partitions for $k \geq 3$.

By [2, Theorem 1], for $c = (a + (k - 1)b)/k$, the models $G_{n,c/n}$ and $G_{n,k,p,q}$ are contiguous if $(a - b)^2 < 2ck^2 \ln(k-1)/(k-1)$. Let $a = c + x\sqrt{c}$ and $b = c - x\sqrt{c}/(k-1)$, where $0 < x < f(k)$, and where we take x near the upper bound. We now argue much as when $k = 2$. We see that $G_{n,c/n}$ and $G_{n,k,p,q}$ are contiguous, and use Chebyshev's inequality repeatedly to check that, for the planted partition $\mathcal{A}$, $q_{\mathcal{A}}(G_{n,k,p,q}) = f(k)/\sqrt{c} + o(1)$ whp; finally, we shuffle isolated vertices to obtain a balanced partition.

8. The modularity of random graphs embeddable in a surface

The modularity of a random planar graph is close to 1 whp. More generally, let us consider the modularity of a random graph embeddable in a given surface S . Let $\mathcal{E}^S$ be the class of graphs embeddable in S . We then have the following theorem.

Theorem 8.1. *Let S be any surface, and let $R_n \in_u \mathcal{E}^S$. Then $q^*(R_n) = 1 - o(1)$ whp.*

How can we prove this? The *tree-width* $\text{tw}(G)$ of a graph G is a measure of how similar it is to a tree (see for example [21]); recall that $\Delta(G)$ is the maximum vertex-degree. Consider the following deterministic result [33, Corollary 1.12]. For $m = 1, 2, \dots$, let G_m be a graph with m edges. If $\text{tw}(G_m) \cdot \Delta(G_m) = o(m)$, then $q^*(G_m) \rightarrow 1$ as $m \rightarrow \infty$. But $\text{tw}(R_n) = O(\sqrt{n})$, whp $\Delta(R_n) = O(\log n)$, and whp $e(R_n) = \Theta(n)$, and so Theorem 8.1 follows. Theorem 8.1 can also be deduced easily from the following result of Łasoń and Sulkowska [26].

Theorem 8.2. *Let the class $\mathcal{A}$ of graphs have an excluded minor, and suppose that the maximum degree of any graph G_m in $\mathcal{A}$ with m edges is $o(m)$. Then whp $q^*(G_m) = 1 - o(1)$.*

To deduce Theorem 8.1 from Theorem 8.2, let $\varepsilon(m) = 1/\log m$, say, and let $\mathcal{B}$ be the class of graphs $G \in \mathcal{E}^S$ for which $\Delta(G) \leq \varepsilon(m)m$, where $m = e(G)$. Then Theorem 8.2 applied to $\mathcal{B}$, together with the results noted above that whp $\Delta(R_n) = O(\log n)$ and $e(R_n) = \Theta(n)$, again show that $q^*(R_n) = 1 - o(1)$ whp.

We can extend Theorem 8.1 by considering order-dependent surfaces (see [32]). Let $g = g(n) \geq 0$, and let $\mathcal{E}^g$ be the class of graphs G such that, if G has order n , then $G \in \mathcal{E}^S$ for some surface S with Euler genus at most $g(n)$.

Theorem 8.3.

- (i) *Let $g(n) = o(n/\log^3 n)$ and let $R_n \in_u \mathcal{E}^g$; then $q^*(R_n) = 1 - o(1)$ whp.*
- (ii) *Let $g(n) = o(n/\log n)$ be non-decreasing, and let $R_n \in_u \mathcal{E}^g$; then there is a subsequence $1 \leq n_1 < n_2 < \dots$ for which $q^*(R_{n_k}) = 1 - o(1)$ whp.*

To prove part (i), we can show that $\Delta(R_n)$ is still $O(\log n)$ whp (see [32, Theorem 8(a)]). Call a set F of edges in a graph G a *planarising edge-set* if $G \setminus F$ is planar. A result of Djidjev and Venkatesan (see [32, Lemma 27]) shows that, for any $n \geq 2$ and $h \geq 0$, any connected graph $G \in \mathcal{E}_n^h$ has a planarising edge-set of size at most $4\sqrt{h(n+h)\Delta(G)}$. So for R_n , whp we have a planarising edge-set F_n of size $o(n)$, and so $|F_n|/e(R_n) = o(1)$. For the planar graph $R_n^- = R_n \setminus F_n$, whp

$q^*(R_n^-) = 1 + o(1)$, by the proof of Theorem 8.1, and by the robustness result Theorem 2.1, we have $q^*(R_n) = q^*(R_n^-) + o(1) = 1 + o(1)$ whp.

