

On graphs with modularity zero or near-zero

Colin McDiarmid^a

Fiona Skerman^b

Submitted: Dec 22, 2023; Accepted: Jul 7, 2026; Published: Aug 28, 2026

© The authors. Released under the CC BY-ND license (International 4.0).

Abstract

It is known that complete graphs and complete multipartite graphs have modularity zero. We show that the least number of edges we may delete from the complete graph K_n to obtain a graph with non-zero modularity is $\lfloor n/2 \rfloor + 1$. Similarly we determine the least number of edges we may delete from or add to a complete bipartite graph to reach non-zero modularity. We give some corresponding results for complete multipartite graphs, and a short proof that complete multipartite graphs have modularity zero.

We also analyse the modularity of very dense random graphs, and in particular we find that there is a transition to modularity zero when the average degree of the complementary graph drops below 1.

Finally we consider some natural variants of the definition of modularity; and investigate which graphs have corresponding modularity value 0, and the least number of edges we may delete from the complete graph K_n to obtain a graph with non-zero modularity.

Mathematics Subject Classifications: 05C69, 68R10, 91D30

1 Introduction

The modularity $q^*(G)$ of a graph G was introduced by Newman and Girvan in 2004 [31], to give a measure of how well G can be divided into ‘communities’, and now modularity is at the heart of the most popular algorithms used to cluster real data [17], including the Louvain [5] and Leiden [36] algorithms. See Section 1.2 below for precise definitions and further discussion.

We are interested here in the question of which graphs have modularity zero or near zero. We saw in [26] that for suitable random graphs R_n , if the average degree of R_n tends to ∞ as $n \rightarrow \infty$ then $q^*(R_n) \rightarrow 0$ in probability – details are given below. But when is $q^*(R_n) > 0$ whp and when is it actually 0 whp? (For a sequence of events A_n we say that A_n holds *with high probability (whp)* if $\mathbb{P}(A_n) \rightarrow 1$ as $n \rightarrow \infty$.) A key step in understanding this is deterministic.

^aDepartment of Statistics, University of Oxford, United Kingdom (cmcd@ox.ac.uk).

^bDepartment of Mathematics, Uppsala University, Sweden (fiona.skerman@math.uu.se).

It is known [8] and easy to see that complete graphs have modularity zero. Graphs which are close to a complete graph in that there are very few missing edges also have modularity zero. We give the precise threshold in Theorem 1 below.

Each graph G with at most 3 vertices has modularity $q^*(G) = 0$. (By convention a graph G with no edges has $q^*(G) = 0$.) The 4-vertex path has modularity $\frac{1}{6}$, and the 4-vertex graph consisting of two disjoint edges has modularity $\frac{1}{2}$; and indeed for each $n \geq 4$ there is an n -vertex graph G with $q^*(G) > 0$. Given an integer $n \geq 4$, let $f^q(n)$ be the least number of edges missing from any n -vertex graph G with $q^*(G) > 0$.

Theorem 1. *For each $n \geq 4$ we have $f^q(n) = \lfloor n/2 \rfloor + 1$.*

This theorem corrects Lemma 9 and the corresponding part of Theorem 2 in [37], which in the notation here says that $f^q(n) \geq n - 2$: see Section 1.3 below for details. We also give results on how slight changes to the definition of modularity affect the result in Theorem 1 – see the table in Figure 3 on page 29 for a summary.

The following result gives a characterisation of the set of graphs which are extremal in the sense of Theorem 1. We first consider the case $n = 4$, before covering the general case $n \geq 5$ in the proposition. By Theorem 1 we have $f^q(4) = 3$. It is easy to see that a set F of 3 edges in K_4 satisfies $q^*(K_4 - F) > 0$ if and only if F forms a path. Note that if F is a star (which is a bipartite graph), then $K_4 - F = \text{star} - \text{star} = \text{isolated vertex}$, i.e., K_3 plus an isolated vertex, which has modularity zero.

Proposition 2. *Let $n \geq 5$. For graphs G on $[n]$ with $e(\overline{G}) = f^q(n)$, $q^*(G) > 0$ if and only if $\overline{G}$ is bipartite.*

A result related to this proposition, see Remark 13, will be useful to us in Example 20.

Given a graph H , let $\delta^-(H)$ be the least number of edges we can delete to obtain a subgraph H' with $q^*(H') > 0$. If there is no subgraph H' with $q^*(H') > 0$ we set $\delta^-(H) = \infty$. (The only graph with $m \geq 4$ edges with this property is the m -edge star $K_{1,m}$. This can be seen by noting that $\delta^-(H)$ is finite as soon as H contains two disjoint edges.) In this notation, Theorem 1 says that $\delta^-(K_n) = \lfloor n/2 \rfloor + 1$ for $n \geq 4$, so large complete graphs have modularity 0 ‘robustly’.

To recap, any complete graph has modularity zero, and this is robust in the sense that graphs ‘near’ complete graphs also have modularity zero. In contrast, while any complete multipartite graph H has modularity zero [6, 23] (see Section 3 below for a short proof), here there is no robustness in general. We focus on complete bipartite graphs, where we can tell the complete story about robustness reasonably quickly: we find that, except for a few small examples, it is enough to add or remove a single edge to obtain a graph with positive modularity. In Section 2.5 we give some corresponding partial robustness results for complete multipartite graphs, and in Section 3 we use a recent result on modularity and graph expansion [20] to give a short and simple proof that complete multipartite graphs have modularity 0.

To state the results, let us define natural counterparts to δ^- : given a graph H , let $\delta^+(H)$ be the least number of edges we can *add* (between already existing vertices) to

obtain a graph H' with $q^*(H') > 0$ (where $\delta^+(H) = \infty$ if there is no such graph H'). Likewise let $\delta(H)$ be the minimum number of edits, either adding or removing edges, to obtain a graph H' with positive modularity (where $\delta(H) = \infty$ if there is no such graph H' , that is if $v(H) \leq 3$).

Theorem 3. *Let $G = K_{s,t}$ be a complete bipartite graph with $s \leq t$.*

- (a) *If $s = 1$ then $\delta^-(G) = \infty$, and if $s \geq 2$ then $\delta^-(G) = 1$.*
- (b) *If $s = 1$ and $t \geq 4$, or if $s \geq 2$ and $t \geq 3$, then $\delta^+(G) = 1$. In the other cases (namely when $(s, t) = (1, 1), (1, 2), (1, 3)$ or $(2, 2)$), we have $\delta^+(G) = \infty$.*
- (c) *If $s = 1$ and $t \geq 4$, or if $s \geq 2$, then $\delta(G) = 1$. If $(s, t) = (1, 1)$ or $(1, 2)$ then $\delta(G) = \infty$, and in the remaining case $(s, t) = (1, 3)$ we have $\delta(G) = 2$.*

Theorems 1 and 3 suggest that complete graphs K_n may be the ‘furthest’ from graphs with positive modularity.

Question 4. Is it true that, for each graph H with $n \geq 4$ vertices, we have $\delta(H) \leq \delta(K_n)$?

Now consider random graphs. The modularity of random graphs $G_{n,p}$ and $G_{n,m}$ (see Section 1.2 below for definitions) is quite fully analysed in [26], except for the very dense case. By Theorem 1.3 of [26], when $p = p(n)$ satisfies $p \geq 1/n$ and p is bounded away from 1, we have $q^*(G_{n,p}) = \Theta((np)^{-1/2})$ whp; and by Theorem 4.1 of the same paper, when $1/n \leq p \leq 1 - c_0/n$ for a suitable (large) constant c_0 we have $q^*(G_{n,p}) = \Omega\left(\sqrt{\frac{1-p}{np}}\right)$ whp. We focus here on the very dense case, when $p = 1 - \Theta(1/n)$. Note that in this case $\sqrt{\frac{1-p}{np}} = \Theta(1/n)$. It follows easily from Theorem 1 that when $1 - p$ is very small, say $1 - p \leq (1 - \varepsilon)/n$ (for some $\varepsilon > 0$), then whp $q^*(G_{n,p}) = 0$. We shall see that when $1 - p = 1/n$, two different behaviours can occur each with probability about $\frac{1}{2}$, namely $q^*(G_{n,p}) = 0$ and $q^*(G_{n,p}) = \Theta(n^{-3/2})$; and if $1 - p$ is just a little larger, say $1 - p \geq (1 + \varepsilon)/n$ (for some $\varepsilon > 0$), then whp $q^*(G_{n,p}) = \Theta(1/n)$.

Theorem 5. (a) *If $p = p(n)$ 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 (i) $\mathbb{P}(q^*(G_{n,p}) = 0) = \frac{1}{2} + o(1)$, and (ii) given $\varepsilon > 0$ there exist $0 < \alpha < \beta$ such that $\mathbb{P}(\alpha n^{-3/2} < q^*(G_{n,p}) < \beta n^{-3/2}) \geq \frac{1}{2} - \varepsilon$.*

- (c) *Given $1 < c_1 < c_2$, there exist $0 < \alpha < \beta$ 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.*

The lower bound in part (c) of Theorem 5 allows us to extend the range of p in Theorem 4.1 of [26] up to nearly $1 - 1/n$ (and part (a) shows that we cannot go much further). Using Theorem 4.1 of [26] for the lower parts of the range we obtain the following result.

Theorem 6. *Given $\varepsilon > 0$ there exists $\alpha > 0$ such that, if $p = p(n)$ satisfies $1/n \leq p \leq 1 - (1 + \varepsilon)/n$, then $q^*(G_{n,p}) \geq \alpha \sqrt{\frac{1-p}{np}}$ whp.*

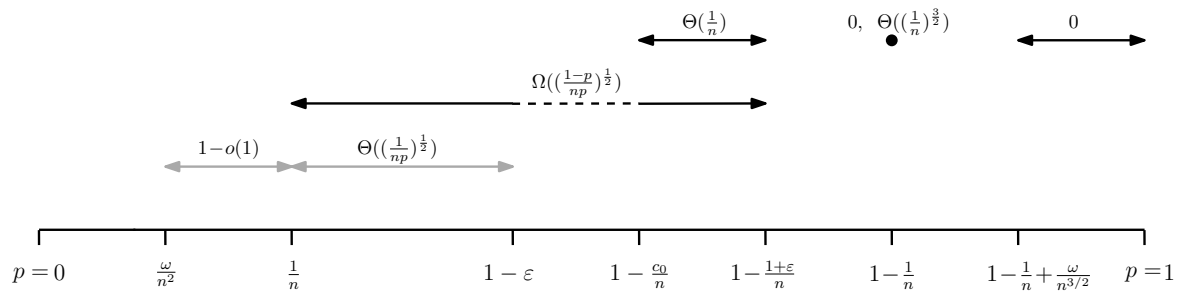


Figure 1: An illustration of the bounds known to hold whp on $q^*(G_{n,p})$ for different ranges of $p = p(n)$. On the top level, we depict the results of Theorem 5; the left interval corresponds to part (a), the middle to part (b) and the right to part (c). For $p = 1 - 1/n$, the modularity of $G_{n,p}$ has probability about a half of being zero, and about a half of being of order $n^{-3/2}$, see part (b) of Theorem 5. On the middle level, we depict the result of Theorem 6, which extends earlier results from [26] shown in grey on the bottom level.

1.1 Outline of the paper

In the remainder of this section, we give definitions and preliminaries concerning modularity, and describe some relevant previous work. In Section 2, we give proofs for the new deterministic results above (Theorem 1, Proposition 2, and Theorem 3), and for some related results on complete multipartite graphs; in Section 3 we give a short new proof (based on a recent expansion result) that complete multipartite graphs have modularity zero; in Section 4 we consider random graphs and prove Theorems 5 and 6 above; in Section 5 we discuss some natural variants of the definition of modularity (for example where we insist that the random graph with the given degrees is simple), and how these changes would impact our results; and in the final section we make a few concluding remarks.

1.2 Definitions and preliminaries

Given a non-empty graph G (with at least one edge), we give a modularity score to each vertex-partition (or ‘clustering’): the modularity $q^*(G)$ (sometimes called the ‘maximum modularity’) of G is defined to be the maximum of these scores over all vertex-partitions. For a set A of vertices, let the *volume* $\text{vol}(A)$ or $\text{vol}_G(A)$ be the sum over the vertices v in A of the degree d_v , also let $e(A)$ or $e_G(A)$ denote the number of edges within A ; and for B disjoint from A , let $e(A, B)$ or $e_G(A, B)$ denote the number of edges between A and B . Given a vertex-partition $\mathcal{A}$, let $\delta_{ij}(\mathcal{A})$ be the indicator that vertices i and j are in the same part in the partition (so $\delta_{ii}(\mathcal{A})$ is always 1). Also, denote by $e_{\mathcal{A}}^{\text{int}}(G)$ the number of edges in G with both endpoints in the same part of $\mathcal{A}$.

Definition 7 (Newman & Girvan [31], see also Newman [30]). Let $G = (V, E)$ be a graph with $m \geq 1$ edges. For a partition $\mathcal{A}$ of V , the modularity score of $\mathcal{A}$ on G is

$$q_{\mathcal{A}}(G) = \frac{1}{2m} \sum_{i,j \in V} \left(\mathbf{1}_{ij \in E} - \frac{d_i d_j}{2m} \right) \delta_{ij}(\mathcal{A}) = \frac{1}{m} \sum_{A \in \mathcal{A}} e(A) - \frac{1}{4m^2} \sum_{A \in \mathcal{A}} \text{vol}(A)^2;$$

and the modularity of G is $q^*(G) = \max_{\mathcal{A}} q_{\mathcal{A}}(G)$, where the maximum is over all partitions $\mathcal{A}$ of V .

Isolated vertices are irrelevant. We need to give empty graphs (with no edges) some modularity value: conventionally we set $q^*(G) = 0$ for each such graph G , and we set $q_{\mathcal{A}}(G) = 0$ for every partition $\mathcal{A}$ of $V(G)$. The second equation for $q_{\mathcal{A}}(G)$ expresses modularity as the difference of two terms, the *edge contribution* or *coverage* $q_{\mathcal{A}}^E(G) = \sum_A e(A)/m = e_{\mathcal{A}}^{\text{int}}(G)/m$, and the *degree tax* $q_{\mathcal{A}}^D(G) = \sum_A \text{vol}(A)^2/(4m^2)$. Since, for any non-empty graph, we have $q_{\mathcal{A}}^E(G) \leq 1$ and $q_{\mathcal{A}}^D(G) > 0$, we have $q_{\mathcal{A}}(G) < 1$ for any such graph G . Also, the trivial partition $\mathcal{A}_0$ with all vertices in one part has $q_{\mathcal{A}_0}^E(G) = q_{\mathcal{A}_0}^D(G) = 1$, so $q_{\mathcal{A}_0}(G) = 0$. Thus we have

$$0 \leq q^*(G) < 1.$$

Suppose that we use the configuration model (see for example [7, 12, 14]) to pick a ‘nearly uniform’ random pseudograph R (where multiple edges and loops are allowed) with degree sequence $(d_1, \dots, d_n)$ where $\sum_i d_i = 2m$. Then the expected number of edges in R between distinct vertices i and j is $d_i d_j / (2m - 1)$; and it follows (see (5.5)) that, if m is large, the degree tax is approximately the expected value of $e_{\mathcal{A}}^{\text{int}}(R)/m$, see (5.4) for exact calculations. This is the original rationale for the definition of the modularity score: whilst rewarding the partition for capturing a high proportion of edges within the parts, we should penalise by (approximately) the expected proportion of such edges. In Section 5 we amplify this discussion, and consider other possible distributions for the random (pseudo)graph R . Then we investigate when we get modularity 0 with some natural variant versions of this definition.

Random graphs

Let n be a positive integer. Given $0 \leq p \leq 1$, the (binomial) random graph $G_{n,p}$ has vertex set $[n] := \{1, \dots, n\}$ 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 m -edge graphs on vertex set $[n]$. These two random graphs are closely related when $m \approx \binom{n}{2}p$, see for example [12, 22].

Recall, that for a sequence of events A_n we say that A_n holds *with high probability* (*whp*) if $\mathbb{P}(A_n) \rightarrow 1$ as $n \rightarrow \infty$. For a sequence of random variables X_n and a real number a , we write $X_n \xrightarrow{p} a$ if X_n converges in probability to a as $n \rightarrow \infty$ (that is, if for each $\varepsilon > 0$ we have $|X_n - a| < \varepsilon$ whp). For $x = x(n)$ and $y = y(n)$ we write $x \sim y$ to indicate that $x = (1 + o(1))y$ as $n \rightarrow \infty$.

More on modularity

Let $G = (V, E)$ be a nonempty graph, and let $\nu = \text{vol}(G)$. For $U \subseteq V$, let $p_G(U) = 2e(U)\nu - \text{vol}(U)^2$. Observe that $p_G(\emptyset) = p_G(V) = 0$. If $\mathcal{A}$ is a partition of $V(G)$, then $q_{\mathcal{A}}(G) = \nu^{-2} \sum_{A \in \mathcal{A}} p_G(A)$.

There are two useful alternative expressions for $p_G(U)$. Since $\text{vol}(U) = 2e(U) + e(U, \bar{U})$

(where $\overline{U}$ denotes $V \setminus U$), we have

$$\begin{aligned} p_G(U) &= 2e(U)\nu - \text{vol}(U)^2 = (\text{vol}(U) - e(U, \overline{U}))\nu - \text{vol}(U)(\nu - \text{vol}(\overline{U})) \\ &= -e(U, \overline{U})\nu + \text{vol}(U)\text{vol}(\overline{U}). \end{aligned}$$

Thus

$$p_G(U) = \text{vol}(U)\text{vol}(\overline{U}) - e(U, \overline{U})\nu. \quad (1.1)$$

Similarly

$$p_G(U) = 4e(U)(e(U) + e(U, \overline{U}) + e(\overline{U})) - (2e(U) + e(U, \overline{U}))^2,$$

and so

$$p_G(U) = 4e(U)e(\overline{U}) - e(U, \overline{U})^2. \quad (1.2)$$

These expressions for $p_G(U)$ are symmetric in U and $\overline{U}$, so $p_G(U) = p_G(\overline{U})$. Now, if $\mathcal{A}$ is a bipartition with parts U and $\overline{U}$, then

$$q_{\mathcal{A}}(G) = \nu^{-2}(p_G(U) + p_G(\overline{U})) = 2\nu^{-2}p_G(U).$$

Thus there is a bipartition $\mathcal{A}$ with $q_{\mathcal{A}}(G) > 0$ if and only if there is a set U of vertices with $p_G(U) > 0$, if and only if $q^*(G) > 0$. We have thus shown the following result, which we record as a lemma.

Lemma 8. *Let G be a graph with at least one edge. Then the following are equivalent : (i) $q^*(G) = 0$, (ii) $q_{\mathcal{A}}(G) \leq 0$ for each bipartition $\mathcal{A}$ of $V(G)$, and (iii) $p_G(U) \leq 0$ for each $U \subseteq V(G)$.*

Recall the very nice result of Dinh and Thai [11], that for any $k \geq 2$ there is a partition $\mathcal{A}$ with at most k parts such that $q_{\mathcal{A}}(G) \geq q^*(G)(1 - 1/k)$. Note that this implies the equivalence of (i) and (ii) in the lemma (that is, $q^*(G) > 0$ if and only if there is a bipartition $\mathcal{A}$ with $q_{\mathcal{A}}(G) > 0$).

Robustness

The following ‘robustness’ result will be used in the proof of Theorem 5. It is essentially Lemma 5.1 in [26]. Since we set $q^*(G) = 0$ if G has no edges, we can rephrase the lemma there as the following.

Lemma 9. *If $G = (V, E)$ is a graph, E_0 is a non-empty subset of E , and $G' = (V, E \setminus E_0)$, then*

$$|q^*(G) - q^*(G')| < 2|E_0|/|E|.$$

1.3 Relevant previous work on modularity

As discussed above the following graphs are known to have zero modularity: complete graphs [8] and complete multipartite graphs [6, 23]. There are some general lower bounds which can show strictly positive modularity : if we let $\bar{d}$ denote the average degree of a graph, then connected graphs with maximum degree $o(n)$ have modularity at least $2/\bar{d} - o(1)$ [33], and both preferential attachment graphs [1, 33] and deterministic graphs with mild assumptions on the degree sequence [1] have modularity $\Omega(\bar{d}^{-1/2})$.

