

# New results for MaxCut in $H$ -free graphs

Stefan Glock <sup>\*</sup>

Oliver Janzer <sup>†</sup>

Benny Sudakov <sup>‡</sup>

## Abstract

The MaxCut problem asks for the size  $\text{mc}(G)$  of a largest cut in a graph  $G$ . It is well known that  $\text{mc}(G) \geq m/2$  for any  $m$ -edge graph  $G$ , and the difference  $\text{mc}(G) - m/2$  is called the *surplus* of  $G$ . The study of the surplus of  $H$ -free graphs was initiated by Erdős and Lovász in the 70s, who in particular asked what happens for triangle-free graphs. This was famously resolved by Alon, who showed that in the triangle-free case the surplus is  $\Omega(m^{4/5})$ , and found constructions matching this bound. We prove several new results in this area.

- (i) Confirming a conjecture of Alon, Krivelevich and Sudakov, we show that for every fixed odd  $r \geq 3$ , any  $C_r$ -free graph with  $m$  edges has surplus  $\Omega_r(m^{\frac{r+1}{r+2}})$ . This is tight, as is shown by a construction of pseudorandom  $C_r$ -free graphs due to Alon and Kahale. It improves previous results of several researchers, and complements a result of Alon, Krivelevich and Sudakov which is the same bound when  $r$  is even.
- (ii) Generalizing the result of Alon, we allow the graph to have triangles, and show that if the number of triangles is a bit less than in a random graph with the same density, then the graph has large surplus. For regular graphs our bounds on the surplus are sharp.
- (iii) We prove that an  $n$ -vertex graph with few copies of  $K_r$  and average degree  $d$  has surplus  $\Omega_r(d^{r-1}/n^{r-3})$ , which is tight when  $d$  is close to  $n$  provided that a conjectured dense pseudorandom  $K_r$ -free graph exists. This result is used to improve the best known lower bound (as a function of  $m$ ) on the surplus of  $K_r$ -free graphs.

Our proofs combine techniques from semidefinite programming, probabilistic reasoning, as well as combinatorial and spectral arguments.

## 1 Introduction

*MaxCut* is a central problem in discrete mathematics and theoretical computer science. Given a graph  $G$ , a *cut* is a partition of the vertex set into two parts, and its size is the number of edges going across. The aim is to determine the maximum size of a cut, denoted here by  $\text{mc}(G)$ . This problem has received a lot of attention in the last 50 years, both from an algorithmic perspective in theoretical computer science, where the aim is to approximate  $\text{mc}(G)$  well for a given graph, and from an extremal perspective in combinatorics, where we are mainly interested in good bounds on  $\text{mc}(G)$  in terms of the number of vertices and/or edges of  $G$ .

A folklore observation is that every graph with  $m$  edges has a cut of size at least  $m/2$ . This can be easily seen using a probabilistic argument, or a greedy algorithm. Therefore, it is a fundamental question by how much this trivial bound can be improved. Since it was already demonstrated by Erdős [11] in the 60s that the factor  $1/2$  cannot be improved in general, even if we consider very restricted families such as graphs of large girth, the natural parameterization is to consider the so-called *surplus* of a graph  $G$ , denoted  $\text{sp}(G)$ , which is the difference  $\text{mc}(G) - \frac{m}{2}$  of the optimal cut size to the greedy bound. By a classical result of Edwards [9, 10], every graph  $G$  with  $m$  edges has surplus

$$\text{sp}(G) \geq \frac{\sqrt{8m+1} - 1}{8}, \tag{1.1}$$

---

<sup>\*</sup>Fakultät für Informatik und Mathematik, Universität Passau, Germany. Email: [stefan.glock@uni-passau.de](mailto:stefan.glock@uni-passau.de). Research supported by Dr. Max Rössler, the Walter Haefner Foundation and the ETH Zürich Foundation.

<sup>†</sup>Department of Pure Mathematics and Mathematical Statistics, University of Cambridge, United Kingdom. Email: [oj224@cam.ac.uk](mailto:oj224@cam.ac.uk). Research supported by an ETH Zürich Postdoctoral Fellowship 20-1 FEL-35.

<sup>‡</sup>Department of Mathematics, ETH Zürich, Switzerland. Email: [benny.sudakov@gmail.com](mailto:benny.sudakov@gmail.com). Research supported in part by SNSF grant 200021\_196965.

and this is tight whenever  $G$  is a complete graph with an odd number of vertices.

Although the bound  $\Omega(\sqrt{m})$  on the surplus is optimal in general, it can be significantly improved for graphs that are “far” from complete. One natural way to enforce this is to forbid the containment of a fixed subgraph. The study of MaxCut in  $H$ -free graphs was initiated by Erdős and Lovász (see [12]) in the 70s, and has received significant attention since then (e.g. [2, 3, 5, 6, 14, 21, 22, 24]). For a graph  $H$ , define  $\text{sp}(m, H)$  as the minimum surplus  $\text{sp}(G) = \text{mc}(G) - m/2$  over all  $H$ -free graphs  $G$  with  $m$  edges.