We can prove part (ii) similarly; for by [32, Theorem 8(b)], there are a subsequence $n_1 < n_2 < \dots$ and a constant c such that $\Delta(R_{n_k}) \leq c \log(n_k)$ whp (as $k \rightarrow \infty$).

9. Some known modularity values

We conclude with a list of some known modularity values of graphs, sorted into four classes, those with modularity near to 1, modularity bounded away from 1 and from 0, modularity near to 0, and modularity exactly 0. Knowing the modularity for classes of graphs may help us to understand the behaviour of the modularity function. We list results on the modularity of the binomial random graph $G_{n,p}$: as we saw earlier, there are similar results for the Erdős–Rényi random graph $G_{n,m}$. We always use n for the number of vertices and m for the number of edges. In the second column we give extra conditions, if these are necessary, and in the last column we refer to a result in this chapter or to the appropriate reference.

maximally modular			
cycle		$q^*(C_n) = 1 - 2/\sqrt{n} (1 + o(1))$	[8]
tree	$\Delta(T_n) = o(n)$	$q^*(T_n) \geq 1 - 2(2\Delta/n)^{1/2}$	[33]
tree-like (low tree-width)	$\Delta(G_m)\text{tw}(G_m) = o(m)$	$q^*(G_m) \geq 1 - 2((\text{tw} + 1)\Delta/m)^{1/2}$	[33]
minor-free graphs	fixed forbidden minor & $\Delta(G_m) = o(m)$	$q^*(G_m) = 1 + o(1)$	Thm 8.2
binomial / Erdős-Rényi	$n^2p \rightarrow \infty$ & $np \leq 1 + o(1)$	whp $q^*(G_{n,p}) = 1 + o(1)$	Thm 3.1
random hyperbolic		whp $q^*(G_n) = 1 + o(1)$	Thm 5.3
random 2-regular		whp $q^*(G_{n,2}) = 1 - 2/\sqrt{n} + o(1)$	Eqn (4.2)
random geometric		whp $q^*(G_n) = 1 + o(1)$	[10]
random planar (or any fixed surface)		whp $q^*(G_n) = 1 + o(1)$	Thm 8.1

critically modular			
binomial / Erdős-Rényi	$np = c > 1 \Rightarrow \exists \varepsilon > 0 :$ $np = c \geq 1$	whp $\varepsilon < q^*(G_{n,p}) < 1 - \varepsilon$ whp $0.668/\sqrt{c} < q^*(G_{n,p}) < 3.06/\sqrt{c}$	Thm 3.1 Eqn (3.1), Thm 7.4
random cubic		whp $0.667 < q^*(G_{n,3}) < 0.790$	Thm 4.1.
random r -regular	for fixed $r \geq r_0$	whp $0.76/\sqrt{r} < q^*(G_{n,r}) < 2/\sqrt{r}$	Thm 4.2
preferential attachment	$h \geq 2$ edges per step	whp $\Omega(1/\sqrt{h}) = q^*(G_n^h) = O(\sqrt{\log h}/\sqrt{h})$	Thm 5.1.
stochastic block model (k balanced blocks)	$0 \leq q \leq p \leq 1$ & $np \rightarrow \infty$	whp $q^*(G_n) = \frac{(p-q)(1-1/k)}{p+(k-1)q} + o(1)$	Thm 7.1

minimally modular			
binomial / Erdős-Rényi	$np \rightarrow \infty$	whp $q^*(G_{n,p}) = o(1)$	Thm 3.1

non-modular			
complete		$q^*(K_n) = 0$	[8]
complete multipartite		$q^*(K_{n_1, n_2, \dots, n_k}) = 0$	[6, 36, 37]
nearly complete	$e(G) \geq \binom{n}{2} - n/2$	$q^*(G) = 0$	[36]
binomial / Erdős-Rényi	$p \geq 1 - c/n, \quad c < 1$	whp $q^*(G_{n,p}) = 0$	Thm 3.5