At the other extreme, which graphs have modularity near 1? In [29], de Montgolfier, Soto and Viennot define a sequence of graphs to be *maximally modular* if their modularity values tend to 1 as their number m of edges tends to infinity. The following are known to be maximally modular: trees with maximum degree $o(m)$, graphs of bounded genus with maximum degree $o(m)$ [18], graphs where treewidth times maximum degree is $o(m)$ [25], lattices [29, 35], whp random graphs $G_{n,p}$ and $G_{n,m}$ with average degree at most $1+o(1)$ [26], whp random hyperbolic graphs [10], and whp spatial preferential attachment graphs [33]. There is also work on when modularity is able to pick up planted structure in random graphs; the stochastic block model (SBM) is considered in [2, 3], the degree corrected SBM in [4, 39] and a related ABCD model in [16].

We consider alternative definitions for the degree tax in the modularity score in Section 5 of this paper. We note that several definitions for the modularity score exist for hypergraphs [15, 32] – see [32] for a comparison of the different definitions.

For a summary of known modularity values, see the table in [24]. Notable results include improved bounds for random regular graphs, which were obtained in [19] – in particular the modularity of random cubic graphs was shown to lie in the interval $[0.667, 0.79]$ whp. Additionally, for the preferential attachment model G_n^h (where $h \geq 2$ edges are added with each new vertex), the modularity satisfies $\Omega(1/\sqrt{h}) = q^*(G_n^h) = O(\sqrt{\log h}/\sqrt{h})$ whp. Here the lower bound is known to hold for generalised preferential attachment models [1, 33], and the upper bound follows from results showing the expansion of a set of vertices is linked to the ‘ages’ of those vertices – see [34].

Lastly, it is known that for any n -vertex base graph G , letting G_p be the ‘percolated’ random graph obtained by keeping each edge with probability p , then $q^*(G_p) \rightarrow q^*(G)$ in probability if $e(G)p/n \rightarrow \infty$ as $n \rightarrow \infty$ [27].

As mentioned earlier, Theorem 1 corrects Lemma 9 and the corresponding part of Theorem 2 in [37], which claimed that $\max_{|\mathcal{A}|=2} q_{\mathcal{A}}(G) < 0$ if $e(G) \in [(\binom{n-1}{2} + 2, \binom{n}{2})]$ (where the maximum is over all bipartitions $\mathcal{A}$). This claim is also restated in Theorem 2 of [37], as the third range of the number m of edges. Recall, for example by Lemma 8, that $q^*(G) > 0$ if and only if there exists a bipartition $\mathcal{A}$ with $q_{\mathcal{A}}(G) > 0$. Note also that $\binom{n-1}{2} + 2 = \binom{n}{2} - n + 3$. Thus, if Lemma 9 of [37] were correct, it would imply that if $e(G) \geq \binom{n}{2} - n + 3$ then $q^*(G) = 0$; that is, that $f^q(n) \geq n - 2$, in contradiction to Theorem 1.

2 Proofs for new deterministic results

In this section we first prove Theorem 1 on $f^q(n) = \delta^-(K_n)$, and then quickly prove Proposition 2 which characterises the corresponding extremal graphs. Then, after a couple of preliminary lemmas, we prove Theorem 3, in which we determine $\delta^-(G)$, $\delta^+(G)$ and $\delta(G)$ for all complete bipartite graphs G . Finally we give a couple of examples and briefly consider multipartite graphs. (Recall that $\delta^+, \delta^-, \delta$ are the minimum numbers of edges that may be added, removed or edited to obtain a graph with positive modularity - see the paragraphs preceding Theorem 3 for definitions.)

2.1 Nearly complete graphs: proof of Theorem 1

We first give a straightforward proof of the upper bound on $f^q(n) (= \delta^-(K_n) = \delta(K_n))$ for $n \geq 4$. We proceed via a slightly stronger result, since it will be helpful to refer back to it later. For disjoint vertex sets A and B , write $E(A, B)$ for the set of edges between A and B .

Claim 10. *Let $n \geq 4$, let $\mathcal{A} = (A, B)$ be a bipartition of $[n]$ where $|A| = \lceil n/2 \rceil$ and $|B| = \lfloor n/2 \rfloor$, and let G be obtained from K_n by deleting any $\lfloor n/2 \rfloor + 1$ edges from $E(A, B)$. Then $q^*(G) > 0$.*

Proof of Claim 10. Let $F = E(A, B)$ be the set of edges between the parts, and let $k = \lfloor n/2 \rfloor + 1$. Note that if n is even then $|F| = n^2/4 \geq n \geq k$, and if n is odd then $|F| = (n^2 - 1)/4 \geq n - 1/4 \geq k$. Thus we may form G by deleting any k edges from F , so G has $m = \binom{n}{2} - k$ edges. We consider separately the cases when n is even, odd.

Suppose first that n is even, so $k = n/2 + 1$ and $m = \frac{1}{2}(n^2 - 2n - 2)$. Since $|A| = |B|$ and all removed edges are between A and B , we have $\text{vol}(A) = \text{vol}(B)$ and thus $q_{\mathcal{A}}^D(G) = 1/2$. Also

$$q_{\mathcal{A}}^E(G) = \frac{2\binom{n/2}{2}}{m} = \frac{n(n-2)}{2(n^2 - 2n - 2)} = \frac{1}{2} + \frac{1}{n^2 - 2n - 2}.$$

Hence $q_{\mathcal{A}}(G) > 0$, as required.

Now suppose that n is odd, so $k = (n+1)/2$ and $m = \frac{1}{2}(n^2 - 2n - 1) > 1$. Then

$$q_{\mathcal{A}}^E(G) = \frac{\binom{\lceil n/2 \rceil}{2} + \binom{\lfloor n/2 \rfloor}{2}}{m} = \frac{(n+1)(n-1) + (n-1)(n-3)}{8m},$$

and we may check that the numerator is $4m + 4$, so

$$q_{\mathcal{A}}^E(G) = \frac{1}{2} + \frac{1}{2m}.$$

Also $\text{vol}(A) = \frac{1}{2}(n+1)(n-1) - k = \frac{1}{2}(n^2 - n - 2)$ and $\text{vol}(B) = \frac{1}{2}(n-1)^2 - k = \frac{1}{2}(n^2 - 3n)$, so

$$\begin{aligned} q_{\mathcal{A}}^D(G) &= \frac{1}{4} \frac{(n^2 - n - 2)^2 + (n^2 - 3n)^2}{4m^2} \\ &= \frac{1}{4} \frac{8m^2 + 4m + 4}{4m^2} = \frac{1}{2} + \frac{m+1}{4m^2}. \end{aligned}$$

Hence $q_A(G) = (m-1)/4m^2 > 0$, and this completes the proof of the claim. $\square$

Proof of Theorem 1. The upper bound in the theorem follows by Claim 10, so it remains to prove the lower bound. We must show that if $n \geq 4$ and we form G by removing $x \leq n/2$ edges from K_n then $q^*(G) = 0$. Let us note first one useful inequality. Let G be a graph on vertex set V , let $U \subseteq V$, and form G' by ‘moving’ an edge currently between U and $\bar{U}$ to the location of a missing edge within U or $\bar{U}$. Then by the expression (1.2) for $p_G(U)$,

$$p_{G'}(U) > p_G(U). \quad (2.1)$$

(This is similar to the ‘local rewiring’ result of [37, Lemma 1].)

Now suppose that the graph H on $V = [n]$ and $\emptyset \neq U \subset V$ are such that $p_H(U)$ maximises the value $p_G(W)$ over all graphs G on V with at most $n/2$ edges missing and all $\emptyset \neq W \subset V$. Since the number of edges missing is at most $n/2 < n-1$ there are edges of H in $(U, \bar{U})$. Hence by inequality (2.1), no edges within U or within $\bar{U}$ are missing.

Suppose that $|U| = j$ (where $1 \leq j \leq n-1$) and x edges are missing from H . Then

$$\begin{aligned} p_H(U) &= \text{vol}_H(U)\text{vol}_H(\bar{U}) - e_H(U, \bar{U})\text{vol}(H) \\ &= (j(n-1) - x)((n-j)(n-1) - x) - (j(n-j) - x)(n(n-1) - 2x) \\ &= -j(n-j)(n-1) + 2xj(n-j) - x^2 \\ &= j(n-j)(2x - (n-1)) - x^2. \end{aligned}$$

Hence $p_H(U) < 0$ if $x \leq (n-1)/2$. Suppose that $x = n/2$. Then $p_H(U) = j(n-j) - (n/2)^2$, so $p_H(U) < 0$ if $j \neq n/2$, and $p_H(U) = 0$ if $j = n/2$. This completes the proof of Theorem 1. $\square$

2.2 Extending Theorem 1, proof of Proposition 2

A proof of Proposition 2 follows quickly from the following simple lemma and the proof of Theorem 1.

Lemma 11. *Let H be a bipartite graph on $[n]$ with $k \leq n/2 + 1$ edges. Then there is a bipartition $\mathcal{A} = (A, B)$ of $[n]$ such that all edges in H are between the parts, and where $|A| = \lceil n/2 \rceil$ and $|B| = \lfloor n/2 \rfloor$; except possibly in the case where n is even and $k = n/2 + 1$, when we may still insist that $n/2 - 1 \leq |A|, |B| \leq n/2 + 1$.*

It is possible to be more precise in the case when n is even and $k = n/2 + 1$, see Proposition 12.

Proof of Lemma 11. Let $\mathcal{A} = (A, B)$ be a bipartition of $[n]$ with $E(H) \subseteq E(A, B)$, with $|A| \geq |B|$, and with $|A|$ as small as possible. Thus $|A| \geq \lceil n/2 \rceil$. If $|A| = \lceil n/2 \rceil$ then of course $|B| = \lfloor n/2 \rfloor$.

Now suppose that $|A| > \lceil n/2 \rceil$. Then A contains no isolated vertex v (since we could move v to B , and make $|A|$ one smaller), so $|A| \leq k$. Now n must be even and $k = n/2 + 1$, and $n/2 - 1 \leq |A|, |B| \leq n/2 + 1$, as required. $\square$

Proof of Proposition 2. Let the graph G on $[n]$ be such that $e(\overline{G}) = \lfloor n/2 \rfloor + 1 = f^q(n)$. In part (a) of the proof, we show that if $\overline{G}$ is bipartite, then $q^*(G) > 0$. In part (b) of the proof, we show that if $q^*(G) > 0$, then $\overline{G}$ is bipartite.

(a) Suppose that $\overline{G}$ is bipartite. We consider the two cases in the conclusion of Lemma 11 separately. For the case when $|A| = \lceil n/2 \rceil$ and $|B| = \lfloor n/2 \rfloor$, $q^*(G) > 0$ follows directly from Claim 10. So it remains to consider the case when n is even, $n \geq 6$, $k = \frac{1}{2}n + 1$, $|A| = \frac{1}{2}n + 1$, $|B| = \frac{1}{2}n - 1$ and all edges in H are between A and B .

We must show that $q_A(G) > 0$. As before, $m = \binom{n}{2} - k = \binom{n}{2} - \frac{1}{2}n - 1 = \frac{1}{2}(n^2 - 2n - 2)$. We may check that

$$\binom{\frac{1}{2}n + 1}{2} + \binom{\frac{1}{2}n - 1}{2} = \frac{1}{4}(n^2 - 2n + 4),$$

and thus

$$q_A^E(G) = \frac{\frac{1}{4}(n^2 - 2n + 4)}{\frac{1}{2}(n^2 - 2n - 2)} = \frac{1}{2} + \frac{3}{n^2 - 2n - 2}.$$

Also

$$\begin{aligned} q_A^D(G) &= \frac{\frac{1}{4}((n^2 - 4n)^2 + (n^2 - 4)^2)}{(n^2 - 2n - 2)^2} \\ &= \frac{\frac{1}{2} \frac{n^4 - 4n^3 + 4n^2 + 8}{n^4 - 4n^3 + 4n + 4}}{n^2 - 2n - 2} \\ &= \frac{1}{2} + \frac{2n^2 - 2n + 2}{n^4 - 4n^3 + 4n + 4}. \end{aligned}$$

Now

$$q_A^E(G) - \frac{1}{2} = \frac{3}{n^2 - 2n - 2} = \frac{3n^2 - 6n - 6}{(n^2 - 2n - 2)^2},$$

so

$$q_A^E(G) - q_A^D(G) = \frac{n^2 - 4n - 8}{(n^2 - 2n - 2)^2} = \frac{(n - 2)^2 - 12}{(n^2 - 2n - 2)^2}$$

which is > 0 for all $n \geq 6$. Thus $q_A(G) > 0$, as required.

(b) Now we show that if $q^*(G) > 0$, then $\overline{G}$ is bipartite. Suppose that $\overline{G}$ is not bipartite: we shall show that $q^*(G) = 0$, and thus complete the proof.

Let $\mathcal{A} = (A, B)$ be any bipartition of $V(G)$. Then $e_{\mathcal{A}}^{\text{int}}(G) \leq \binom{|A|}{2} + \binom{|B|}{2} - 1$. Suppose first that n is even. Then $e_{\mathcal{A}}^{\text{int}}(G) \leq 2\binom{n/2}{2} - 1 = \frac{1}{4}(n^2 - 2n - 4)$; and $e(G) = \binom{n}{2} - \frac{1}{2}n - 1 = \frac{1}{2}(n^2 - 2n - 2)$. Hence

$$q_{\mathcal{A}}^E(G) \leq \frac{\frac{1}{2} \frac{n^2 - 2n - 4}{n^2 - 2n - 2}}{n^2 - 2n - 2} < \frac{1}{2}.$$

Now suppose that n is odd. Then

$$\begin{aligned} e_{\mathcal{A}}^{\text{int}}(G) &\leq \binom{(n-1)/2}{2} + \binom{(n+1)/2}{2} - 1 \\ &= \frac{1}{8}((n-1)(n-3) + (n^2 - 1)) - 1 \\ &= \frac{1}{4}(n^2 - 2n - 3), \end{aligned}$$

and $e(G) = \binom{n}{2} - \frac{1}{2}n - \frac{1}{2} = \frac{1}{2}(n^2 - 2n - 1)$. Hence

$$q_A^E(G) \leq \frac{\frac{1}{2}n^2 - 2n - 3}{n^2 - 2n - 1} < \frac{1}{2}.$$

Hence, in both the even and the odd case, $q_A^E(G) < \frac{1}{2}$, and so $q_A(G) < 0$ (since $q_B^D(G) \geq 1/2$ for any bipartition $\mathcal{B}$). It now follows by Lemma 8 that $q^*(G) = 0$, as required. $\square$

Let us say that a bipartition $\{A, B\}$ is *balanced* if $||A| - |B|| \leq 1$. Lemma 11 establishes that, if $k \leq \lfloor n/2 \rfloor + 1$, then any k -edge bipartite graph H on $[n]$ has a balanced bipartition with all edges between parts, except possibly for the case when n is even and $k = n/2 + 1$. Below, we further consider this last case, and show the following.

Proposition 12. *Let $n \geq 4$ be even, let $k = n/2 + 1$, and let H be a k -edge bipartite graph on $[n]$. Then there is no balanced bipartition of $[n]$ such that the edges of H lie between the parts if and only if H less any isolated vertices and edges is a star with at least 3 edges.*

Proof. Suppose first that H less any isolated vertices and edges is a star with at least 3 edges. Let (A, B) be any bipartition of $[n]$ with $E(H) \subseteq E(A, B)$, and suppose wlog that the centre of the star is in B . Then each vertex in A has degree at most 1, so $|A| \geq k$, and the bipartition is not balanced.

Now suppose that there is no balanced bipartition as required. Let (A, B) be a bipartition of $[n]$ such that the edges of H lie between the parts, with $|A| > n/2$, and with $|A|$ as small as possible. Then there are no isolated vertices v in A (since we could move v to A). Thus, since $|A| = k = |E(H)|$, each vertex in A has degree 1. Hence, the graph H is a disjoint union of proper-stars, edges and isolated vertices; where there is at least one proper-star, and the centres of the stars and the isolated vertices are in B . (Here, we say ‘proper-star’ if the star has at least 2 edges.)

It remains to show that there is at most one proper-star, and then that this proper-star has ≥ 3 edges. Suppose for a contradiction that there are at least 2 proper stars, pick a smallest one S , and swap sides for its vertices, to obtain the amended bipartition (A', B') . Then $|A'| \leq |A| - 1 = n/2$, and $|A'|$ is at least the number of non-isolated vertices in B' . Hence, we can redistribute the isolated vertices of H to obtain a balanced bipartition for H . This contradiction shows there is at most one proper-star. Finally, simply note that if there is one proper-star and it has 2 edges, then there is balanced bipartition, which completes the proof. $\square$

Remark 13. Proposition 2 also implies the following. Let $0 \leq k \leq \lfloor n/2 \rfloor + 1$, and let the graph G be obtained from K_n by deleting the edges of a k -edge bipartite graph H on $[n]$. Then $\delta^-(G) = \lfloor n/2 \rfloor + 1 - k$. We may see this since if $k = \lfloor n/2 \rfloor + 1$ then $q^*(G) > 0$; and if we delete $k \leq \lfloor n/2 \rfloor$ edges of a bipartite graph between A and B with $||A| - |B|| \leq 1$, we may delete a further $\lfloor n/2 \rfloor + 1 - k$ edges from between A and B and apply Claim 10.

2.3 Local graph modifications

We shall use the following lemma in the proof of part (a) of Theorem 3. Note that in the lemma below an edge between the parts is removed or an edge within a part added, whilst in the inequality (2.1) above an edge from between the parts is moved to within the parts.

Lemma 14. *Let G be a graph and let $\mathcal{A} = \{A, \overline{A}\}$ be a bipartition with $q_{\mathcal{A}}(G) \geq 0$.*

(i) *If $m \geq 2$ and G^- is obtained by removing an edge between the parts, then $q_{\mathcal{A}}(G^-) > q_{\mathcal{A}}(G)$.*

(ii) *If $e_G(A) > 0$ and G^+ is obtained by adding an edge within $\overline{A}$, then $q_{\mathcal{A}}(G^+) > 0$.*

Proof. The proof is almost immediate from (1.2), which shows that

$$q_{\mathcal{A}}(G) = \frac{1}{2m^2}(4e_G(A)e_G(\overline{A}) - e_G(A, \overline{A})^2). \quad (2.2)$$

To prove part (i), note that in G^- the quantities $e(A)$ and $e(\overline{A})$ are unchanged from G , but $e(A, \overline{A})$ has decreased by one. Thus by (2.2) it follows that $q_{\mathcal{A}}(G^-) > q_{\mathcal{A}}(G) (\geq 0)$, as claimed.

Part (ii) follows similarly. In this case, $e(A, \overline{A})$ and $e(A) > 0$ are the same in G and G^+ , but $e(\overline{A})$ has increased by one, so by (2.2) we may conclude $q_{\mathcal{A}}(G^+) > 0$. $\square$

Given an n -vertex graph G , by adding $((\binom{n}{2} - \frac{n}{2} - e(G))^+)$ edges we may form a graph G_0 with at most $\frac{n}{2}$ edges missing, and then $q^*(G_0) = 0$ since $f^q(n) > n/2$ by Theorem 1. Hence, by Lemma 9 (the robustness lemma), we have the following result, which extends the bound from Theorem 1 that $f^q(n) > n/2$. The notation x^+ means $\max\{x, 0\}$.

Lemma 15. *For each $n \geq 2$ and each nonempty n -vertex graph G*

$$q^*(G) \leq \frac{2(\binom{n}{2} - \frac{n}{2} - e(G))^+}{e(G)}.$$

We will use Lemma 15 to prove the upper bound in Theorem 5 (b).

2.4 Nearly complete bipartite graphs: proof of Theorem 3

We commented earlier that any complete multipartite graph G satisfies $q^*(G) = 0$, but there is no robustness result for complete multipartite graphs corresponding to Theorem 1. In this section we prove the three parts of Theorem 3 on $\delta^-(G)$, $\delta^+(G)$ and $\delta(G)$ for complete bipartite graphs G , in Lemmas 16, 17 and 18. In Section 2.5 we briefly consider complete multipartite graphs.

We start with δ^- , and see in Lemma 16 that $\delta^-(G) = 1$ for all complete bipartite graphs G , with the exception of stars.

Lemma 16 (part (a) of Theorem 3). *Let $G = K_{s,t}$ be a complete bipartite graph with $s \leq t$. If $s = 1$ then $\delta^-(G) = \infty$; and if $s \geq 2$ then $\delta^-(G) = 1$.*

Proof. Suppose first that $s = 1$, so G is the star $K_{1,t}$. As we delete edges we get stars $K_{1,t'}$ for $t' < t$ (together with isolated vertices) and eventually an empty graph. All these graphs have modularity zero and hence $\delta^-(K_{1,t}) = \infty$.