References

1. V. Agdur, N. Kamčev and F. Skerman, Universal lower bound for community structure of sparse graphs, arXiv:2307.07271, (2023).
2. J. Banks, C. Moore, J. Neeman and P. Netrapalli, Information-theoretic thresholds for community detection in sparse networks, *Conference on Learning Theory* (2016), 383–416.
3. P. J. Bickel and A. Chen, A nonparametric view of network models and Newman–Girvan and other modularities, *Proc. Nat. Acad. Sci.* **106** (50) (2009), 21068–21073.
4. P. J. Bickel, A. Chen, Y. Zhao, E. Levina and J. Zhu, Correction to the proof of consistency of community detection, *Ann. Statist.* **43** (2015), 462–466.
5. V. D. Blondel, J. Guillaume, R. Lambiotte and E. Lefebvre, Fast unfolding of communities in large networks, *J. Stat. Mech. Theory Exp.* **2008** (10) (2008), P10008.
6. M. Bolla, B. Bullins, S. Chaturapruek, S. Chen and K. Friedl, Spectral properties of modularity matrices, *Linear Algebra Appl.* **473** (2015), 359–376.
7. C. Borgs, J. T. Chayes, L. Lovász, V. T. Sós and K. Vesztegombi, Convergent sequences of dense graphs I: Subgraph frequencies, metric properties and testing, **219** (6) (2008), 1801–1851.
8. U. Brandes, D. Delling, M. Gaertler, R. Gorke, M. Hoefer, Z. Nikoloski and D. Wagner, On modularity clustering, *Knowledge and Data Engineering, IEEE* **20** (2) (2008), 172–188.
9. J. Chellig, N. Fountoulakis and F. Skerman, The modularity of random graphs on the hyperbolic plane, *J. Complex Networks* **10** (1) (2022).
10. E. Davis and S. Sethuraman, Consistency of modularity clustering on random geometric graphs, *Ann. Appl. Probab.* **28** (4) (2018), 2003–2062.
11. A. Dembo, A. Montanari and S. Sen, Extremal cuts of sparse random graphs, *Ann. Probab.* **45** (2) (2017), 1190–1217.
12. J. Díaz and D. Mitsche, The cook-book approach to the differential equation method, *Comp. Sci. Rev.* **4** (3) (2010), 129–151.
13. T. N. Dinh and M. T. Thai, Finding community structure with performance guarantees in scale-free networks, *Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third Internat. Conf. on Social Computing (Social-Com)*, IEEE (2011), 888–891.
14. D. Elgesem, L. Steskal and N. Diakopoulos, Structure and content of the discourse on climate change in the blogosphere: The big picture, *Environmental Communication* (2014), 21–40.
15. A. Ferber, M. Kwan, B. Narayanan, A. Sah and M. Sawhney, Friendly bisections of random graphs, *Comm. Amer. Math. Soc.* **210** (2022), 380–416.
16. S. Fortunato and M. Barthélemy, Resolution limit in community detection, *Proc. Nat. Acad. Sci.* **104** (1) (2007), 36–41.
17. S. Fortunato and D. Hric, Community detection in networks: A user guide, *Phys. Rep.* **659** (2016), 1–44.
18. N. Fountoulakis, P. Van Der Hoorn, T. Müller and M. Schepers, Clustering in a hyperbolic model of complex networks, *Electron. J. Probab.* **26** (2021), 1–132.
19. J. Friedman, A proof of Alon’s second eigenvalue conjecture and related problems, *Mem. Amer. Math. Soc.* **195** (910) (2008), viii–100.
20. S. Janson, T. Luczak and A. Rucinski, *Random Graphs*, Wiley, 2011.
21. T. Kloks, *Treewidth: Computations and Approximations*, Springer, 1994.
22. M. Koshelev, Modularity in planted partition model, *Computational Management Science* **20** (34) (2023).
23. D. Krioukov, F. Papadopoulos, M. Kitsak, A. Vahdat and M. Boguná, Hyperbolic geometry of complex networks, *Phys. Rev. E* **82** (3) 036106 (2010).
24. R. Lambiotte and M. T. Schaub, *Modularity and Dynamics on Complex Networks*, Cambridge Univ. Press, 2021.
25. A. Lancichinetti and S. Fortunato, Limits of modularity maximization in community detection, *Phys. Rev. E.* **84** (6) 066122 (2011).
26. M. Lasoń and M. Sulkowska, Modularity of minor-free graphs, *J. Graph Theory* **102** (4) (2023), 728–736.

27. L. Lichev and D. Mitsche, On the modularity of 3-regular random graphs and random graphs with given degree sequences, *Random Structures Algorithms* **61** (4) (2022), 754–802.
28. A. Liebenau and N. Wormald, Asymptotic enumeration of graphs by degree sequence, and the degree sequence of a random graph, *J. Europ. Math. Soc.* **26** (1) (2024), 1–40.
29. B. Louf, C. McDiarmid and F. Skerman, Modularity and graph expansion, 15th Innovations in Theoretical Computer Science Conference (ITCS 2024) (2024), 78:1–78:21.
30. L. Lovász, *Large Networks and Graph Limits*, Vol. 60, *Amer. Math. Soc. Colloq. Publ.*, 2012.
31. D. Lusseau, The emergent properties of a dolphin social network, *Proc. Roy. Soc. London, Series B: Biological Sciences* **270** (2003), S186–S188.
32. C. McDiarmid and S. Saller, Random graphs embeddable in order-dependent surfaces, *Random Structures Algorithms* **64** (4) (2023), 1–46.
33. C. McDiarmid and F. Skerman, Modularity of regular and treelike graphs, *J. Complex Networks* **6** (4) (2018), 596–619.
34. C. McDiarmid and F. Skerman, Modularity of Erdős–Rényi random graphs, *Random Structures Algorithms* **57** (1) (2020), 211–243.
35. C. McDiarmid and F. Skerman, Modularity and partially observed graphs, arXiv:2112.13190, (2021).
36. C. McDiarmid and F. Skerman, Modularity of nearly complete graphs and bipartite graphs, arXiv:2311.06875, (2023).
37. S. Majstorovic and D. Stevanovic, A note on graphs whose largest eigenvalues of the modularity matrix equals zero, *Electron. J. Linear Algebra* **27** (1) (2014), 611–618.
38. K. Meeks and F. Skerman, The Parameterised Complexity of Computing the Maximum Modularity of a Graph, *Algorithmica*, **82** (8) (2020), 2174–2199.
39. E. Mossel, J. Neeman and A. Sly, Reconstruction and estimation in the planted partition model, *Probab. Theory & Related Fields* **162** (2015), 431–461.
40. M. E. J. Newman and M. Girvan, Finding and evaluating community structure in networks, *Phys. Rev. E* **69** (2) (2004), 026113.
41. M. A. Porter, J. Onnela and P. J. Mucha, Communities in networks, *Notices Amer. Math. Soc.* **56** (9) (2009), 1082–1097.
42. L. O. Prokhorenkova, A. Raigorodskii and P. Prałat, Modularity of complex networks models, *Internet Math.* (2017), 115–126.
43. J. Reichardt and S. Bornholdt, When are networks truly modular?, *Phys. D.* **224** (1) (2006), 20–26.
44. R. Robinson and N. C. Wormald, Almost all cubic graphs are Hamiltonian, *Random Structures Algorithms* **3** (2) (1992), 117–125.
45. K. Rybarczyk and M. Sulkowska, Modularity of preferential attachment graphs, arXiv:2501.06771, (2025).
46. K. Rybarczyk and M. Sulkowska, New bounds on the modularity of $G(n, p)$, arXiv:2504.16254, (2025).
47. F. Skerman, *Modularity of Networks*, D.Phil. Thesis, University of Oxford, 2016.
48. V.A. Traag, L. Waltman and N. J. Van Eck, From Louvain to Leiden: guaranteeing well-connected communities, *Scientific reports* **9** (1) (2019), 1–12.
49. Y. Zhao, E. Levina and J. Zhu, Consistency of community detection in networks under degree-corrected stochastic block models, *Ann. Statist.* **40** (4) (2012), 2266–2292.