From now on let $s \geq 2$. Note that $q^*(K_{s,t}) = 0$ (shown in [6, 23], see also Section 3). Thus we must show that removing an edge from $K_{s,t}$ yields a graph with positive modularity value. (Recall by Lemma 8 that $q^*(G) > 0$ if and only if there is a bipartition $\mathcal{A}$ with $q_{\mathcal{A}}(G) > 0$, and hence $q^*(K_{s,t}) = 0$ follows also from the claim on line (2.3) below.)

There are two cases. We first consider the case that s and t have a common factor and show that in this case there is a bipartition $\mathcal{A} = \{A, \bar{A}\}$ with $q_{\mathcal{A}}(K_{s,t}) = 0$ and $e(A, \bar{A}) \geq 1$. Thus by Lemma 14 part (i), for $K_{s,t}^-$ the graph obtained from $K_{s,t}$ by removing an edge between parts A and $\bar{A}$ we have $q_{\mathcal{A}}(K_{s,t}^-) > 0$. After that we consider the case when s and t are coprime which is only a little more complicated.

Before splitting into cases we give an expression for the modularity score of $K_{s,t}$ for bipartitions which will be useful for both cases. Write S, T for the two parts of the bipartite graph of sizes s, t respectively, and let $\sigma = |A \cap S|/s$ and $\tau = |A \cap T|/t$ be the proportions of the vertex set A in each of the bipartite parts. We *claim* that for $\mathcal{A} = \{A, \bar{A}\}$

$$q_{\mathcal{A}}(K_{s,t}) = -\frac{1}{2}(\sigma - \tau)^2. \quad (2.3)$$

Proof of Claim. The proof is by direct calculation. Note that

$$e(A) = |A \cap S| |A \cap T| = \sigma \tau st,$$

similarly, $e(\bar{A}) = (1 - \sigma)(1 - \tau)st$, and thus

$$q_{\mathcal{A}}^E(K_{s,t}) = \sigma \tau + (1 - \sigma)(1 - \tau) = 1 - (\sigma + \tau) + 2\sigma \tau. \quad (2.4)$$

Also each vertex in part S (resp. T) has degree t (resp. s) and so

$$\text{vol}(A) = |A \cap S|t + |A \cap T|s = (\sigma + \tau)st,$$

which together with the corresponding expression for $\text{vol}(\bar{A})$ yields

$$q_{\mathcal{A}}^D(K_{s,t}) = \left(\frac{1}{2}(\sigma + \tau)\right)^2 + \left(1 - \frac{1}{2}(\sigma + \tau)\right)^2 = 1 - (\sigma + \tau) + \frac{1}{2}(\sigma + \tau)^2.$$

Thus we have

$$q_{\mathcal{A}}(K_{s,t}) = 2\sigma \tau - \frac{1}{2}(\sigma + \tau)^2 = -\frac{1}{2}(\sigma - \tau)^2$$

which completes the proof of the claim.

Case 1: s and t have a common factor.

Recall that by Lemma 14 part (i), it is sufficient to find a bipartition $\mathcal{A} = \{A, \bar{A}\}$ such that $e(A, \bar{A}) \geq 1$ and $q_{\mathcal{A}}(K_{s,t}) = 0$. Let $\ell > 1$ be a common factor of s and t . Then take A to have $|S|/\ell$ vertices from S and $|T|/\ell$ vertices from T . Clearly $e(A, \bar{A}) \geq 1$; and since $\sigma = |A \cap S|/s = 1/\ell = |A \cap T|/t = \tau$, by the claim on line (2.3) we have $q_{\mathcal{A}}(K_{s,t}) = 0$ and we are done.

Case 2: s and t are co-prime.

The following relation between graphs G , G^- and the modularities will be useful. Let $\mathcal{A} = \{A, \bar{A}\}$ be a vertex bipartition of G such that $e(A, \bar{A}) \geq 1$ and let G^- be any graph formed from G by removing an edge between the parts A and $\bar{A}$. Then

$$e(G^-)^2 q_{\mathcal{A}}(G^-) = e(G)^2 q_{\mathcal{A}}(G) - e(G) q_{\mathcal{A}}^E(G) + e(G) - 1/2. \quad (2.5)$$

To see this note that because the removed edge is between parts in $\mathcal{A}$, $e(G^-) q_{\mathcal{A}}^E(G^-) = e(G) q_{\mathcal{A}}^E(G)$. Similarly one may check that $e(G^-)^2 q_{\mathcal{A}}^D(G^-) = e(G)^2 q_{\mathcal{A}}^D(G) - e(G) + 1/2$ and that these together give (2.5).

Applied to our graph $G = K_{s,t}$ and $G^- = K_{s,t}^-$ obtained by removing an edge from between parts A and $\bar{A}$, we have

$$(st - 1)^2 q_{\mathcal{A}}(K_{s,t}^-) = s^2 t^2 \cdot q_{\mathcal{A}}(K_{s,t}) - st \cdot (q_{\mathcal{A}}^E(K_{s,t}) - 1) - 1/2. \quad (2.6)$$

Note from (2.4)

$$q_{\mathcal{A}}^E(K_{s,t}) - 1 = -(\sigma + \tau) + 2\sigma\tau = -(\sigma - \tau)^2 - \sigma(1 - \sigma) - \tau(1 - \tau).$$

Hence, substituting this and the expression for the modularity score in (2.3) we have

$$(st - 1)^2 q_{\mathcal{A}}(K_{s,t}^-) = -st(\sigma - \tau)^2 (\frac{1}{2}st - 1) + st(\sigma(1 - \sigma) + \tau(1 - \tau)) - 1/2. \quad (2.7)$$

Observe that since s and t are co-prime the linear diophantine equation $at - bs = 1$ has solutions (a, b) with a and b integers. Note also that if (a, b) is a solution then $(a + s, b + t)$ is also a solution. Hence there is a solution (a', b') with $1 \leq a' \leq s - 1$ (we cannot have $a' = 0$). Thus $b's = a't - 1$ satisfies $t - 1 \leq b's \leq (s - 1)t - 1$, and so $1 \leq b' \leq t - 1$.

Our construction is to take A with a' vertices in S and b' vertices in T then

$$|\sigma - \tau| = \left| \frac{|A \cap S|}{s} - \frac{|A \cap T|}{t} \right| = \left| \frac{a'}{s} - \frac{b'}{t} \right| = \frac{1}{st}. \quad (2.8)$$

Since $1 \leq a' \leq s - 1$, σ satisfies $1/s \leq \sigma \leq 1 - 1/s$ and hence $\sigma(1 - \sigma) \geq 1/s^2$. Similarly $\tau(1 - \tau) \geq 1/t^2$.

Hence by (2.7), substituting $|\sigma - \tau| = 1/(st)$, $\sigma(1 - \sigma) \geq 1/s^2$ and $\tau(1 - \tau) \geq 1/t^2$ we get

$$(st - 1)^2 q_{\mathcal{A}}(K_{s,t}^-) \geq \frac{1}{st} + \frac{t}{s} + \frac{s}{t} - 1 > 0 \quad (2.9)$$

and we are done. $\square$

Lemma 17 (part (b) of Theorem 3). *Let G be a complete bipartite graph $K_{s,t}$ with $s \leq t$. If $s = 1$ and $t \geq 4$, or if $s \geq 2$ and $t \geq 3$, then $\delta^+(G) = 1$. In the other cases (namely when $(s, t) = (1, 1), (1, 2), (1, 3)$ or $(2, 2)$), we have $\delta^+(G) = \infty$.*

Proof. Let $s = 1$ and form G^+ by adding an edge to the star $K_{1,t}$. If t is 1 or 2 then $q^*(G^+) = 0$ since G^+ has at most 3 vertices. If $t = 3$ then G^+ is K_4 less 2 edges, so $q^*(G^+) = 0$ by Theorem 1. But for all $t \geq 4$

$$q^*(G^+) > 0. \quad (2.10)$$

To see this, suppose that the added edge e is $\{a, b\}$, and let $\mathcal{A}$ be the partition of the vertex set V into $\{a, b\}$ and $V \setminus \{a, b\}$. Then

$$q_{\mathcal{A}}^E(G^+) = \frac{t-1}{t+1}$$

and

$$q_{\mathcal{A}}^D(G^+) = \frac{4^2 + (2t-2)^2}{4(t+1)^2} = \frac{4 + (t-1)^2}{(t+1)^2}.$$

Thus

$$(t+1)^2 q_{\mathcal{A}}(G^+) = (t^2 - 1) - 4 - (t^2 - 2t + 1) = 2t - 6 :$$

so $q_{\mathcal{A}}(G^+) > 0$ for $t \geq 4$. This completes the proof of (2.10), and thus of the case $s = 1$.

Suppose now that $s \geq 2$. Note first that if $s = t = 2$ (so G is $K_{2,2}$) then at most 2 edges are missing from K_4 , and thus $\delta^+(G) = \infty$ by Theorem 1. Now assume also that $t \geq 3$. It remains to show that $\delta^+(G) = 1$.

The following relation between graphs G , G^+ and the modularity scores will be useful. Let $\mathcal{A} = \{A, \bar{A}\}$ be a vertex bipartition of G such that $e(A) < |A|(|A| - 1)/2$ and let G^+ be any graph formed from G by adding an edge within the part A . Since the extra edge is added inside a part in $\mathcal{A}$, $e(G^+)q_{\mathcal{A}}^E(G^+) = e(G)q_{\mathcal{A}}^E(G) + 1$. For the degree tax we may calculate

$$e(G^+)^2 q_{\mathcal{A}}^D(G^+) = \frac{1}{4}(\text{vol}_G(A) + 2)^2 + \frac{1}{4}\text{vol}_G(\bar{A})^2 = e(G)^2 q_{\mathcal{A}}^D(G) + \text{vol}_G(A) + 1.$$

Thus we get an expression similar to (2.5),

$$\begin{aligned} e(G^+)^2 q_{\mathcal{A}}(G^+) &= e(G^+)^2 q_{\mathcal{A}}^E(G^+) - e(G^+)^2 q_{\mathcal{A}}^D(G^+) \\ &= (e(G) + 1)(e(G)q_{\mathcal{A}}^E(G) + 1) - e(G)^2 q_{\mathcal{A}}^D(G) - \text{vol}_G(A) - 1 \\ &= e(G)^2 q_{\mathcal{A}}(G) + e(G) + e(G)q_{\mathcal{A}}^E(G) - \text{vol}_G(A). \end{aligned} \quad (2.11)$$

Note that making the following substitutions for total edges : $e(G) = e_G(A) + e_G(A, \bar{A}) + e_G(\bar{A})$, for internal edges : $e(G)q_{\mathcal{A}}^E(G) = e_G(A) + e_G(\bar{A})$ and for volume of A : $\text{vol}_G(A) = 2e_G(A) + e_G(A, \bar{A})$ we get

$$e(G^+)^2 q_{\mathcal{A}}(G^+) = e(G)^2 q_{\mathcal{A}}(G) + 2e_G(\bar{A}). \quad (2.12)$$

Case 1: s and t have a common factor.

Let $\ell > 1$ be a common factor of s and t . Take $\overline{A}$ to be the part with $|S|/\ell$ vertices from S and $|T|/\ell$ vertices from T (we want to add an edge within the vertex set A , so we take $|A| \geq |\overline{A}|$). Then $|A \cap T| \geq 2$ since $t \geq 3$, and thus we may add an edge within part A - call this new graph $K_{s,t}^+$.

Similarly to in the proof of Lemma 16, $\sigma = |A \cap S|/s = 1 - 1/\ell = |A \cap T|/t = \tau$ and we have $q_{\mathcal{A}}(K_{s,t}) = 0$ for the bipartition $\mathcal{A} = \{A, \overline{A}\}$. Thus by (2.12)

$$e(K_{s,t}^+)^2 q_{\mathcal{A}}(K_{s,t}^+) = 2e_{K_{s,t}}(\overline{A}).$$

But now, since $\overline{A}$ has non-empty intersection with both S and with T we have $e_{K_{s,t}}(\overline{A}) \geq 1$ and thus $q_{\mathcal{A}}(K_{s,t}^+) > 0$ as required.

Case 2: s and t are co-prime.

As in the proof of Lemma 16 we may assume that $a't - b's = 1$ has a solution with $1 \leq a' \leq s - 1$ and $1 \leq b' \leq t - 1$ positive integers. Now, if either $a' > 1$ or $b' > 1$ then take A with a' vertices in S and b' vertices in T , otherwise take A to be the complement of that, i.e., with $s - a'$ vertices in S and $t - b'$ vertices in T . Let $\sigma = |A \cap S|/s$ and $\tau = |A \cap T|/t$. In either case, as in (2.8),

$$|\sigma - \tau| = 1/(st). \quad (2.13)$$

Notice also by our choice of A that the set A intersects either S or T in at least two vertices, and hence we have a place to add an edge within the set A , again we call the resulting graph $K_{s,t}^+$.

Now by (2.12) for bipartition $\mathcal{A} = \{A, \overline{A}\}$

$$e(K_{s,t}^+)^2 q_{\mathcal{A}}(K_{s,t}^+) = e(K_{s,t})^2 q_{\mathcal{A}}(K_{s,t}) + e_{K_{s,t}}(\overline{A}) = -\frac{1}{2} + e_{K_{s,t}}(\overline{A})$$

where the second equality follows because by the claim in (2.3) $q_{\mathcal{A}}(K_{s,t}) = -\frac{1}{2}(\sigma - \tau)^2$, by (2.13) and because the total number of edges is $e(K_{s,t}) = st$. Again, by construction $\overline{A}$ has non-empty intersection both with S and with T and hence $e_{K_{s,t}}(\overline{A}) \geq 1$, and $q_{\mathcal{A}}(K_{s,t}^+) > 0$ as required. $\square$

The above results show that for a complete bipartite graph G , always $\delta^-(G)$ and $\delta^+(G)$ are 1 or ∞ : that is not quite the case for $\delta(G)$.

Lemma 18 (part (c) of Theorem 3). *Let G be a complete bipartite graph $K_{s,t}$ with $s \leq t$. If $s = 1$ and $t \geq 4$, or if $s \geq 2$, then $\delta(G) = 1$. If $(s, t) = (1, 1)$ or $(1, 2)$ then $\delta(G) = \infty$, and in the remaining case $(s, t) = (1, 3)$ we have $\delta(G) = 2$.*

Proof. If $s = 1$ and $t \geq 4$, or if $s \geq 2$, then $1 \leq \delta(G) \leq \delta^+(G) = 1$ by Lemma 17; and if $s + t \leq 3$ then $\delta(G) = \infty$. This leaves only the case $(s, t) = (1, 3)$. When $G = K_{1,3}$ we saw that $\delta^-(G) = \delta^+(G) = \infty$, and so $\delta(G) \geq 2$. But if H is the 4-vertex path then $q^*(H) = \frac{1}{6} > 0$, and so $\delta(G) = 2$. $\square$

2.5 Nearly complete multipartite graphs - a partial story

We start this section by showing that complete multipartite graphs G whose part sizes have a non-trivial common divisor have $\delta^-(G) = \delta^+(G) = \delta(G) = 1$; though some other complete multipartite graphs G have $\delta^+(G) = \infty$, see Example 20. Finally we give an example of a complete tripartite graph G with $\delta^+(G) = 2$, see Example 21. We do not aim here to give a complete story but rather give a few partial results and examples which indicate that for complete multipartite graphs the results will be a little more involved than for complete bipartite graphs.

Proposition 19. *Let $G = K_{s_1, \dots, s_k}$ be a complete multipartite graph where $k \geq 2$; and let $\ell | s_i$ for each i for some integer $\ell \geq 2$. Then $\delta^-(G) = \delta^+(G) = \delta(G) = 1$, except that $\delta^+(G) = \infty$ if each $s_i = 2$.*

Proof. Denote the parts of the complete multipartite graph by $S_1, \dots, S_k$. Construct the vertex set A by including $|S_i|/\ell$ vertices from S_i for each i . We prove first that for the bipartition $\mathcal{A} = \{A, \bar{A}\}$ we have $q_{\mathcal{A}}(G) = 0$. Note that

$$e_G(A) = \sum_{i < j} \frac{|S_i|}{\ell} \frac{|S_j|}{\ell} = \frac{1}{\ell^2} e(G);$$

that is, part A contains a $1/\ell^2$ proportion of the edges of G . Similarly, part $\bar{A}$ contains a $(\ell - 1)^2/\ell^2$ proportion of the edges; and thus $q_{\mathcal{A}}^E(G) = (1 + (\ell - 1)^2)/\ell^2$. Also, part A has a $1/\ell$ proportion of the volume of the graph, and part $\bar{A}$ has a $(\ell - 1)/\ell$ proportion; and so $q_{\mathcal{A}}^D(G) = q_{\mathcal{A}}^E(G)$. Hence we have $q_{\mathcal{A}}(G) = 0$, as required.

Recall that $q^*(G) = 0$, by [6, 37]. Since $e_G(A, \bar{A}) \geq 1$, we can form G^- by removing an edge between A and $\bar{A}$; and by Lemma 14 part (i), $q^*(G^-) > 0$ and so $\delta^-(G) = 1$. It follows that $\delta(G) = 1$.

Now consider $\delta^+(G)$. Observe that $e_G(A) \geq 1$. Suppose first that some $s_i \neq 2$: then $|\bar{A} \cap S_i| \geq s_i/2 > 1$, so we can form G^+ by adding an edge within $\bar{A}$. Now by Lemma 14 part (ii), $q^*(G^+) > 0$ and so $\delta^+(G) = 1$. On the other hand, if each $s_i = 2$ and $n = \sum_i s_i = 2k$ then G is K_n less $n/2$ edges, so $\delta^+(G) = \infty$ by Theorem 1. $\square$

Example 20. Let $j, k \geq 0$ with $j + k \geq 1$, and let $G = G_{j,k}$ be the complete multipartite graph $K_{2, \dots, 2, 1, \dots, 1}$ with j partite sets of size 2 and k of size 1 (and so with $n = 2j + k$ vertices), and note that $q^*(G) = 0$. We shall show that $\delta^+(G) = \infty$, and $\delta(G) = \delta^-(G) = \lfloor k/2 \rfloor + 1$ if $n \geq 4$.

Consider $\delta^+(G)$. If $j = 0$ it is not possible to add an edge. If $j \geq 1$ we may add $\ell = j$ edges to yield the complete graph K_{2j+k} . If $j \geq 2$ we may also add $\ell < j$ edges which yields the complete multipartite graph $G_{j-\ell, k+2\ell}$. However, all such graphs have modularity zero, and so $\delta^+(G) = \infty$ as claimed.

Now let $n = 2j + k \geq 4$ and consider $\delta^-(G)$. Note that G is the complete graph K_n less the bipartite graph formed by j disjoint edges, and $j \leq n/2$. Hence by Proposition 2 we have $\delta^-(G) = \lfloor n/2 \rfloor + 1 - j = \lfloor k/2 \rfloor + 1$ if $n \geq 5$, and the case $n = 4$ is easily checked.

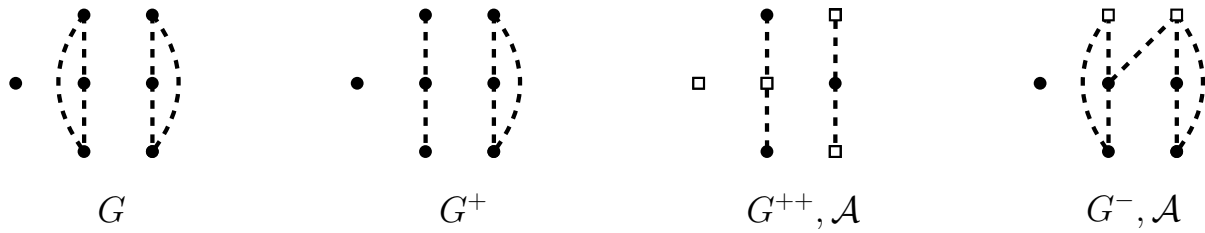


Figure 2: An illustration of the graphs considered in Example 21, the dotted lines indicate the non-edges in the graph. For G^{++} and G^- the vertices are depicted with dots and squares to illustrate a bipartition which achieve positive modularity.

Example 21. Let G be the complete tripartite graph $K_{1,3,3}$. We shall see that $\delta^+(G) = 2$ and $\delta(G) = \delta^-(G) = 1$.

Note that $q^*(G) = 0$ and thus $\delta^+(G), \delta^-(G) \geq 1$.

We first show that $\delta^+(G) \geq 2$ and then that $\delta^+(G) = 2$. All graphs obtained by adding a single edge to G are isomorphic: denote this 16-edge graph by G^+ . Observe that G^+ is the complete graph K_7 less a disjoint triangle and two-edge path: a total of five missing edges. By Lemma 8, to show that $q^*(G^+) = 0$ it is enough to consider an arbitrary bipartition $\mathcal{A} = \{A, \bar{A}\}$ for G^+ , and show that $q_{\mathcal{A}}(G^+) \leq 0$. Without loss of generality, $|A| \leq |\bar{A}|$. Note that for every graph H and every bipartition $\mathcal{B}$ with a part of size 1, we have $q_{\mathcal{B}}(H) \leq 0$. Thus we may assume that $|A| \geq 2$, so $|A|$ is 2 or 3.

Of the five edges missing in G^+ at most four can be between the parts A and $\bar{A}$, and thus $e_{G^+}(A, \bar{A}) \geq |A||\bar{A}| - 4$. Firstly, if $|A| = 3$ and $|\bar{A}| = 4$ then $e_{G^+}(A, \bar{A}) \geq |A||\bar{A}| - 4 = 8$. Hence $q_{\mathcal{A}}^E(G^+) \leq 8/16 = 1/2$, and since the degree tax of any bipartition is always at least $1/2$, in this case $q_{\mathcal{A}}(G^+) \leq 0$.

To show that $\delta^+(G) \geq 2$, it remains to check the case $|A| = 2$ and $|\bar{A}| = 5$, when $e_{G^+}(A, \bar{A}) \geq 6$. If $e_{G^+}(A, \bar{A}) \geq 8$ then as in the case $|A| = 3$ above we have $q_{\mathcal{A}}(G^+) \leq 0$. If $e_{G^+}(A, \bar{A}) = 7$ then $\text{vol}_{G^+}(A) = 2e_{G^+}(A) + e_{G^+}(A, \bar{A}) \leq 2 + 7 = 9$. The degree tax is minimised by having volumes as equal as possible, so $q_{\mathcal{A}}^D(G^+) \geq (9^2 + 23^2)/32^2$. But $q_{\mathcal{A}}^E(G^+) = 1 - 7/16 = 9/16$ so $q_{\mathcal{A}}(G^+) \leq 9/16 - 305/512 < 0$. Finally, suppose that $e_{G^+}(A, \bar{A}) = 6$. Then arguing as before we see that $\text{vol}_{G^+}(A) \leq 2 + 6 = 8$, so $q_{\mathcal{A}}^D(G^+) \geq (8^2 + 24^2)/32^2 = 5/8$. But $q_{\mathcal{A}}^E(G^+) = 1 - 6/16 = 5/8$, and so $q_{\mathcal{A}}(G^+) = 0$.

We have now seen that $\delta^+(G) \geq 2$. To show that $\delta^+(G) = 2$, let G^{++} be the 17 edge graph obtained from $K_{1,3,3}$ by adding two disjoint edges. Note that G^{++} is the complete graph less two disjoint two-edge paths. Let $\mathcal{A}$ be a bipartition with part sizes 3 and 4 such that all four missing edges are between the parts, so $e_{G^{++}}(A, \bar{A}) = 8$. Then $q_{\mathcal{A}}^E(G^{++}) = 8/17$ and the degree tax is $q_{\mathcal{A}}^D(G^{++}) = (14^2 + 20^2)/34^2$ which gives $q_{\mathcal{A}}(G^{++}) > 0$ and thus $\delta^+(G) = 2$.

Finally, we show that $\delta^-(G) = 1$ and thus $\delta(G) = 1$. Form G^- from G by deleting an edge, ij say, between the partite sets of size 3. Then let $A = \{i, h\}$ where $h \neq j$ is in the same partite set as j . Note that the 7 ‘missing’ edges in G^- form two K_3 ’s joined by an edge, and 5 of these ‘missing’ edges are between A and $\bar{A}$ so there are 5 edges in G^- between A and $\bar{A}$. Thus for the bipartition $\mathcal{A} = \{A, \bar{A}\}$ we have $q_{\mathcal{A}}^E(G^-) = 1 - 5/14 =$

9/14. For the degree tax of $\mathcal{A}$, note that $d_i = 3$ and $d_h = 4$ so $\text{vol}(A) = 7$, $\text{vol}(\bar{A}) = 21$ and thus $q_{\mathcal{A}}^D(G^-) = (7^2 + 21^2)/28^2 = (1^2 + 3^2)/4^2 = 5/8$. Hence $q_{\mathcal{A}}(G^-) = 9/14 - 5/8 = 1/56 > 0$, and so we have $\delta^-(G) = 1$ as required.

3 Short proof that complete multipartite graphs have modularity zero

We noted that complete multipartite graphs have modularity zero ([6, 23]). The proofs given for this result are based on properties of the eigenvalues of related matrices. Here we give a shorter simple proof based on graph expansion.

Theorem 22. *If G is a complete multipartite graph then $q^*(G) = 0$.*

Proof. We use the following recent expansion result [20]: for every graph G

$$q^*(G) = 0 \quad \text{if and only if} \quad e(A, \bar{A}) \geq 2\sqrt{e(A)e(\bar{A})} \quad \forall A \subseteq V(G).$$

Here $\bar{A} = V(G) \setminus A$, and recall that $q^*(G) = 0$ if G has no edges.

Fix a complete multipartite graph G on parts $V_1, \dots, V_k$ for some $k \geq 2$. Let $\emptyset \neq A \subsetneq V(G)$, let $x_i = |A \cap V_i|$ and $y_i = |\bar{A} \cap V_i|$; and let $x = \sum_i x_i = |A|$ and $y = \sum_i y_i = |\bar{A}|$. Observe that $2e(A) = \sum_i x_i(x - x_i) = x^2 - \sum_i x_i^2$, similarly $2e(\bar{A}) = y^2 - \sum_i y_i^2$, and $e(A, \bar{A}) = \sum_i x_i(y - y_i) = xy - \sum_i x_i y_i$. Thus $e(A, \bar{A}) \geq 2\sqrt{e(A)e(\bar{A})}$ if and only if

$$(xy - \sum_i x_i y_i)^2 \geq (x^2 - \sum_i x_i^2)(y^2 - \sum_i y_i^2). \quad (3.1)$$

We shall prove the theorem by showing that (3.1) holds. Let $u_i = x_i/x$ so $\sum_i u_i = 1$, and let $v_i = y_i/y$ so $\sum_i v_i = 1$. Dividing both sides of (3.1) by $x^2 y^2$ shows that it is equivalent to

$$(1 - \sum_i u_i v_i)^2 \geq (1 - \sum_i u_i^2)(1 - \sum_i v_i^2). \quad (3.2)$$

Let $s = (\sum_i u_i^2)^{1/2}$ and $t = (\sum_i v_i^2)^{1/2}$, so $0 < s, t \leq 1$. By the Cauchy-Schwarz inequality we have $\sum_i u_i v_i \leq st$. Let $\alpha = (\sum_i u_i v_i)/st$, so $0 < \alpha \leq 1$. We may now write the inequality (3.2) as

$$(1 - \alpha st)^2 \geq (1 - s^2)(1 - t^2),$$

that is

$$s^2 + t^2 - 2\alpha st \geq (1 - \alpha^2)s^2 t^2.$$

The LHS here equals

$$s^2 - 2st + t^2 + 2(1 - \alpha)st = (s - t)^2 + 2(1 - \alpha)st \geq 2(1 - \alpha)st.$$

Thus to prove (3.1) it suffices to show that

$$2(1 - \alpha)st \geq (1 - \alpha^2)s^2 t^2. \quad (3.3)$$

But both sides in (3.3) are non-negative (since $0 < \alpha \leq 1$ and $st > 0$), $st \geq s^2t^2$ (since $0 < s, t \leq 1$), and $2(1 - \alpha) \geq (1 - \alpha^2)$ (since $2(1 - \alpha) - (1 - \alpha^2) = (1 - \alpha)^2 \geq 0$). Hence (3.3) holds, and the proof is complete. $\square$

4 Proofs for results on random graphs

Most of this section is taken up with proving Theorem 5. After that we quickly prove Theorem 6.

4.1 Proof of Theorem 5

Part (a). This part follows easily from Theorem 1. Let $p = p(n)$ satisfy $p \geq 1 - 1/n + \omega(n)n^{-3/2}$. The number of edges missing in $G_{n,p}$ is at most X , where $X \sim \text{Bin}(\binom{n}{2}, 1/n - \omega(n)n^{-3/2})$. Now $\mathbb{E}[X] \leq n/2 - \omega\sqrt{n}/2$, and $\text{var}(X) \leq n/2$. Thus, by Chebyshev's inequality,

$$\mathbb{P}(X \geq n/2) \leq \mathbb{P}(|X - \mathbb{E}[X]| \geq \omega\sqrt{n}/2) \leq \frac{n/2}{\omega^2 n/4} = o(1).$$

Hence $q^*(G_{n,p}) = 0$ whp by Theorem 1. This completes the proof of Theorem 5 (a).

Part (b). Let $p = p(n) = 1 - 1/n$. For subpart (i), note that the number X of edges missing in $G_{n,p}$ satisfies $X \sim \text{Bin}(\binom{n}{2}, 1/n)$, and X is asymptotically normal with mean and variance $n/2$ (see for example [7]). Thus $\mathbb{P}(X \leq n/2) = \frac{1}{2} + o(1)$, and so by Theorem 1 as above, $\mathbb{P}(q^*(G_{n,p}) = 0) \geq \frac{1}{2} + o(1)$. It is now enough to appeal to subpart (ii) to conclude that we must have $\mathbb{P}(q^*(G_{n,p}) = 0) = \frac{1}{2} + o(1)$. Hence we have finished the proof of (b) subpart (i), conditional on the proof of subpart (ii).

Now consider subpart (ii). For X as above, there is a constant $c > 0$ such that $\mathbb{P}(X > \lfloor n/2 \rfloor + c\sqrt{n}) < \varepsilon/2$. Let $\hat{G}$ be any n -vertex graph with x edges missing, where $x \leq \lfloor n/2 \rfloor + c\sqrt{n}$; and let H be a graph obtained by deleting $\min\{x, \lfloor n/2 \rfloor\}$ of the missing edges of $\hat{G}$ from K_n . Then $q^*(H) = 0$ by Theorem 1; and by Lemma 9 (the robustness lemma),

$$q^*(\hat{G}) = |q^*(H) - q^*(\hat{G})| \leq 2(X - \lfloor n/2 \rfloor)^+/e(H) \leq (1 + o(1))2c\sqrt{n}/(n^2/2) < 5cn^{-3/2}$$

for n sufficiently large. Now let $\beta = 5c$, and observe that

$$\mathbb{P}(q^*(G_{n,p}) \geq \beta n^{-3/2}) \leq \mathbb{P}(X > \lfloor n/2 \rfloor + c\sqrt{n}) < \varepsilon/2.$$

It remains to show that there is a constant $\alpha > 0$ such that $\mathbb{P}(q^*(G_{n,p}) \leq \alpha n^{-3/2}) < \varepsilon/2$. There is a constant $c_0 > 0$ such that $\mathbb{P}(X < n/2 + c_0\sqrt{n}) < \frac{1}{2} + \varepsilon/2$. Note that $\mathbb{P}(n/2 + c_0\sqrt{n} \leq X \leq n/2 + c\sqrt{n}) > \frac{1}{2} - \varepsilon$. Let $m = m(n)$ satisfy $n/2 + c_0\sqrt{n} \leq m \leq n/2 + c\sqrt{n}$. Consider the random graph $G_{n,m}$ and its complement $\overline{G}_{n,m}$. It suffices to show that $q^*(\overline{G}_{n,m}) \geq (c_0 + o(1))n^{-3/2}$ whp.

Let us recall some definitions. If the graph G has n vertices, m edges and k components, then the *cyclomatic number* (or cycle rank) $\text{cyclo}(G)$ is $m - n + k$, which is the least number

of edges that can be deleted to yield a tree or forest. Thus $\text{cyclo}(G)$ is 0 if G is a tree, and is 1 if G is a unicyclic graph. Also, let $\text{isol}(G)$ be the number of isolated vertices; and let $L_1(G)$ be the maximum order of a component. We shall use one simple deterministic lemma concerning these quantities.

Lemma 23. *Suppose that the graph G satisfies*

$$L_1(G) \leq \text{isol}(G) + 1. \quad (4.1)$$

Then G has a balanced bipartition A, B such that

$$e_G(A) + e_G(B) \leq \text{cyclo}(G). \quad (4.2)$$

Proof. To prove the inequality (4.2), initially set $A = B = \emptyset$. Consider in turn the components C in G other than the isolated vertices. Given C , delete a set D of at most $\text{cyclo}(C)$ edges to obtain a connected bipartite subgraph, with (vertex) parts C_1, C_2 . Add C_1 to A and C_2 to B , or C_1 to B and C_2 to A , so as to keep the maximum of $|A|$ and $|B|$ as small as possible. Observe that the edges in D and no other edges are added to $e(A) \cup e(B)$. At the end of this part of the process, $||A| - |B|| \leq L_1$ and $e_G(A) + e_G(B) \leq \text{cyclo}(G)$. Now consider the (many) isolated vertices, and add them in turn to a smaller of A and B . Finally we have $||A| - |B|| \leq 1$, which completes the proof of (4.2). $\square$

Now let us return to the random graph $G = G_{n,m}$. Whp it satisfies (1) $\text{cyclo}(G) = O(\log n)$, (2) $L_1(G) \leq n^{2/3+o(1)}$, and (3) $\text{isol}(G) = (1/e + o(1))n$. In more detail we have the following. Suppose that $0 \leq s \leq n^{2/3}$ and $m = n/2 + s$. Then $\text{cyclo}(G_{n,m}) = (\frac{2}{3} + o(1)) \log n$ whp, by [21] Theorem 10 (i); and $L_1(G_{n,m}) \leq \omega n^{2/3}$ whp, see for example [14] Theorem 5.23. Finally, $\text{isol}(G_{n,m})$ is asymptotically $\text{Po}(n/e)$, see for example [7].

Hence whp (4.1) holds for G , and so there is a balanced bipartition $\mathcal{A} = (A, B)$ such that the inequality (4.2) holds. We use this partition for the complementary graph $\overline{G} = \overline{G}_{n,m}$. Now

$$\begin{aligned} q_{\mathcal{A}}^E(\overline{G}) &\geq \left(\binom{|A|}{2} + \binom{|B|}{2} - \text{cyclo}(G) \right) / m \\ &\geq (\tfrac{1}{4}n(n-2) - O(\log n)) / (\tfrac{1}{2}n(n-2) - c_0\sqrt{n}) \\ &= \tfrac{1}{2} + (\tfrac{1}{2}c_0\sqrt{n} - O(\log n)) / (\tfrac{1}{2}n(n-2) - c_0\sqrt{n}) \\ &= \tfrac{1}{2} + c_0n^{-3/2} - O(n^{-2} \log n). \end{aligned}$$

For the degree tax, note that if n is odd then $2(\binom{(n+1)/2}{2} - \binom{(n-1)/2}{2}) = n - 1$, and so by (4.2)

$$|\text{vol}_{\overline{G}}(A) - \text{vol}_{\overline{G}}(B)| \leq n - 1 + 2\text{cyclo}(G) \leq n + 2m \leq 3n.$$

Letting $\nu = \text{vol}(\overline{G})$, we have

$$q_{\mathcal{A}}^D(\overline{G}) \leq \nu^{-2}((\tfrac{\nu}{2} + \tfrac{3}{2}n)^2 + (\tfrac{\nu}{2} - \tfrac{3}{2}n)^2) = \nu^{-2}(\tfrac{\nu^2}{2} + \tfrac{9}{2}n^2) = \tfrac{1}{2} + O(n^{-2}).$$

Hence $q_{\mathcal{A}}(\overline{G}) \geq c_0 n^{-3/2} - O(n^{-2} \log n)$, which completes the proof of part (b).

Part (c). In order to prove part (c) of Theorem 5 we start with three deterministic lemmas, Lemmas 24 and 26 which we will apply to the complementary graph, together with Lemma 25 which will be used to prove Lemma 26. We shall need here only the case $k = 2$ of Lemma 24.

Lemma 24. *Let $k \geq 2$ be fixed, and let the n -vertex graph G satisfy $e(G) = o(n^{3/2})$. Let $\mathcal{A} = (A_1, \dots, A_k)$ be a partition of $V(G)$ such that $\max_i |A_i| \leq n/k + o(\sqrt{n})$. Then for the complementary graph $\overline{G}$ we have $q_{\mathcal{A}}^D(\overline{G}) = \frac{1}{k} + o(1/n)$.*

Proof. For each i

$$\text{vol}_{\overline{G}}(A_i) \leq \left(\frac{n}{k} + o(\sqrt{n})\right)(n-1) = \frac{2}{k} \binom{n}{2} + o(n^{3/2}) = \frac{2}{k} e(\overline{G}) + x$$

where $x = \frac{2}{k} e(G) + o(n^{3/2}) = o(n^{3/2})$. Hence, by the convexity of $f(x) = x^2$,

$$\begin{aligned} \sum_{i=1}^k \text{vol}_{\overline{G}}(A_i)^2 &\leq (k-1) \left(\frac{2}{k} e(\overline{G}) + x\right)^2 + \left(\frac{2}{k} e(\overline{G}) - (k-1)x\right)^2 \\ &= \frac{4}{k} e(\overline{G})^2 + k(k-1)x^2 = \frac{4}{k} e(\overline{G})^2 + o(n^3), \end{aligned}$$

and so $q_{\mathcal{A}}^D(\overline{G}) \leq \frac{1}{k} + o(n^{-1})$. Also $q_{\mathcal{A}}^D(\overline{G}) \geq \frac{1}{k}$ since $\mathcal{A}$ has k parts, which completes the proof. $\square$

Lemma 25. *Let F be an n -vertex forest with maximum degree at most d , and let $\rho \geq 1/n$. Then we can find a set of less than d/ρ edges such that deleting these edges yields a forest in which each component tree has at most ρn vertices.*

Proof. Let x be the total number of vertices in components with more than ρn vertices. Thus $x \leq n$, and we may assume that $x > 0$. Let the tree T be a component with more than ρn vertices. Pick a leaf r in T , and orient the edges away from r . Call an edge uv *open* if when we delete it the component containing the terminal vertex v has order more than ρn . Start at r and follow a path of open edges until we reach a vertex z such that no edge out of z is open. Remove from T all the edges incident with z . This removes at most d edges, and decreases x by more than ρn . We need to repeat this less than $1/\rho$ times, and so we delete in total less than d/ρ edges. $\square$

Let G be a graph with n vertices and m edges. The *excess* of G is $m - n$. The *core* of G is the graph obtained by repeatedly deleting leaves, and (if it is non-empty) it is the unique maximal subgraph of G with minimum degree at least 2. We call a bipartition $\mathcal{A} = (A, B)$ *balanced* when $||A| - |B|| \leq 1$.

Lemma 26. *Let G be an n -vertex graph with maximum degree at most d , let $\rho \geq 1/n$, and let G have at least $\rho n - 1$ isolated vertices; and suppose that the core of G has excess at most s , and has at most j components. Then there is a balanced bipartition (A, B) of $V(G)$ such that $e(G) - e(A, B) < s + j + d/\rho$.*

Proof. By removing at most $s + j$ edges from G we may reduce it to a forest F . By Lemma 25, by removing less than d/ρ further edges we may form a forest F' where each (tree) component has at most ρn vertices. We may find a proper 2-colouring (A', B') of F' less its isolated vertices such that $||A'| - |B'|| \leq \rho n - 1$. Now by adding the isolated vertices we can form a proper 2-colouring (A, B) of F' such that $||A| - |B|| \leq 1$. To complete the proof note that $e(A, B) \geq e(F') > e(G) - (s + j + d/\rho)$. $\square$

Now consider the random graph $G \sim G_{n,c/n}$. (We shall be interested in the complementary graph $\overline{G}$.) For each $c \in [1, \infty)$ there is a unique $x = x(c) \in (0, 1]$ such that $xe^{-x} = ce^{-c}$. Here we are following Frieze and Karonski [12]. The function $x(c)$ is continuous and strictly decreasing on $[1, \infty)$.

Lemma 27 (Lemma 2.16 of [12]). *Let $c > 1$ and let $x = x(c)$. Then whp the core of $G_{n,c/n}$ has $(1 - x)(1 - x/c)n + o(n)$ vertices and $(1 - x/c)^2 cn/2 + o(n)$ edges.*

We now use the above lemmas to prove the lower bound in Theorem 5 (c) in the special case when the complementary graph is $G_{n,c/n}$ with c small (precisely, when $1 < c \leq 2$).

Lemma 28. *Let $1 < c \leq 2$, and let $G \sim G_{n,c/n}$. Then there exists $\delta > 0$ such that whp there is a balanced bipartition $\mathcal{A} = (A, B)$ of $V(G)$ with $q_{\mathcal{A}}(\overline{G}) \geq \delta/n + o(1/n)$.*

Proof. Let $x = x(c)$ as above. We shall see that we may take

$$\delta = \frac{1}{2c}((c - 1)(2 - c) + 2(x - 1)^2). \quad (4.3)$$

It is easy to handle the degree tax for any balanced bipartition $\mathcal{A}$. Since $e(G) = cn/2 + o(n) = o(n^{3/2})$ whp, by Lemma 24 we have $q_{\mathcal{A}}^D(\overline{G}) = \frac{1}{2} + o(1/n)$ whp. We need to work to find $\mathcal{A}$ with sufficiently high edge-contribution $q_{\mathcal{A}}^E(\overline{G})$.

By Lemma 27, the core of G whp has $(1 - x)(1 - x/c)n + o(n)$ vertices and $(1 - x/c)^2 cn/2 + o(n)$ edges. So whp the excess of the core is

$$\begin{aligned} (1 - \tfrac{x}{c})^2 \tfrac{cn}{2} - (1 - x)(1 - \tfrac{x}{c})n + o(n) &= (1 - \tfrac{x}{c})(\tfrac{c}{2} - \tfrac{x}{2} - (1 - x))n + o(n) \\ &= \tfrac{1}{2}(1 - \tfrac{x}{c})(c + x - 2)n + o(n). \end{aligned}$$

(It is easy to check that $c + x > 2$.) Also, whp the maximum degree in G is $O(\log n / \log \log n)$, there is exactly one complex component (with more than one cycle), there are $o(n)$ (actually many fewer) unicyclic components, and there are $\Omega(n)$ isolated vertices [12]. Thus, by Lemma 26, whp there is a balanced bipartition (A, B) of $V(G)$ such that

$$e(G) - e_G(A, B) \leq \tfrac{1}{2}(1 - \tfrac{x}{c})(c + x - 2)n + o(n).$$

But $e(G) = \frac{1}{2}cn + o(n)$ whp, so whp

$$\begin{aligned} e_G(A, B) &\geq \frac{1}{2}cn - \frac{1}{2}(1 - \frac{x}{c})(c + x - 2)n + o(n) \\ &= \frac{n}{2}(c - (c + x - 2 - x - \frac{x^2}{c} + \frac{2x}{c})) + o(n) \\ &= (1 - \frac{x(2-x)}{2c})n + o(n). \end{aligned}$$

Since $|A||B| \leq n^2/4$, we may see that

$$\begin{aligned} q_A^E(\overline{G}) &\geq 1 - \frac{\frac{n^2}{4} - e_G(A, B)}{e(\overline{G})} \\ &= 1 - \frac{\frac{1}{2}e(\overline{G}) + (\frac{n^2}{4} - \frac{1}{2}e(\overline{G}) - e_G(A, B))}{e(\overline{G})} \\ &= \frac{1}{2} - \frac{\frac{n^2}{4} - \frac{1}{2}\binom{n}{2} - (e_G(A, B) - \frac{1}{2}e(G))}{e(\overline{G})} \\ &= \frac{1}{2} + \frac{e_G(A, B) - \frac{1}{2}e(G) - \frac{1}{4}n}{e(\overline{G})}. \end{aligned}$$

But by the above, whp

$$e_G(A, B) - \frac{1}{2}e(G) - \frac{1}{4}n \geq \delta n/2 + o(n)$$

where

$$\delta = 2 - \frac{x(2-x)}{c} - \frac{1}{2}(1 + c) = \frac{3}{2} - \frac{x(2-x)}{c} - \frac{1}{2}c.$$

Thus $q_A^E(\overline{G}) \geq \frac{1}{2} + \frac{(\delta/2)n}{e(\overline{G})} + o(1/n)$. Hence, recalling that $q_A^D(\overline{G}) = \frac{1}{2} + o(1/n)$, we have

$$q_A(\overline{G}) \geq \frac{(\delta/2)n}{e(\overline{G})} + o(1/n) = \frac{\delta}{n} + o(1/n) \text{ whp.}$$

We may rewrite δ as

$$\delta = \frac{1}{2c}(3c - 2x(2 - x) - c^2) = \frac{1}{2c}((c - 1)(2 - c) + 2(x - 1)^2)$$

as in (4.3), so clearly $\delta > 0$ since $1 < c \leq 2$; and this completes the proof. $\square$

Observe from (4.3) that if $c_1 = 1 + \varepsilon$ with ε small, then $\delta = \varepsilon/2 + O(\varepsilon^2)$ as $\varepsilon \rightarrow 0$.

The next lemma uses Lemma 28 and completes the proof of the lower bound in part (c) of Theorem 5. Given a graph G and $0 \leq p \leq 1$, let the *percolated* random graph G_p be the subgraph of G obtained by considering each edge independently and keeping it with probability p (and otherwise deleting it). Thus the binomial random graph $G_{n,p}$ could be written as $(K_n)_p$, where K_n is the n -vertex complete graph.

Lemma 29. *Let $1 < c \leq 2$ and let $\delta > 0$ be as in (4.3). Let the probability $p = p(n)$ satisfy $p \rightarrow 1$ as $n \rightarrow \infty$ and $p \leq 1 - c/n$ for n sufficiently large. Then $q^*(G_{n,p}) \geq \delta/n + o(1/n)$ whp.*

Proof. Consider large n . Let $\rho = \rho(n) = p/(1 - c/n)$. We may form $G_{n,p}$ from K_n in two steps. First delete edges independently with probability c/n to give $G_{n,1-c/n}$, and then keep the remaining edges independently with probability ρ . By Lemma 28 there is a balanced bipartition $\mathcal{A}$ of $[n]$ such that whp $q_{\mathcal{A}}(G_{n,1-c/n}) \geq \delta/n + o(1/n)$. Note that also $e(G_{n,1-c/n}) \geq \binom{n}{2} - cn$ whp.

Let H be a graph on $[n]$ with $e(H) \geq \binom{n}{2} - cn$, and suppose that $\mathcal{A}$ is a balanced bipartition of $[n]$ with $q_{\mathcal{A}}(H) \geq \delta/n + o(1/n)$. It suffices to show that the random subgraph H_ρ of H must satisfy $q_{\mathcal{A}}(H_\rho) \geq \delta/n + o(1/n)$ whp. Note that $\rho \rightarrow 1$ as $n \rightarrow \infty$: let $\omega = \omega(n) \rightarrow \infty$ as $n \rightarrow \infty$ sufficiently slowly that $\omega(1 - \rho)^{1/2} \rightarrow 0$.

Consider $e(H_\rho)$, which has the binomial distribution $\text{Bin}(e(H), \rho)$. By Chebyshev's inequality

$$\mathbb{P}(|e(H_\rho) - e(H)\rho| \geq x(e(H)\rho(1 - \rho))^{1/2}) \leq x^{-2}.$$

Thus whp

$$\begin{aligned} e(H_\rho) &= e(H)\rho + O(\omega(e(H)\rho(1 - \rho))^{1/2}) \\ &= e(H)\rho(1 + O((\omega((1 - \rho)/e(H))^{1/2})) \\ &= e(H)\rho(1 + O(\omega(1 - \rho)^{1/2}/n)), \end{aligned}$$

and so

$$e(H_\rho) = e(H)\rho(1 + o(1/n)) \quad \text{whp.} \quad (4.4)$$

For a graph H' on $[n]$ let $e^{\text{int}}(H') = e_{\mathcal{A}}^{\text{int}}(H')$ denote the number of internal edges within the parts of $\mathcal{A}$. Observe that $e^{\text{int}}(H) = (1 + o(1))n^2/4$, since $\mathcal{A}$ is a balanced bipartition and at most $cn = o(n^2)$ edges are missing from H . Thus, much as above, we have

$$e^{\text{int}}(H_\rho) = e^{\text{int}}(H)\rho(1 + o(1/n)) \quad \text{whp.}$$

Hence

$$q_{\mathcal{A}}^E(H_\rho) = \frac{e^{\text{int}}(H_\rho)}{e(H_\rho)} = q_{\mathcal{A}}^E(H)(1 + o(1/n)) = q_{\mathcal{A}}^E(H) + o(1/n) \quad \text{whp.} \quad (4.5)$$

The degree tax is only a little more complicated. Let the parts of $\mathcal{A}$ be A and B , and let $X = \text{vol}_{H_\rho}(A)$. Suppose that H has i edges within A and j edges between A and B , so $\text{vol}_H(A) = 2i + j$. Then X has variance

$$\text{var}(X) = (4i + j)\rho(1 - \rho) \leq 2\text{vol}_H(A)\rho(1 - \rho).$$

Hence, by Chebyshev's inequality as before, whp

$$\begin{aligned} X &= \text{vol}_H(A)\rho + O(\omega(\text{vol}_H(A)\rho(1 - \rho))^{1/2}) \\ &= \text{vol}_H(A)\rho(1 + O((\omega((1 - \rho)/\text{vol}_H(A))^{1/2})) \\ &= \text{vol}_H(A)\rho(1 + O(\omega(1 - \rho)^{1/2}/n)), \end{aligned}$$

since $\text{vol}_H(A) = \Theta(n^2)$. Thus $\text{vol}_{H_\rho}(A) = \text{vol}_H(A)\rho(1 + o(1/n))$ whp; and similarly $\text{vol}_{H_\rho}(B) = \text{vol}_H(B)\rho(1 + o(1/n))$ whp. Hence, using also (4.4), whp

$$\begin{aligned} q_{\mathcal{A}}^D(H_\rho) &= (\text{vol}_{H_\rho}(A)^2 + \text{vol}_{H_\rho}(B)^2)/\text{vol}(H_\rho)^2 \\ &= q_{\mathcal{A}}^D(H)(1 + o(1/n)) = q_{\mathcal{A}}^D(H) + o(1/n). \end{aligned}$$

But now, using also (4.5), whp

$$\begin{aligned} q_{\mathcal{A}}(H_{\rho}) &= q_{\mathcal{A}}^E(H_{\rho}) - q_{\mathcal{A}}^D(H_{\rho}) = q_{\mathcal{A}}^E(H) - q_{\mathcal{A}}^D(H) + o(1/n) \\ &= q_{\mathcal{A}}(H) + o(1/n) \geq \delta/n + o(1/n), \end{aligned}$$

which completes the proof of Lemma 29. $\square$

It remains in this subsection to prove the upper bound in Theorem 5 (c). Let $c > 0$ and let $p \geq 1 - c/n$. Then whp $e(G_{n,p}) \geq \binom{n}{2} - (c + o(1))n/2$, and so $\binom{n}{2} - n/2 - e(G_{n,p}) \leq (c - 1 + o(1))n/2$. Hence by Lemma 15

$$q^*(G_{n,p}) \leq \frac{(c - 1 + o(1))n}{e(G_{n,p})} = 2(c - 1 + o(1))/n \quad \text{whp.}$$

Thus we may set $\beta = 2c_2$ to complete the proof. (Indeed, given any $\varepsilon > 0$, we may set $\beta = 2(c_2 - 1) + \varepsilon$, which shows that we may take β small if c_2 is not much bigger than 1.)

4.2 Proof of Theorem 6

By Theorem 4.1 of [26] there exist $\alpha_1 > 0$ and c_0 such that, if $p = p(n)$ satisfies $1/n \leq p \leq 1 - c_0/n$ for n sufficiently large, then $q^*(G_{n,p}) \geq \alpha_1 \sqrt{\frac{1-p}{np}}$ whp. We shall choose $0 < \alpha \leq \alpha_1$, so we need not consider this range of p further. We may assume that $c_0 \geq 2$, say.

Now let $\varepsilon > 0$, and let p satisfy $1 - c_0/n \leq p \leq 1 - 1/n - \varepsilon/n$. Then

$$\sqrt{\frac{1-p}{np}} \leq \sqrt{\frac{c_0/n}{n(1-c_0/n)}} = (\sqrt{c_0} + o(1/n))/n.$$

By Theorem 5 there exists $\alpha_2 > 0$ such that $q^*(G_{n,p}) \geq \alpha_2/n$ whp. Let $\alpha > 0$ satisfy $\alpha \leq \alpha_1$ and $\alpha < \alpha_2/\sqrt{c_0}$. Then $\alpha_2/n \geq \alpha \sqrt{\frac{1-p}{np}}$ for n sufficiently large, so $q^*(G_{n,p}) \geq \alpha \sqrt{\frac{1-p}{np}}$ whp, as required.

5 Variant forms of degree tax and when modularity is zero

Let G be a graph with $m \geq 1$ edges, and let $\mathcal{A}$ be a partition of $V(G)$. In the definition of the modularity score $q_{\mathcal{A}}(G)$, the degree tax is based on the model of a random pseudograph R which is approximately uniformly distributed over the pseudographs on $V(G)$ with the same degrees as G , where R is generated by the configuration model. Note that R may have loops and multiple edges. As we noted earlier, the expected number of edges in R between distinct vertices i and j is $d_i d_j / (2m - 1)$. If we do not worry too much about loops, this yields

$$\mathbb{E}[e_{\mathcal{A}}^{\text{int}}(R)]/m \approx \sum_A \sum_{i,j \in A} d_i d_j / (4m^2) = (1/4m^2) \sum_A \text{vol}(A)^2 = q_{\mathcal{A}}^D(G), \quad (5.1)$$

and so

$$q_{\mathcal{A}}(G) = q_{\mathcal{A}}^E(G) - q_{\mathcal{A}}^D(G) \approx (1/m) (e_{\mathcal{A}}^{\text{int}}(G) - \mathbb{E}[e_{\mathcal{A}}^{\text{int}}(R)]) . \quad (5.2)$$

This approximation motivated the original definition of modularity [31].

A referee of an earlier version of this paper commented that allowing loops in the random pseudograph R leads to bias in particular against partitions with small parts, and questioned how changing the definition of R to exclude loops would change our results on very dense graphs. In this (new) section we consider five natural variants of the Newman–Girvan definition of modularity, and we investigate which graphs have corresponding modularity value 0. Recall that ‘graph’ here always means ‘simple graph’.

Given a graph G and a vertex-partition $\mathcal{A}$, each modularity variant uses the original edge contribution $q_{\mathcal{A}}^E(G)$, and has its own penalty (corresponding to the degree-tax); and the modularity of G is the maximum of the scores for each vertex-partition. In each case, the modularity value lies in $[0, 1]$; there are graphs with modularity 0 (for example, complete graphs); there are graphs with modularity arbitrarily close to 1 (for example, disjoint unions of many k -cliques for any $k \geq 2$); and isolated vertices are irrelevant, except for modified modularity and Erdős–Rényi modularity.

In Subsection 5.1 we define (precise) ‘configuration modularity’ cq (where R is distributed exactly according to the configuration model), ‘simple modularity’ (or ‘exact modularity’) sq (where R is simple), ‘modified modularity’ $\tilde{q}$ (a ‘local’ version of simple modularity), and ‘loop-free modularity’ lfq (where R is loop-free), and finally we recall the well-established Erdős–Rényi modularity ERq . We note that simple/exact modularity is considered independently in the recent paper [9], and modified modularity is introduced in that paper. Corresponding to $f^q(n)$, we let $f^{cq}(n)$ be the least number of edges missing from the complete graph K_n in a graph with strictly positive configuration modularity; and similarly for the other variants. We characterise those graphs G with simple modularity value $sq^*(G) = 0$, those with modified modularity value $\tilde{q}^*(G) = 0$, and those with loop-free modularity value $lfq^*(G) = 0$; and these results yield easily the values $f^{sq}(n)$, $f^{\tilde{q}}(n)$, and $f^{lfq}(n)$. In Subsection 5.2, we complete the story concerning these values by determining $f^{cq}(n)$ and $f^{ERq}(n)$. See also the table in Figure 3 on page 29.

5.1 Definitions of modularity and natural variants

Calculating the left hand side of (5.1) precisely, we have

$$\begin{aligned} \mathbb{E}[e_{\mathcal{A}}^{\text{int}}(R)]/m &= \frac{1}{m} \sum_A \left(\sum_{i \neq j \in A} \frac{d_i d_j}{2(2m-1)} + \sum_{i \in A} \frac{d_i(d_i-1)}{2(2m-1)} \right) \\ &= \frac{1}{m} \sum_A \left(\sum_{i, j \in A} \frac{d_i d_j}{2(2m-1)} - \sum_{i \in A} \frac{d_i}{2(2m-1)} \right) \\ &= \frac{1}{2m(2m-1)} \sum_A (\text{vol}^2(A) - \text{vol}(A)). \end{aligned}$$

From the last equation,

$$\mathbb{E}[e_{\mathcal{A}}^{\text{int}}(R)]/m = \sum_{A \in \mathcal{A}} \binom{\text{vol}(A)}{2} / \binom{2m}{2} : \quad (5.3)$$

we shall use this observation in equation (5.6) below. Continuing from the previous equations, since $\sum_A \text{vol}(A) = 2m$ we have

$$\mathbb{E}[e_{\mathcal{A}}^{\text{int}}(R)]/m = \frac{4m}{2(2m-1)} \left(\frac{1}{4m^2} \sum_A \text{vol}^2(A) - \frac{1}{2m} \right) = \frac{2m}{2m-1} (q_{\mathcal{A}}^D(G) - \frac{1}{2m}).$$

Thus

$$\mathbb{E}[e_{\mathcal{A}}^{\text{int}}(R)]/m = q_{\mathcal{A}}^D(G) - \frac{1}{2m-1} (1 - q_{\mathcal{A}}^D(G)). \quad (5.4)$$

But $0 < q_{\mathcal{A}}^D(G) \leq 1$ for each vertex-partition $\mathcal{A}$, and so

$$q_{\mathcal{A}}^D(G) - \frac{1}{2m-1} < \mathbb{E}[e_{\mathcal{A}}^{\text{int}}(R)]/m \leq q_{\mathcal{A}}^D(G), \quad (5.5)$$

with equality in the second inequality if and only if $\mathcal{A}$ has a single part (ignoring isolated vertices), in which case both sides equal 1. The inequalities (5.5) yield a more precise version of the approximation (5.2), see also (5.6) below.

*Configuration modularity cq^**

Suppose that we define the (precise) *configuration modularity score* $cq_{\mathcal{A}}(G)$ to be the right hand side of (5.2) (without using the approximation (5.1), though still using the configuration model to generate R). Then, by (5.3),

$$cq_{\mathcal{A}}(G) = (1/m)(e_{\mathcal{A}}^{\text{int}}(G) - \mathbb{E}[e_{\mathcal{A}}^{\text{int}}(R)]) = e_{\mathcal{A}}^{\text{int}}(G)/m - \sum_{A \in \mathcal{A}} \binom{\text{vol}(A)}{2} / \binom{2m}{2}, \quad (5.6)$$

and thus

$$cq_{\mathcal{A}}(G) = q_{\mathcal{A}}(G) + (1 - q_{\mathcal{A}}^D(G))/(2m-1) \quad (5.7)$$

by (5.4). Observe that isolated vertices are irrelevant. Also let the (precise) *configuration modularity* $cq^*(G)$ of G be the maximum value of $cq_{\mathcal{A}}(G)$ over all vertex-partitions $\mathcal{A}$. By (5.7), we have

$$q_{\mathcal{A}}(G) \leq cq_{\mathcal{A}}(G) < q_{\mathcal{A}}(G) + \frac{1}{2m-1},$$

and so

$$q^*(G) \leq cq^*(G) < q^*(G) + \frac{1}{2m-1}. \quad (5.8)$$

Also, if G_m denotes the graph consisting of m isolated edges, then $q^*(G) \leq q^*(G_m) = 1 - 1/m$ (see [35, Cor 1.4.2]), and so $0 \leq cq^*(G) < 1$ for all graphs G . (If $m = 1$ then $cq^*(G) = 0$.) Further, if $q^*(G) > 0$, or $q_{\mathcal{A}}(G) = 0$ for some partition $\mathcal{A}$ with at least two parts of positive volume, then there is a q^* -optimal partition $\mathcal{A}$ with $q_{\mathcal{A}}^D(G) < 1$, and so

variant		degree tax		$f(n)$	
$q_{\mathcal{A}}$	(standard)	$\sum_A \text{vol}(A)^2 / (2m)^2$	Def (7)	$\lfloor n/2 \rfloor + 1$	Thm 1
$cq_{\mathcal{A}}$	configuration	$\sum_A \binom{\text{vol}(A)}{2} / \binom{2m}{2}$	eqn (5.6)	$\lceil n/2 \rceil$	Prop 35
$sq_{\mathcal{A}}$	simple/exact	$\mathbb{E}[e_{\mathcal{A}}^{\text{int}}(S)]$ where S is simple	eqn (5.9)	2	eqn (5.11)
$\tilde{q}_{\mathcal{A}}$	modified/local	(see section on page 33)	eqn (5.14)	2	eqn (5.16)
$lfq_{\mathcal{A}}$	loop-free	$\mathbb{E}[e_{\mathcal{A}}^{\text{int}}(R_0)]$ where R_0 is loop-free	eqn (5.19)	1	eqn (5.20)
$ERq_{\mathcal{A}}$	Erdős-Rényi	$\sum_A A ^2 / n^2$	eqn (5.22)	2	Prop 37

Figure 3: Table showing how variations in the modularity definition would change the results in this paper. For each modularity variant, we write $f(n)$ for the least number of edges missing from any n -vertex graph G with positive modularity value, and list these values (note that we may for example need $n \geq 4$ or similar) along with a reference to where to find the result in the paper.

$cq^*(G) > q^*(G)$ by equation (5.7). For example, consider the complete graph K_n where n is even, and let G be this graph less a perfect matching (so G is the complement of the n -vertex 1-regular graph). Then $q^*(G) = 0$ by Theorem 1, and we saw in the proof of the theorem that there is a bipartition $\mathcal{A}$ with $q_{\mathcal{A}}(G) = 0$; and hence $cq^*(G) \geq cq_{\mathcal{A}}(G) > 0$. See Proposition 35 for the edit-distance $f^{cq}(n)$ from K_n to a graph with positive configuration modularity.

Simple modularity sq^*

As we saw above, in the standard definition of the modularity score $q_{\mathcal{A}}(G)$, the degree tax ‘penalty’ is based on the expected number of internal edges for a random pseudograph R which is nearly uniformly distributed over the graphs on $V(G)$ with the same degrees as G , as generated by the configuration model. What if we insist on the random pseudograph R not having loops or multiple edges (that is, on R being a simple graph), and on R being *exactly* uniformly distributed on the graphs with the same degrees as G ? For a vertex-partition $\mathcal{A}$ of G , define the corresponding ‘simple modularity’ (or ‘exact modularity’) score $sq_{\mathcal{A}}(G)$ by

$$sq_{\mathcal{A}}(G) = (1/m) (e_{\mathcal{A}}^{\text{int}}(G) - \mathbb{E}[e_{\mathcal{A}}^{\text{int}}(R)]) = \frac{1}{m} \sum_{i < j} (\mathbf{1}_{ij \in E(G)} - \mathbb{P}(ij \in E(R))) \delta_{ij}(\mathcal{A}), \quad (5.9)$$

where the simple graph R is uniformly distributed on the graphs on $V(G)$ with the same

degrees as G . (Recall that $\delta_{ij}(\mathcal{A})$ is the indicator that vertices i and j are in the same part in $\mathcal{A}$.)

Let $sq^*(G)$ be the maximum value of $sq_{\mathcal{A}}(G)$ over all vertex-partitions $\mathcal{A}$, and observe that $0 \leq sq^*(G) < 1$. This seems like a natural variant of the definition of modularity, though now there are no simple general formulae like the degree-tax to help us to calculate or approximate a modularity score $sq_{\mathcal{A}}(G)$.

Simple / exact modularity sq is introduced also in Chang and Van Mieghem [9], where it is called exact modularity. That paper notes (as above) the lack of a general formula to calculate the modularity score $sq_{\mathcal{A}}(G)$. (The same paper also introduces a simplified ‘local’ version of sq^* called ‘modified modularity’ $\tilde{q}^*$ – which we discuss next, see page 33).

To get a feel for simple/exact modularity sq , consider first the easy case when G is r -regular, so $rn = 2m$. By symmetry, each edge appears in R with probability $p = m/\binom{n}{2} = r/(n-1)$, so

$$\mathbb{E}[e_{\mathcal{A}}^{\text{int}}(R)] = \sum_A \binom{|A|}{2} p = \sum_A \binom{|A|}{2} r/(n-1).$$

Hence

$$\mathbb{E}[e_{\mathcal{A}}^{\text{int}}(R)]/m = \sum_A \binom{|A|}{2} / \binom{n}{2},$$

the proportion of all $\binom{n}{2}$ possible edges which are internal in $\mathcal{A}$. Also

$$\sum_A \binom{|A|}{2} / \binom{n}{2} = \sum_{A \in \mathcal{A}} |A|^2/n^2 - \frac{1}{n-1} \left(1 - \sum_{A \in \mathcal{A}} |A|^2/n^2\right) = q_{\mathcal{A}}^D(G) - \frac{1}{n-1}(1 - q_{\mathcal{A}}^D(G)),$$

where $q_{\mathcal{A}}^D(G) = \sum_A |A|^2/n^2$ is the usual degree tax (since G is regular). Thus, for a regular graph G with a non-trivial vertex partition $\mathcal{A}$,

$$q_{\mathcal{A}}^D(G) - \frac{1}{n-1} < \mathbb{E}[e_{\mathcal{A}}^{\text{int}}(R)]/m < q_{\mathcal{A}}^D(G).$$

(For the ‘Erdős-Rényi modularity’, we always use $\sum_A |A|^2/n^2$ as the degree-tax, see later in this section.) For example, if G is the 4-cycle C_4 () which is also $K_{2,2}$, then the set of labelled simple graphs with the same degree sequence is the set of graphs isomorphic to G (i.e., $\{\text{---}, \text{---}, \text{---}\}$). Now if $\mathcal{A}$ is a bipartition with $e_{\mathcal{A}}^{\text{int}}(G) = 2$, then

$$q_{\mathcal{A}}(G) = \frac{2}{4} - \frac{4^2 + 4^2}{8^2} = 0 \quad \text{and} \quad sq_{\mathcal{A}}(G) = \frac{2}{4} - \frac{2}{6} = \frac{1}{6}, \quad (5.10)$$

and both these values are optimal (that is, $q^*(G) = 0$ and $sq^*(G) = \frac{1}{6}$).

For a non-regular example, let G be a 3-edge path (). If $\mathcal{A}$ is the bipartition with $e_{\mathcal{A}}^{\text{int}}(G) = 2$, then

$$q_{\mathcal{A}}(G) = \frac{2}{3} - \frac{3^2 + 3^2}{6^2} = \frac{1}{6},$$

and noting that there are exactly two (simple) labelled graphs with the same degrees as G (i.e., , ), we may see that

$$sq_{\mathcal{A}}(G) = \frac{2}{3} - \frac{\frac{1}{2}(2+0)}{3} = \frac{1}{3}.$$

Both these values are optimal (that is, $q^*(G) = \frac{1}{6}$ and $sq^*(G) = \tilde{q}^*(G) = \frac{1}{3}$).

We have been interested in the graphs G for which $q^*(G) = 0$: but when is $sq^*(G) = 0$? We say that a graph G is *determined by its degrees* if G is the unique graph on $V(G)$ with the same vertex degrees as G .

Proposition 30. *For every graph G , $sq^*(G) = 0$ if and only if G is determined by its degrees.*

Proof. If there is a unique graph on $V(G)$ with the degrees of G , then R must be G ; hence $\mathbb{E}[e_{\mathcal{A}}^{\text{int}}(R)] = e_{\mathcal{A}}^{\text{int}}(G)$ for each vertex-partition $\mathcal{A}$, and so $sq^*(G) = 0$. If G is not determined by its degrees, then there must be an edge vw of G such that vw is an edge in R with probability < 1 . Let $\mathcal{A}$ be the partition of $V(G)$ with one part $\{v, w\}$ and each other part a singleton. Then $e_{\mathcal{A}}^{\text{int}}(G) = 1$ and $\mathbb{E}[e_{\mathcal{A}}^{\text{int}}(R)] < 1$, so $sq^*(G) \geq sq_{\mathcal{A}}(G) > 0$. $\square$

We can tell easily if a graph G on $[n]$ is determined by its degrees. This is always the case for $n \leq 3$. Let $n \geq 4$ and let G be a graph on $[n]$. Let a, b, c, d be four distinct vertices of G ; and suppose that ab and cd are edges, and bc and ad are non-edges. (Thus a, b, c, d form an ‘alternating 4-cycle’ – edge, non-edge, edge, non-edge.) The corresponding 2-switch is the replacement of edges ab and cd by new edges bc and ad . By a theorem of Berge (1973) (see for example Theorem 1.3.33 in [38]) the degrees of G determine G if and only if G has no 2-switches. Observe also that G has a 2-switch if and only if the complementary graph $\overline{G}$ has a 2-switch. We can say more.

Theorem 31. *For every graph G with at least one edge, the following are equivalent.*

- (a) G is determined by its degrees.
- (b) G has no 2-switch.
- (c) Repeatedly deleting an isolated or universal vertex destroys G (leaving nothing).

It is straightforward to prove this result directly, without using Berge’s result; and we do so now.

Proof. (a) $\implies$ (b). Performing a 2-switch on G yields a distinct graph G' with the same degrees as G , so this part of Berge’s result is immediate.

(b) $\implies$ (c). Suppose that $\delta(G), \delta(\overline{G}) \geq 1$. It suffices to show that G has a 2-switch. Suppose wlog that $\delta := \delta(G) \leq \delta(\overline{G})$. Let vertex a have degree δ in G . Fix a neighbour b of a in G , and let $Z = N_G(a) \setminus \{b\}$. Then b has at least $\delta = |Z| + 1$ neighbours in $\overline{G}$, so in $\overline{G}$ vertex b has a neighbour $c \notin Z$. Note that vertices a, b, c are distinct, $ab \in E(G)$, and $ac, bc \in E(\overline{G})$. Vertex c has degree in G at least $\delta = |Z| + 1$ so c has a neighbour $d \notin Z$ in G . Now $d \notin \{a, b, c\}$ and $ad \in E(\overline{G})$, so we have a 2-switch, as required.

(c) $\implies$ (a). This is trivial if $v(G) \leq 3$. Let $k \geq 3$ and suppose it holds for all graphs with k vertices. Let graph G have $v(G) = k + 1$, and let graph H on $V(G)$ have the same degrees as G . We must show that $H = G$.

Let vertex v in G be isolated or universal, let $G' = G - v$ and $H' = H - v$. Consider a vertex w in $V(G') = V(H')$: if v was isolated in G then in G' and H' , each vertex w has

degree $d_G(w)$; and if v was universal in G then in G' and H' , each vertex w has degree $d_G(w) - 1$. Thus H' has the same degrees as G' , so $H' = G'$ by the induction hypothesis, and so $H = G$ as required. $\square$

We may read off various results from Proposition 30 and Theorem 31. Recall that $f^{\text{sq}}(n)$ is the minimum number of edges that may be deleted from K_n to form a graph G with $sq^*(G) > 0$. If G is the complete graph K_n , or G is K_n less one edge, then G is determined by its degrees, so $sq^*(G) = 0$. If G is K_n less two disjoint edges, then G has a 2-switch and so is not determined by its degrees, and $sq^*(G) > 0$ (for $n \geq 4$). Hence

$$f^{\text{sq}}(n) = 2 \quad \text{for each } n \geq 4. \quad (5.11)$$

If G is a complete multipartite graph with at least 2 parts of size at least 2, then again G has a 2-switch, so G is not determined by its degrees, and thus $sq^*(G) > 0$ (whereas $q^*(G) = 0$ in contrast). If G is K_n less all the edges in a set $A \subset [n]$, or if G is K_n less the edges of a star, then G is determined by its degrees (for example, part (c) of Theorem 31 clearly holds) so $sq^*(G) = 0$.

Finally here, let us observe some features of sq^* not shared by q^* , concerning a graph G and its complement $\overline{G}$. We saw above that if G is C_4 then $q^*(G) = 0$; and the complementary graph $\overline{G}$ consists of two disjoint edges, so $q^*(\overline{G}) = \frac{1}{2}$. But clearly a graph G is determined by its degrees if and only if this holds for the complementary graph $\overline{G}$. Hence, by Proposition 30

$$sq^*(G) = 0 \quad \text{if and only if} \quad sq^*(\overline{G}) = 0. \quad (5.12)$$

For clarity, for $ij \in E(G)$ we shall write $ij \in G$, etc. We may take $R(\overline{G})$ to be the complement $\overline{R(G)}$ of $R(G)$. Thus

$$\mathbb{P}(ij \in R(G)) + \mathbb{P}(ij \in R(\overline{G})) = 1.$$

Hence

$$\begin{aligned} sq_{\mathcal{A}}(G) + sq_{\mathcal{A}}(\overline{G}) &= \frac{1}{m} \sum_{i < j} (\mathbf{1}_{ij \in G} - \mathbb{P}(ij \in R(G)) + \mathbf{1}_{ij \in \overline{G}} - \mathbb{P}(ij \in R(\overline{G}))) \delta_{ij}(\mathcal{A}) \\ &= \frac{1}{m} \sum_{i < j} (1 - 1) \delta_{ij}(\mathcal{A}) = 0. \end{aligned}$$

Thus, for every graph G and vertex-partition $\mathcal{A}$,

$$sq_{\mathcal{A}}(G) + sq_{\mathcal{A}}(\overline{G}) = 0. \quad (5.13)$$

Now we may amplify (5.12) as follows.

Proposition 32. *For every graph G , the following four statements are equivalent: (a) $sq^*(G) = 0$, (b) $sq^*(\overline{G}) = 0$, (c) $sq_{\mathcal{A}}(G) = 0$ for every vertex-partition $\mathcal{A}$, (d) $sq_{\mathcal{A}}(\overline{G}) = 0$ for every vertex-partition $\mathcal{A}$.*

Modified modularity $\tilde{q}^*$

Modified modularity may be considered to be a ‘local’ version of simple modularity. It was introduced in [9], together with simple modularity.

For modified modularity, we consider the pairs of vertices $i < j$ in V separately, and let $\tilde{R}_{ij}$ be a random graph on V picked uniformly from the set of all graphs with the same number of edges as G , and with degrees d_i and d_j the same as in G . Then we set

$$\tilde{q}_{\mathcal{A}}(G) = \frac{1}{m} \sum_{i < j} \left(\mathbf{1}_{ij \in E(G)} - \mathbb{P}(ij \in E(\tilde{R}_{ij})) \right) \delta_{ij}(\mathcal{A}) \quad (5.14)$$

(recall $\delta_{ij}(\mathcal{A})$ is the indicator that vertices i and j are in the same part in the partition $\mathcal{A}$). Also, we let $\tilde{q}^*(G)$ be the maximum value of $\tilde{q}_{\mathcal{A}}(G)$ over all vertex partitions $\mathcal{A}$, and note that $0 \leq \tilde{q}^*(G) < 1$. Observe that, to determine the ‘penalty term’ for vertices i, j in (5.14), we consider a graph chosen uniformly at random from those which *locally* match the degrees of G , as well as matching the total number of edges. For a graph G with vertex set V and pair of distinct vertices i, j , we denote by $\mathcal{G}(G, ij)$ the set of labelled simple graphs G' with $V' = V$, $d'_i = d_i$, $d'_j = d_j$ and $m' = m$ (where V' , d'_i and m' denote the vertex-set, degree of vertex i , and number of edges in G' respectively), so that $\tilde{R}_{ij}$ is chosen uniformly from $\mathcal{G}(G, ij)$.

We again begin with some examples to get a feel for the definition. First, suppose that G is the 4-cycle C_4 ( with vertex labels a, b, c, d clockwise from bottom-left). The set of labelled simple graphs with $d_a = d_b = 2$ and $m = 4$ edges is $\mathcal{G}(\text{img}, ab) = \{\text{img}, \text{img}, \text{img}, \text{img}, \text{img}\}$. Thus $\tilde{R}_{ab}$ is chosen uniformly from these five graphs, and the probability that $ab \in E(\tilde{R}_{ab})$ is $\frac{4}{5}$; and this is the same for all pairs of vertices by symmetry. Now let $\mathcal{A}$ be the bipartition $\{\{a, b\}, \{c, d\}\}$, with edges ab and cd as internal edges. Thus

$$\tilde{q}_{\mathcal{A}}(G) = \frac{1}{4} \left(\left(1 - \frac{4}{5}\right) + \left(1 - \frac{4}{5}\right) \right) = \frac{1}{10} > 0,$$

and this is optimal, so $\tilde{q}^*(G) = \frac{1}{10}$. (To check optimality, note that including any non-edge within a part would lead to a contribution of $-\frac{1}{5}$.) Note that the modified modularity score $\tilde{q}_{\mathcal{A}}(G)$ for this partition is different from the scores for both usual modularity $q_{\mathcal{A}}(G) = 0$, and simple modularity $sq_{\mathcal{A}}(G) = 1/6$ both given in (5.10).

We now consider a non-regular example – see also Fig. 4 (left) on page 34. Let G be the 3-edge path on vertices $V = \{a, b, c, d\}$ with edges ab, bc, cd (i.e.,  labelled clockwise from bottom-left). Then $\tilde{R}_{ab}$ is chosen uniformly from the set of graphs on V with $d_a = 1$, $d_b = 2$ and $m = 3$. There are exactly four such graphs, i.e., $\mathcal{G}(\text{img}, ab) = \{\text{img}, \text{img}, \text{img}, \text{img}\}$, and so the probability that ab is an edge in $\tilde{R}_{ab}$ is $\frac{1}{2}$. Note that by symmetry, this is the same for $\tilde{R}_{ac}$ (that is, the probability that ac is an edge in $\tilde{R}_{ac}$ is $\frac{1}{2}$), for $\tilde{R}_{bd}$, and for $\tilde{R}_{cd}$. Similarly, $\tilde{R}_{bc}$ is chosen uniformly from the set of graphs on V with $d_b = 2$, $d_c = 2$ and $m = 3$ (i.e., $\{\text{img}, \text{img}, \text{img}, \text{img}\}$), and each of these graphs have the edge bc . Lastly $\tilde{R}_{ad}$ is chosen uniformly from $\{\text{img}, \text{img}\}$, and neither of these graphs have the edge ad .

We may now see that ab, cd are the only pairs that can make a positive contribution in (5.14), and ac, bd make a negative contribution; – see also last column of Fig. 4. Hence,

$$G = \text{---}\text{---}\text{---}$$

$$G = \text{---}\text{---}\text{---}$$

i, j	$\mathcal{G}(\text{---}\text{---}\text{---}, ij)$	p_{ij}	w_{ij}
a, b		$\frac{1}{2}$	$\frac{1}{2}$
c, d		$\frac{1}{2}$	$\frac{1}{2}$
b, c		1	0
a, c		$\frac{1}{2}$	$-\frac{1}{2}$
b, d		$\frac{1}{2}$	$-\frac{1}{2}$
a, d		0	0

i, j	$\mathcal{G}(\text{---}\text{---}\text{---}, ij)$	p_{ij}	w_{ij}
a, b		1	0
a, c		0	0
a, d		0	0
b, c		1	0
b, d		1	0
c, d		$\frac{4}{5}$	$\frac{1}{5}$

Figure 4: Calculation details for the modified modularity score of the 3-edge path (left), and the graph consisting of a triangle with a pendant edge (right), where for both graphs the vertex sets are labelled clockwise from bottom-left. Here, $p_{ij} = \mathbb{P}(ij \in E(\tilde{R}_{ij}))$ and $w_{ij} = \mathbf{1}_{ij \in E(G)} - \mathbb{P}(ij \in E(\tilde{R}_{ij}))$. See also the discussions preceding (5.15) and (5.17).

the bipartition $\mathcal{A} = \{\{a, b\}, \{c, d\}\}$ is the unique optimal partition for $\tilde{q}$, and

$$\tilde{q}^*(G) = \tilde{q}_{\mathcal{A}}(G) = \frac{1}{3} \left(\left(1 - \frac{1}{2}\right) + \left(1 - \frac{1}{2}\right) \right) = \frac{1}{3}. \quad (5.15)$$

Recall that $q_{\mathcal{A}}(G) = \frac{1}{6}$ and $sq_{\mathcal{A}}(G) = \frac{1}{3}$; and that $\mathcal{A}$ is an optimal partitions for both standard and simple modularities.

When do we have $\tilde{q}^*(G) = 0$? We say that the graph G is *locally edge-determined by its degrees* if, for every edge ij in G , every graph G' in $\mathcal{G}(G, ij)$ contains the edge ij . (The set $\mathcal{G}(G, ij)$ is defined just below (5.14).) The following result shows that we can tell easily if $\tilde{q}^*(G) = 0$.

Theorem 33. *For every graph G with at least one edge, the following are equivalent.*

- (a) $\tilde{q}^*(G) = 0$.
- (b) G is locally edge-determined by its degrees.
- (c) Either every edge of G is incident with a universal vertex, or G less any isolated vertices is a triangle or a star.

From Theorem 33 we may quickly determine $f^{\tilde{q}}$, and we do so now before proceeding with the proof of the theorem. Let $n \geq 4$. If G is a complete graph K_n or K_n less one edge, then every edge of G is incident to a universal vertex, so $\tilde{q}^*(G) = 0$. If G is K_n less

any two edges, then some edge is not incident to any universal vertex, G has no isolated vertices, and G is not a triangle or a star, and so $\tilde{q}^*(G) > 0$ by Theorem 33. Hence

$$f^{\tilde{q}}(n) = 2 \quad \text{for each } n \geq 4. \quad (5.16)$$

Proof of Theorem 33. Let G have m edges. Note first that for all distinct vertices i, j in G we have $d_i + d_j \leq m + 1$ (since otherwise G would have more than m edges). Also, if $d_i + d_j = m + 1$ then G and every other graph in $\mathcal{G}(G, ij)$ must contain the edge ij . We shall show that $(b) \implies (a)$, then $(c) \implies (b)$ and finally $(a) \implies (c)$.

$(b) \implies (a)$. Let G be locally edge-determined by its degrees. First note that for any graph, the singletons partition which places each vertex in a separate part has modified modularity score 0, so it suffices to prove that $\tilde{q}^*(G) \leq 0$. For each edge ij of G , the random graph $\tilde{R}_{ij}$ must contain the edge ij , that is $\mathbb{P}(ij \in E(\tilde{R}_{ij})) = 1$. Thus, whatever the partition $\mathcal{A}$, if there is a corresponding summand in (5.14) it is 0. Also, any summand corresponding to a non-edge is ≤ 0 and so $\tilde{q}_{\mathcal{A}}(G) \leq 0$. Thus $\tilde{q}^*(G) = 0$, as required.

$(c) \implies (b)$. If an edge ij of G is incident with a universal vertex, then it must be in every graph in $\mathcal{G}(G, ij)$. Thus if each edge of G is incident with a universal vertex, then G is locally determined by its degrees.

Now let G^- be G less any isolated vertices, and suppose that G^- is a triangle or a star. It remains to show that G is locally determined by its degrees. If ij is an edge in G , then $d_i + d_j = m + 1$, so each graph in $\mathcal{G}(G, ij)$ contains the edge ij . Hence, again G is locally determined by its degrees, as required.

It remains to show that $(a) \implies (c)$. Assume that (a) holds, let G^- be G less any isolated vertices, and assume for a contradiction that G^- is not a triangle or a star, and ij is an edge in G with neither i nor j universal. Suppose that $d_i + d_j \leq m$. Then there is a graph G' in $\mathcal{G}(G, ij)$ with no edge ij , so the partition $\mathcal{A}$ with one part $\{i, j\}$ and each other part a singleton has $\tilde{q}_{\mathcal{A}}(G) > 0$, and thus (a) fails. Thus we must have $d_i + d_j = m + 1$.

We may assume wlog that $d_i \geq d_j$. If $d_j = 1$ then $d_i = m$, so G^- is a star, contradicting our assumptions. Hence $d_j \geq 2$, and there is an edge jh where $h \neq i$. Now h can be adjacent only to i and j , so $d_h \leq 2$. Suppose that $m \geq 4$. Then $d_j + d_h \leq (m + 1)/2 + 2 = m + (5 - m)/2 \leq m + 1/2$, and so $d_j + d_h \leq m$. This gives a contradiction as above, and thus completes the proof of the case $m \geq 4$. Finally suppose that $m \leq 3$. If G^- is not a triangle or a star, then either G^- is the 3-edge path (i.e., $\vdash$) and $\tilde{q}^*(G) = 1/3$ (as we saw above), or G has edges uv and $u'v'$ in distinct components (so there is an alternating cycle C_4). In the latter case, let $\mathcal{A}$ be the vertex-partition with parts $\{u, v\}$ and $\{u', v'\}$ and all other parts singletons. By considering the graph with edges uv and $u'v'$ replaced by uu' and vv' we see that $\tilde{q}_{\mathcal{A}}(G) > 0$. This contradiction completes the proof that $(a) \implies (c)$, and thus completes the entire proof. $\square$

Observe that if a graph G is locally edge-determined by its degrees, then every graph with all the same degrees as G contains all the edges of G , and so it is G : thus G is determined by its degrees. Hence, by Theorem 33 and Proposition 30, $\tilde{q}^*(G) = 0$ implies $sq^*(G) = 0$.

However, there are graphs G with $sq^*(G) = 0$ but such that $\tilde{q}^*(G) > 0$. For example, let the 4-edge graph G on $\{a, b, c, d\}$ have leaf vertex a adjacent to vertex b in the triangle

b, c, d ($\mathfrak{N}$ labelled clockwise from bottom-left) – see also Fig. 4 (right). Observe that G is the unique graph with vertex degrees 1, 3, 2, 2 (and so $sq^*(G) = 0$). But consider vertices c, d and $\tilde{R}_{cd}$. Given that there are $m = 4$ edges and $d_c = 2, d_d = 2$, there are exactly five possible graphs, namely the set $\mathcal{G}(\mathfrak{N}, cd) = \{\mathfrak{N}, \mathfrak{M}, \mathfrak{L}, \mathfrak{X}, \mathfrak{K}\}$. Thus cd is an edge in $\tilde{R}_{cd}$ with probability $\frac{4}{5}$. Hence, letting $\mathcal{A}$ be the partition with parts $\{a\}$, $\{b\}$, and $\{c, d\}$, we have

$$\tilde{q}^*(G) = \tilde{q}_{\mathcal{A}}(G) = (1 - \frac{4}{5})/4 = \frac{1}{20} > 0 \quad (5.17)$$

where the optimality follows since cd is the only pair ij for which we have a positive contribution to the score $w_{ij} = \mathbf{1}_{ij \in E(G)} - \mathbb{P}(ij \in E(\tilde{R}_{ij}))$, see Fig 4 (right).

Two undesirable properties of modified modularity $\tilde{q}$ are that the trivial one-part partition can have modularity score > 0 , and that isolated vertices can make a difference. Both these properties are illustrated by the graph $K_4 + v = \mathfrak{K} \bullet$, which is K_4 plus an isolated vertex v . For $\tilde{q}^*(K_4) = 0$ (for example by Theorem 33), but the trivial one-part partition $\mathcal{T}$ of $K_4 + v$ has modified modularity score $\tilde{q}_{\mathcal{T}}(K_4 + v) > 0$, as we see in the next paragraph.

Indeed, let $n \geq 4$, start with the complete graph K_n , and remove t edges incident with one vertex v , where $t = 2$ if $n = 4$ and $2 \leq t \leq n - 1$ if $n \geq 5$; and call the resulting graph G . (When $n = 4$, G is the graph $\mathfrak{N}$, which we discussed above; and when $n = 5$ and $t = 4$, G is $K_4 + v$.) Then the trivial one-part partition $\mathcal{T}$ gives $\tilde{q}_{\mathcal{T}}(G) > 0$. To see this, consider a ‘missing edge’ vw (not in G), and note that no graph $\tilde{R}_{vw}$ can contain this edge, since then it would have too few edges. Thus $\tilde{q}_{\mathcal{T}}(G) \geq 0$. Further, let the vertex x be such that vx is also a missing edge (with $x \neq w$): then some graph $\tilde{R}_{wx}$ in $\mathcal{G}(G, wx)$ does not contain the edge wx , since we can obtain such a graph by deleting the edge wx from G , adding edges vw and vx , and finally deleting an edge not incident with w or x . This yields a strictly positive contribution to the score, and so $\tilde{q}_{\mathcal{T}}(G) > 0$.

To conclude here, we note that (unlike sq^*) we can have $\tilde{q}^*(G) > 0$ and $\tilde{q}^*(\overline{G}) = 0$. Again, take $G = \mathfrak{N}$, and we have $q^*(G) = 1/20 > 0$. Note that $\overline{G} = \mathfrak{Z}$ is a star on a, c, d with centre a , and an isolated vertex b . Thus by Theorem 33 we have $q^*(\overline{G}) = 0$.

Observe that $\mathfrak{Z}$ is locally edge-determined, i.e., for both edges ac and ad in $\mathfrak{Z}$, the sets $\mathcal{G}(\mathfrak{Z}, ac) = \{\mathfrak{Z}, \mathfrak{V}\}$ and $\mathcal{G}(\mathfrak{Z}, ad) = \{\mathfrak{Z}, \mathfrak{I}\}$ contain only graphs which contain the edges ac and ad respectively. However, for the non-edge cd in $\mathfrak{Z}$, we have $\mathcal{G}(\mathfrak{Z}, cd) = \{\mathfrak{Z}, \mathfrak{N}, \mathfrak{X}, \mathfrak{L}, \mathfrak{K}\}$ (i.e., the complements of the graphs in $\mathcal{G}(\mathfrak{N}, cd)$), which includes a graph with cd an edge, and a graph with cd a non-edge. In general, the graphs in $\mathcal{G}(\overline{G}, ij)$ are exactly the complements of the graphs in $\mathcal{G}(G, ij)$. Thus for $\tilde{R}_{ij}(\overline{G})$ we may take the complement of $\tilde{R}_{ij}(G)$; and so

$$\mathbb{P}(ij \in E(\tilde{R}_{ij}(G))) + \mathbb{P}(ij \in E(\tilde{R}_{ij}(\overline{G}))) = 1.$$

Hence, as for simple modularity sq in (5.13), for any vertex-partition $\mathcal{A}$ we have

$$\tilde{q}_{\mathcal{A}}(G) + \tilde{q}_{\mathcal{A}}(\overline{G}) = 0. \quad (5.18)$$

Loop-free modularity

Above, we introduced simple modularity $sq^*(G)$ and modified modularity $\tilde{q}^*$, where we ‘penalised’ by considering a random *simple* graph R which is uniformly distributed on the graphs with the same degrees as G , or ‘locally’ the same degrees as G ; and explored when we have $sq^*(G) = 0$ and when we have $\tilde{q}^*(G) = 0$. What if we insist that the random graph has no loops but it is allowed multiple edges?

Let R_0 be a multigraph, with no loops, uniformly distributed on such multigraphs on $V(G)$ with the same degrees as G . For a vertex-partition $\mathcal{A}$ of G , define the corresponding ‘loop-free modularity’ $lfq_{\mathcal{A}}^*(G)$ by

$$lfq_{\mathcal{A}}(G) = (1/m) \left(e_{\mathcal{A}}^{\text{int}}(G) - \mathbb{E}[e_{\mathcal{A}}^{\text{int}}(R_0)] \right). \quad (5.19)$$

Let $lfq^*(G)$ be the maximum value of $lfq_{\mathcal{A}}(G)$ over all vertex-partitions $\mathcal{A}$, and observe that $0 \leq lfq^*(G) < 1$. Let S be chosen uniformly at random over all simple graphs with the same degrees as G . When G is a regular graph with m edges, the expected number of edges between any distinct pair of vertices is $m/\binom{n}{2}$ for both the simple graph S and the loop-free graph R_0 . Hence for regular graphs G , we have $\mathbb{E}[e_{\mathcal{A}}^{\text{int}}(R_0)] = \mathbb{E}[e_{\mathcal{A}}^{\text{int}}(S)]$, and so $lfq^*(G) = sq^*(G)$. We saw in Proposition 30 exactly which graphs G have $sq^*(G) = 0$: a small subset of these graphs have $lfq^*(G) = 0$.

Proposition 34. *For any graph G , $lfq^*(G) = 0$ if and only if G less any isolated vertices is a star or a complete graph.*

Proof. Sufficiency is easy. Let G^- denote G less any isolated vertices. If G^- is a star then R_0 must be G , so $lfq_{\mathcal{A}}(G) = 0$ for each vertex-partition $\mathcal{A}$, and $lfq^*(G) = 0$. If G^- is complete and $e \in E(G^-)$, then by symmetry the expected number of times e appears in R_0 is 1. Thus $lfq_{\mathcal{A}}(G) = 0$ for each vertex-partition $\mathcal{A}$, and $lfq^*(G) = 0$.

To prove necessity we first introduce a weakened version of a 2-switch. If a, b, c, d are distinct vertices in a graph G and $ab, cd \in E(G)$ and $bc \in E(\overline{G})$ we call this a 2^- -switch (or an ‘alternating P_3 ’). (There is no restriction on ad .) Observe that a 2-switch is a 2^- -switch but not necessarily conversely. Now we may argue as in the proof of Proposition 30. Suppose that G has a 2^- -switch on a, b, c, d . Then we may form a multigraph G' with the same degrees as G by deleting the edges ab, cd and adding 1 to the number of copies of bc and ad . Thus the probability that bc is an edge in R_0 is > 0 , so there must be an edge vw of G such that $\mathbb{E}[e_{R_0}(vw)] < 1$. Let $\mathcal{A}$ be the partition of $V(G)$ with one part $\{v, w\}$ and each other part a singleton. Then $e_{\mathcal{A}}^{\text{int}}(G) = 1$ and $\mathbb{E}[e_{\mathcal{A}}^{\text{int}}(R_0)] < 1$, so $lfq^*(G) \geq lfq_{\mathcal{A}}(G) > 0$.

Now suppose that $lfq^*(G) = 0$, and so G has no 2^- -switch. Again, let G^- denote G less any isolated vertices. We may assume that G^- has at least 4 vertices, and is not a star. Thus there are two disjoint edges; and since there is no 2^- -switch, any two disjoint edges are part of a complete graph K_4 . In particular G^- is connected. Let W with $|W| \geq 3$ be a maximal set of vertices in G^- which induce a complete graph. If some vertex is not in W then there is an edge ab with $a \notin W$ and $b \in W$. Let $c \in W$, $c \neq b$. Then $ac \in E(G^-)$ since there exists $d \in W \setminus \{b, c\}$ and a, b, c, d do not form a 2^- -switch. Thus vertex a is

adjacent to each vertex in W , contradicting the maximality of W . Hence G^- is complete, and we are done. $\square$

Recall that $f^{\text{lfq}}(n)$ is the minimum number of edges that may be deleted from K_n to form a graph G with $\text{lfq}^*(G) > 0$. An easy corollary of Proposition 34 is that

$$f^{\text{lfq}}(n) = 1 \quad \text{for each } n \geq 4. \quad (5.20)$$

Consider an m -edge graph G , with a given vertex-partition. In each of the above four variant forms of modularity – namely (precise) configuration modularity cq , simple modularity sq , modified modularity $\tilde{q}$, and loop-free modularity lfq – the score is penalised by using the expected behaviour of a suitable random graph or pseudograph. We specify a set $\mathcal{H}$ of pseudographs on V , each with the same number m of edges as G , and also with the same vertex-degrees as G (except that, for modified modularity, we have only ‘locally’ the same degrees). Then we let the random pseudograph R be uniformly distributed on $\mathcal{H}$ (for sq , $\tilde{q}$ and lfq), or at least nearly uniformly distributed on $\mathcal{H}$ (for cq , assuming suitable vertex degrees). For any partition $\mathcal{A}$ of V , the corresponding modularity score is defined to be

$$q_{\mathcal{A}}^R(G) = \frac{1}{m}(e_{\mathcal{A}}^{\text{int}}(G) - \mathbb{E}[e_{\mathcal{A}}^{\text{int}}(R)]). \quad (5.21)$$

The fifth and final modularity variant we consider here is quite different, with a simple form of degree tax.

Erdős-Rényi modularity

This form of modularity has been studied and used for several years, see for example [13, 39]. For a graph G and vertex-partition $\mathcal{A}$, the Erdős-Rényi modularity score is defined to be

$$\text{ER}q_{\mathcal{A}}(G) = e_{\mathcal{A}}^{\text{int}}(G)/m - \sum_A |A|^2/n^2; \quad (5.22)$$

and the Erdős-Rényi modularity $\text{ER}q^*(G)$ is the maximum value of this score over all vertex-partitions $\mathcal{A}$. Observe that $0 \leq \text{ER}q^*(G) < 1$. If G is regular then clearly $\text{ER}q_{\mathcal{A}}(G) = q_{\mathcal{A}}(G)$ for each vertex-partition $\mathcal{A}$, and $\text{ER}q^*(G) = q^*(G)$. In particular we see that $\text{ER}q^*(K_n) = 0$, and if n is even and G is obtained from K_n by deleting a perfect matching then $\text{ER}q^*(G) = 0$.

However, we will see later, in Proposition 37, that it is enough to delete just two edges from K_n to obtain a graph with positive Erdős-Rényi modularity. For an easy example, let $n = 3$ and let G be obtained by deleting two edges from K_3 , so G consists of an edge and isolated vertex, $G = \bullet\text{---}\bullet$; and let $\mathcal{A}$ be the bipartition which takes the adjacent vertices as one part and the isolated vertex as the other. Then $\text{ER}q_{\mathcal{A}}(G) = 1 - (1/3)^2 - (2/3)^2 = 4/9$. Note that this last example shows that adding an isolated vertex can change the Erdős-Rényi modularity, since for G' a single edge $\bullet\text{---}\bullet$ we have $\text{ER}q^*(G') = 0$.

The standard modularity score $q_{\mathcal{A}}(G)$ and the Erdős-Rényi modularity score $\text{ER}q_{\mathcal{A}}(G)$ are both examples of a ‘modularity score with weighted degree tax’, or a ‘modularity

score penalised by weighted part sizes'. Let the graph G have vertex set $V = [n]$, and let $\mathbf{c} = (c_1, \dots, c_n)$ be given vertex weights $c_j > 0$. For $A \subseteq V$ let $c(A) = \sum_{j \in A} c_j$. Given a partition $\mathcal{A}$ of V , the corresponding 'weighted tax on parts' or ' $\mathbf{c}$ -penalty' is $\sum_{A \in \mathcal{A}} c(A)^2 / c(V)^2$; and the corresponding modularity score is

$$q_{\mathcal{A}}^{\mathbf{c}}(G) = \frac{1}{m} e_{\mathcal{A}}^{\text{int}}(G) - \sum_{A \in \mathcal{A}} c(A)^2 / c(V)^2.$$

Observe that, if each c_j is the degree of vertex j then $q_{\mathcal{A}}^{\mathbf{c}}(G)$ is the standard modularity score $q_{\mathcal{A}}(G)$; and if each $c_j = 1$ then $q_{\mathcal{A}}^{\mathbf{c}}(G)$ is the Erdős-Rényi modularity score $ERq_{\mathcal{A}}(G)$. Thus the versions of modularity we have considered above (including the standard modularity) are particular cases of either an R -penalised modularity or a $\mathbf{c}$ -penalised modularity.

5.2 Edit distance from complete graphs to a positive modularity graph

Recall that $f^{\text{cq}}(n)$ is the least number of edges missing from any n -vertex graph G with $cq^*(G) > 0$ (that is, with positive precise configuration modularity); and similarly for $f^{\text{ERq}}(n)$. In this section we determine f^{cq} and f^{ERq} . Note that analogous results for simple modularity sq , modified modularity $\tilde{q}$, and loop-free modularity lfq have already been established, since they followed directly from other results in the previous subsection, see (5.11), (5.16) and (5.20).

Configuration modularity cq

Proposition 35. *For each $n \geq 4$ we have $f^{\text{cq}}(n) = \lceil n/2 \rceil$.*

We first prove a lemma analogous to Lemma 8. Let $G = (V, E)$ be a graph. Similarly to $p_G(U)$, for $U \subseteq V$, let

$$\tilde{p}_G(U) = -e(U, \bar{U})(\text{vol}(G) - 1) + \text{vol}(U)\text{vol}(\bar{U}).$$

Observe that $\tilde{p}_G(\emptyset) = \tilde{p}_G(V) = 0$ (as for p_G).

Lemma 36. *Let G be a graph with vertex set V and with $m \geq 1$ edges. Then the following three statements are equivalent: (i) $cq^*(G) = 0$, (ii) $cq_{\mathcal{A}}(G) \leq 0$ for each bipartition $\mathcal{A}$ of V , and (iii) $\tilde{p}_G(U) \leq 0$ for each $U \subseteq V$.*

Proof. Observe that for any $A \subseteq V$, $\text{vol}(A) - 1 = 2m - 1 - \text{vol}(\bar{A})$. Thus for any partition $\mathcal{A}$ of V ,

$$\begin{aligned} cq_{\mathcal{A}}(G) &= 1 - \frac{1}{2m} \sum_A e(A, \bar{A}) - \frac{1}{2m(2m-1)} \sum_A \text{vol}(A)(2m-1 - \text{vol}(\bar{A})) \\ &= -\frac{1}{2m} \sum_A e(A, \bar{A}) + \frac{1}{2m(2m-1)} \sum_A \text{vol}(A)\text{vol}(\bar{A}); \end{aligned}$$

and so, writing $\nu = \text{vol}(G)$ ($= 2m$),

$$cq_{\mathcal{A}}(G) = \nu^{-1}(\nu - 1)^{-1} \sum_{A \in \mathcal{A}} \tilde{p}_G(A).$$

But $\tilde{p}_G(U)$ is symmetric in U and $\bar{U}$, so $\tilde{p}_G(U) = \tilde{p}_G(\bar{U})$, and the lemma follows. $\square$

Proof of Proposition 35. To prove the lower bound we must show that if $n \geq 4$ and we form G by removing $x < \lceil n/2 \rceil$ edges from K_n then $cq^*(G) = 0$. Note that $x \leq (n-1)/2$, and so G has minimum degree $\delta_G \geq (n-1)/2$, and in particular $\delta_G \geq 2$ since $n \geq 4$.

We first note an inequality (corresponding roughly to (2.1)). Let $\emptyset \neq U \subset V$, and note that there is an edge e in G between U and $\bar{U}$ (since $n-1-x \geq (n-1)/2 > 0$). Suppose that there is a missing edge within U or $\bar{U}$, and form G' by ‘moving’ the edge e to the location of such a missing edge. Then

$$\tilde{p}_{G'}(U) > \tilde{p}_G(U). \quad (5.23)$$

To prove this, we may assume that the edge e is added to U . Since $e_{G'}(U, \bar{U}) = e_G(U, \bar{U}) - 1$, $\text{vol}_{G'}(U) = \text{vol}_G(U) + 1$ and $\text{vol}_{G'}(\bar{U}) = \text{vol}_G(\bar{U}) - 1$, we have

$$\begin{aligned} \tilde{p}_{G'}(U) - \tilde{p}_G(U) &= (2m-1)(e_G(U, \bar{U}) - e_{G'}(U, \bar{U})) \\ &\quad + (\text{vol}_G(U) + 1)(\text{vol}_G(\bar{U}) - 1) - \text{vol}_G(U)\text{vol}_G(\bar{U}) \\ &= 2m - 1 + \text{vol}_G(\bar{U}) - \text{vol}_G(U) - 1. \end{aligned}$$

But $\text{vol}_G(\bar{U}) \geq \delta_G \geq 2$, and so $\text{vol}_G(U) \leq 2m - 2$. Hence

$$\tilde{p}_{G'}(U) - \tilde{p}_G(U) \geq 2 > 0,$$

and (5.23) follows.

As in the proof of Theorem 1, suppose that the graph H on $V = [n]$ and $\emptyset \neq U \subset V$ are such that $\tilde{p}_H(U)$ maximises the value $\tilde{p}_G(W)$ over all graphs G on V with at most $(n-1)/2$ edges missing and all $\emptyset \neq W \subset V$. There must be edges between U and $\bar{U}$; so by inequality (5.23), all missing edges are between U and $\bar{U}$.

Suppose that $|U| = j$ (where $1 \leq j \leq n-1$), each missing edge in H is between U and $\bar{U}$, and x edges are missing from H , where $0 \leq x < \lceil n/2 \rceil$. Then

$$\begin{aligned} \tilde{p}_H(U) &= \text{vol}_H(U)\text{vol}_H(\bar{U}) - e_H(U, \bar{U})(\text{vol}(H) - 1) \\ &= (j(n-1) - x)((n-j)(n-1) - x) - (j(n-j) - x)(n(n-1) - 2x - 1) \\ &= j(n-j)(2x - (n-2)) - x^2 - x. \end{aligned}$$

Thus $\tilde{p}_H(U) < 0$ if $x \leq (n-2)/2$. Suppose that n is odd, and consider the case that $x = (n-1)/2$. Then $\tilde{p}_H(U) = -(n-2j+1)(n-2j-1)/4$; and so for all integer j , $\tilde{p}_H(U) \leq 0$ with equality when $j = (n-1)/2$ or $j = (n+1)/2$. Hence for $x < \lceil n/2 \rceil$, we have $\tilde{p}_H(U) \leq 0$, which by Lemma 36 proves the lower bound.

We now prove the upper bound. For odd n , we may take $x = (n+1)/2$ and $j = (n-1)/2$, and then $\tilde{p}_H(U) = (n+1)(n-3)/2$. Thus for odd $n \geq 4$, there is a graph

G with $(n+1)/2$ edges missing and $cq^*(G) > 0$. Similarly, for even $n \geq 4$, we may take $x = j = n/2$, and then $\tilde{p}_H(U) = n(n-2)/4$. Thus, for even $n \geq 4$, there is a graph G with $n/2$ edges missing and $cq^*(G) > 0$. $\square$

Erdős-Rényi modularity ERq^*

Proposition 37. *For each $n \geq 3$ we have $f^{\text{ERq}}(n) = 2$.*

As for configuration modularity, we first prove a lemma analogous to Lemma 8. For a graph $G = (V, E)$ with n vertices and m edges, for $U \subseteq V$, let

$$r_G(U) = -e(U, \bar{U})n^2 + |U||\bar{U}|2m.$$

Observe that $r_G(\emptyset) = r_G(V) = 0$ (as for p_G and $\tilde{p}_G$).

Lemma 38. *Let G be a graph with $m \geq 1$ edges. Then the following three statements are equivalent: (i) $ERq^*(G) = 0$, (ii) $ERq_{\mathcal{A}}(G) \leq 0$ for each bipartition $\mathcal{A}$ of $V(G)$, and (iii) $r_G(U) \leq 0$ for each $U \subseteq V(G)$.*

Proof. Begin by noting that for any vertex-partition $\mathcal{A}$,

$$\begin{aligned} ERq_{\mathcal{A}}(G) &= 1 - \frac{1}{2m} \sum_{A \in \mathcal{A}} e(A, \bar{A}) - \frac{1}{n^2} \sum_{A \in \mathcal{A}} |A|(n - |\bar{A}|) \\ &= -\frac{1}{2m} \sum_{A \in \mathcal{A}} e(A, \bar{A}) + \frac{1}{n^2} \sum_{A \in \mathcal{A}} |A||\bar{A}| \\ &= \frac{1}{2mn^2} \sum_{A \in \mathcal{A}} r_G(A). \end{aligned}$$

Since $r_G(U)$ is symmetric in U and $\bar{U}$, we have $r_G(U) = r_G(\bar{U})$, and the lemma follows. $\square$

Proof of Proposition 37. To prove the lower bound we must show that, if $n \geq 3$ and we form G by removing one or zero edges from K_n , then $ERq^*(G) = 0$.

Again, we note an inequality (corresponding to (2.1)). Let G be a graph on vertex set V , let $\emptyset \neq U \subset V$, and form G' by ‘moving’ an edge currently between U and $\bar{U}$ to the location of a missing edge within U or $\bar{U}$. Then (noting that the sizes of U and $\bar{U}$ do not change),

$$\begin{aligned} r_{G'}(U) &= -e_{G'}(U, \bar{U})n^2 + |U||\bar{U}|2e(G') \\ &= -(e_G(U, \bar{U}) - 1)n^2 + |U||\bar{U}|2e(G), \end{aligned}$$

and so

$$r_{G'}(U) > r_G(U). \tag{5.24}$$

Again, as in the proof of Theorem 1, suppose that the graph H on $V = [n]$ and $\emptyset \neq U \subset V$ are such that $r_H(U)$ maximises the value $r_G(W)$ over all graphs G on V with at most one edge missing. Note that if there is a missing edge then by (5.24) it is between U and $\bar{U}$.

Suppose that $|U| = j$ (where $1 \leq j \leq n - 1$) and x edges are missing from H , where $x \in \{0, 1\}$. Then

$$\begin{aligned} r_H(U) &= -e_H(U, \overline{U}) n^2 + |U| |\overline{U}| 2m \\ &= -(j(n - j) - x) n^2 + j(n - j)(n(n - 1) - 2x) \\ &= -j(n - j)(n + 2x) + n^2 x. \end{aligned}$$

Clearly $r_H(U) < 0$ for $x = 0$; and for $x = 1$, noting that $j(n - j) \geq n - 1$, we have

$$r_H(U) = -j(n - j)(n + 2) + n^2 \leq -(n - 1)(n + 2) + n^2 = -(n - 2) < 0$$

since $n \geq 3$. This proves the lower bound, that $f^{\text{ERq}}(n) \geq 2$.

To see the upper bound, let $n \geq 3$ and consider the graph G formed from K_n by picking a vertex v and deleting two of the incident edges. Let $\mathcal{A}$ be the bipartition with one part $\{v\}$, and note that, since $x = 2$ and $j = |\{v\}| = 1$, we have

$$r_G(\{v\}) = -(n - 1)(n + 4) + 2n^2 = n^2 - 3n + 4 = (n - 1)(n - 2) + 2 > 0;$$

and hence by Lemma 38, the bipartition $\mathcal{A}$ has positive Erdős-Rényi modularity. $\square$

6 Concluding remarks

We considered graphs with modularity zero, with starting point the complete graphs and complete bipartite graphs. For each such complete graph, we found the least number of edges which we may add to or delete from or edit to obtain a graph with non-zero modularity.

Complete multipartite graphs also have modularity zero but some corresponding questions for them are still open - see Section 2.5. It is also open, see Question 4, whether complete graphs are the furthest (in edit distance) from graphs with positive modularity.

We also investigated the modularity of very dense random graphs, and in particular we found that there is a transition to modularity zero when the average degree of the complementary graph drops below 1.

Finally, in Section 5, we considered some natural variant definitions of modularity, found the distance from a complete graph to one with positive modularity value for each variant, and for simple modularity $sq^*(G)$, modified modularity $\tilde{q}^*(G)$, and loop-free modularity $lfq^*(G)$, we found precisely which graphs have corresponding modularity zero. See also [28], where we investigate further properties of these variants, for example when we limit the number of parts.

Acknowledgements

We would like to thank the referees for helpful and interesting comments. One such comment was that the standard definition of modularity penalised partitions with small parts, since the degree tax was based on random graphs which may have loops. This

led to the study in Section 5 of some variant definitions, and when the corresponding modularity value is 0.

FS was partially supported by the Wallenberg AI, Autonomous Systems and Software Program WASP. This material is partially based upon work supported by the National Science Foundation under Grant No. DMS-1928930, while FS was in residence at the Simons Laufer Mathematical Sciences Institute (MSRI) in Berkeley, California, during the Spring 2025 semester.

References

- [1] V. Agdur, N. Kamčev, and F. Skerman. Universal lower bound for community structure of sparse graphs. [arXiv:2307.07271](https://arxiv.org/abs/2307.07271), 2023.
- [2] S. Bhamidi, D. Gamarnik, R. van der Hofstad, N. Litvak, P. Pralat, F. Skerman, and Y. Tousinejad. The stochastic block model has the overlap gap property for modularity. In *53rd International Colloquium on Automata, Languages, and Programming (ICALP 2026)*. Schloss Dagstuhl–Leibniz-Zentrum für Informatik, 2026. [doi:10.4230/LIPIcs.ICALP.2026.28](https://doi.org/10.4230/LIPIcs.ICALP.2026.28)
- [3] P. J. Bickel and A. Chen. A nonparametric view of network models and Newman–Girvan and other modularities. *Proceedings of the National Academy of Sciences*, 106(50), 2009. [doi:10.1073/pnas.0907096106](https://doi.org/10.1073/pnas.0907096106)
- [4] P. J. Bickel, A. Chen, Y. Zhao, E. Levina, and J. Zhu. Correction to the proof of consistency of community detection. *The Annals of Statistics*, 2015. [doi:10.1214/14-AOS1271](https://doi.org/10.1214/14-AOS1271)
- [5] V. Blondel, J. Guillaume, R. Lambiotte, and E. Lefebvre. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10):P10008, 2008. [doi:10.1088/1742-5468/2008/10/P10008](https://doi.org/10.1088/1742-5468/2008/10/P10008)
- [6] M. Bolla, B. Bullins, S. Chaturapruek, S. Chen, and K. Friedl. Spectral properties of modularity matrices. *Linear Algebra and Its Applications*, 473:359–376, 2015. [doi:10.1016/j.laa.2014.10.039](https://doi.org/10.1016/j.laa.2014.10.039)
- [7] B. Bollobás. *Random graphs*. Cambridge Studies in Advanced Mathematics (Book 73), 2nd edition. Cambridge University Press, Cambridge, 2001. [doi:10.1007/978-1-4612-0619-4_7](https://doi.org/10.1007/978-1-4612-0619-4_7)
- [8] U. Brandes, D. Dellling, M. Gaertler, R. Gorke, M. Hoefer, Z. Nikoloski, and D. Wagner. On modularity clustering. *Knowledge and Data Engineering, IEEE Transactions on*, 20(2):172–188, 2008. [doi:10.1109/TKDE.2007.190689](https://doi.org/10.1109/TKDE.2007.190689)
- [9] B. L. Chang and P. Van Mieghem. Modularity with a more accurate baseline model. *Physical Review E*, 111(4):044317, 2025. [doi:10.1103/PhysRevE.111.044317](https://doi.org/10.1103/PhysRevE.111.044317)
- [10] J. Chellig, N. Fountoulakis, and F. Skerman. The modularity of random graphs on the hyperbolic plane. *Journal of Complex Networks*, 10(1):cnab051, 2022. [doi:10.1093/comnet/cnab051](https://doi.org/10.1093/comnet/cnab051)

- [11] T. N. Dinh and M. T. Thai. Finding community structure with performance guarantees in scale-free networks. In *Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third International Conference on Social Computing (SocialCom), 2011 IEEE Third International Conference on*. IEEE, 2011. [doi:10.1109/PASSAT/SocialCom.2011.185](https://doi.org/10.1109/PASSAT/SocialCom.2011.185)
- [12] A. Frieze and M. Karoński. *Introduction to random graphs*. Cambridge University Press, 2015.
- [13] M. Gösgens, R. van der Hofstad, and N. Litvak. The hyperspherical geometry of community detection: modularity as a distance. *Journal of Machine Learning Research*, 24(112):1–36, 2023.
- [14] S. Janson, T. Łuczak, and A. Ruciński. *Random Graphs*, volume 45. John Wiley & Sons, 2011.
- [15] B. Kamiński, P. Misiorek, P. Prałat, and F. Théberge. Modularity based community detection in hypergraphs. *Journal of Complex Networks*, 12(5):cnae041, 2024. [doi:10.1093/comnet/cnae041](https://doi.org/10.1093/comnet/cnae041)
- [16] B. Kamiński, B. Pankratz, P. Prałat, and F. Théberge. Modularity of the ABCD random graph model with community structure. *Journal of Complex Networks*, 10(6):cnac050, 2022. [doi:10.1093/comnet/cnac050](https://doi.org/10.1093/comnet/cnac050)
- [17] A. Lancichinetti and S. Fortunato. Limits of modularity maximization in community detection. *Physical Review E*, 84(6):066122, 2011. [doi:10.1103/PhysRevE.84.066122](https://doi.org/10.1103/PhysRevE.84.066122)
- [18] M. Lason and M. Sulkowska. Modularity of minor-free graphs. *Journal of Graph Theory*, 102(4):728–736, 2023. [doi:10.1002/jgt.22896](https://doi.org/10.1002/jgt.22896)
- [19] 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. [doi:10.1002/rsa.21080](https://doi.org/10.1002/rsa.21080)
- [20] B. Louf, C. McDiarmid, and F. Skerman. Modularity and graph expansion. In *15th Innovations in Theoretical Computer Science Conference (ITCS 2024)*. Schloss-Dagstuhl-Leibniz Zentrum für Informatik, 2024. [doi:10.4230/LIPIcs.ITCS.2024.78](https://doi.org/10.4230/LIPIcs.ITCS.2024.78)
- [21] T. Łuczak. Component behavior near the critical point of the random graph process. *Random Structures & Algorithms*, 1(3):287–310, 1990.
- [22] T. Łuczak. On the equivalence of two basic models of random graph. In *Proceedings of Random graphs*, volume 87, pages 151–159, 1990.
- [23] S. Majstorovic and D. Stevanovic. A note on graphs whose largest eigenvalues of the modularity matrix equals zero. *Electronic Journal of Linear Algebra*, 27(1):256, 2014.
- [24] C. McDiarmid and F. Skerman. Modularity and random graphs. In *Topics in Probabilistic Graph Theory*, edited by R. J. Wilson, L. W. Beineke, and C. McDiarmid. Cambridge University Press.

- [25] C. McDiarmid and F. Skerman. Modularity of regular and treelike graphs. *Journal of Complex Networks*, 6(4), 2018. [doi:10.1093/comnet/cnx046](https://doi.org/10.1093/comnet/cnx046)
- [26] C. McDiarmid and F. Skerman. Modularity of Erdős-Rényi random graphs. *Random Structures & Algorithms*, 57(1):211–243, 2020. [doi:10.1002/rsa.20910](https://doi.org/10.1002/rsa.20910)
- [27] C. McDiarmid and F. Skerman. Modularity and partially observed graphs. [arXiv:2112.13190](https://arxiv.org/abs/2112.13190), 2021.
- [28] C. McDiarmid and F. Skerman. Redefining the degree tax in modularity. *In preparation*, 2026+.
- [29] F. de Montgolfier, M. Soto, and L. Viennot. Asymptotic modularity of some graph classes. In *Algorithms and Computation*, pages 435–444. Springer, 2011. [doi:10.1007/978-3-642-25591-5_45](https://doi.org/10.1007/978-3-642-25591-5_45)
- [30] M. E. J. Newman. *Networks: An Introduction*. Oxford University Press, 2010.
- [31] M. E. J. Newman and M. Girvan. Finding and evaluating community structure in networks. *Physical Review E*, 69(2):026113, 2004.
- [32] V. Poda and C. Matias. Comparison of modularity-based approaches for nodes clustering in hypergraphs. *Peer Community Journal*, 4, 2024. [doi:10.24072/pcjournal.404](https://doi.org/10.24072/pcjournal.404)
- [33] L. O. Prokhorenkova, A. Raigorodskii, and P. Prałat. Modularity of complex networks models. *Internet Mathematics*, 2017.
- [34] K. Rybarczyk and M. Sulkowska. Modularity of preferential attachment graphs. In *Theoretical Aspects of Computer Science (STACS), 2026 43rd International Symposium on*, 2026. [doi:10.4230/LIPIcs.STACS.2026.76](https://doi.org/10.4230/LIPIcs.STACS.2026.76)
- [35] F. Skerman. *Modularity of Networks*. PhD thesis, University of Oxford, 2016. [doi:10.5287/ora-z56gx5pd9](https://doi.org/10.5287/ora-z56gx5pd9)
- [36] V. Traag, L. Waltman, and N. Van Eck. From Louvain to Leiden: guaranteeing well-connected communities. *Scientific reports*, 9(1):1–12, 2019. [doi:10.1038/s41598-019-41695-z](https://doi.org/10.1038/s41598-019-41695-z)
- [37] S. Trajanovski, H. Wang, and P. Van Mieghem. Maximum modular graphs. *The European Physical Journal B-Condensed Matter and Complex Systems*, 85(7):1–14, 2012. [doi:10.1140/epjb/e2012-20898-3](https://doi.org/10.1140/epjb/e2012-20898-3)
- [38] D. West. *Introduction to graph theory*, volume 2. Prentice Hall Upper Saddle River, 2001.
- [39] Y. Zhao, E. Levina, and J. Zhu. Consistency of community detection in networks under degree-corrected stochastic block models. *The Annals of Statistics*, 40(4):2266, 2012. [doi:10.1214/12-AOS1036](https://doi.org/10.1214/12-AOS1036)