It was shown in [3] that for every fixed graph  $H$ , there exist constants  $\epsilon = \epsilon(H) > 0$  and  $c = c(H) > 0$  such that  $\text{sp}(m, H) \geq cm^{1/2+\epsilon}$  for all  $m$ . This demonstrates that the surplus of  $H$ -free graphs is significantly larger than the bound (1.1) in Edwards’ theorem. Perhaps the main conjecture in the area, due to Alon, Bollobás, Krivelevich and Sudakov [3], is that the constant  $1/2$  in the aforementioned result can be replaced with  $3/4$ , which would be optimal as can be seen by considering a random graph with edge probability  $p = n^{-\delta}$  for some arbitrarily small  $\delta > 0$  and  $H = K_r$  sufficiently large. This conjecture is still wide open. In fact, we do not even know whether there exists a constant  $\alpha > 1/2$  such that uniformly for all  $H$ , we have  $\text{sp}(m, H) \geq c_H m^\alpha$ , where  $c_H > 0$  may depend on  $H$ .

An even more ambitious problem, first explicitly posed in [5], is to determine the asymptotic growth rate of  $\text{sp}(m, H)$  for every fixed graph  $H$ . However, results of the type  $\text{sp}(m, H) = \Theta(m^{q_H})$  for some constant  $q_H$  depending only on  $H$ , are still very rare. The case when  $H$  is a triangle has received particularly much attention. Erdős and Lovász (see [12]) first showed that  $\text{sp}(m, K_3) = \Omega\left(m^{2/3} (\log m / \log \log m)^{1/3}\right)$ . They also noted that it seems unclear what the correct exponent of  $m$  should be, even in this elementary case, which testifies about the difficulty of the problem. The logarithmic factor was later improved by Poljak and Tuza [21]. Shearer [22] proved an important bound on the surplus of any triangle-free graph  $G$  showing that

$$\text{sp}(G) = \Omega\left(\sum_{v \in V(G)} \sqrt{d_G(v)}\right). \quad (1.2)$$

This bound is tight in general and implies that  $\text{sp}(m, K_3) = \Omega(m^{3/4})$ . Finally, Alon [2] proved that  $\text{sp}(m, K_3) = \Theta(m^{4/5})$ . His result is exemplary for many of the challenges and developed methods in the area.

Our contribution in this paper is threefold. First, we determine  $q_H$  for all odd cycles, which improves earlier results of several researchers and adds to the lacunary list of graphs for which  $q_H$  is known. Secondly, we extend the result of Alon in the sense that we prove optimal bounds on the surplus of general graphs in terms of the number of triangles they contain. Thirdly, we study the surplus of graphs with few cliques of size  $r$  and improve the currently best bounds for  $K_r$ -free graphs.

## 1.1 Cycles of odd length

The study of MaxCut in graphs without short cycles goes back to the work of Erdős and Lovász (see [12]). Motivated by one of their conjectures, Alon, Bollobás, Krivelevich and Sudakov [3] showed that  $m$ -edge graphs not containing any cycle of length at most  $r$  have surplus  $\Omega_r(m^{(r+1)/(r+2)})$ . Later, Alon, Krivelevich and Sudakov [5] proved that

$$\text{sp}(m, C_r) = \Omega_r\left(m^{(r+1)/(r+2)}\right) \quad (1.3)$$

for all even  $r$ . They also conjectured that this bound is tight. It is well known (see e.g. [2, 8, 17]) that if a graph  $G$  has smallest eigenvalue  $\lambda_{\min}$ , then

$$\text{sp}(G) \leq |\lambda_{\min}| |G| / 4. \quad (1.4)$$

In fact, essentially all known extremal examples for the MaxCut problem of  $H$ -free graphs come from  $(n, d, \lambda)$ -graphs (i.e.,  $n$ -vertex  $d$ -regular graphs where each nontrivial eigenvalue has absolute value at most  $\lambda$ ). For instance, the well-known Erdős–Rényi graphs [13] are  $C_4$ -free  $(n, d, \lambda)$ -graphs with  $d = \Theta(\sqrt{n})$  and  $\lambda = O(\sqrt{d})$ , which arise as polarity graphs of finite projective planes. By (1.4), these

graphs have surplus at most  $O(\lambda n) = O((nd)^{5/6})$ , which shows the optimality of the bound (1.3) in the case  $r = 4$ . Similar constructions work in the cases  $r \in \{6, 10\}$  (see [5] for more details). However, showing the tightness of (1.3) in general is probably very hard since it seems to go hand in hand with the construction of extremal examples for the Turán problem of even cycles.

Another obvious problem left open by the work of Alon, Krivelevich and Sudakov [5] is to consider odd cycles. In fact, Alon, Krivelevich and Sudakov [5] write that it seems plausible that (1.3) holds for odd cycles too. One advantage here is that the conjectured extremal examples are already known. Indeed, the construction of triangle-free dense pseudorandom graphs due to Alon [1] generalizes to longer odd cycles, as was observed by Alon and Kahale [4] (see also [16]). This yields, for every fixed odd  $r$ ,  $C_r$ -free  $(n, d, \lambda)$ -graphs with  $d = \Theta(n^{2/r})$  and  $\lambda = O(\sqrt{d})$ . By (1.4), these graphs have surplus at most  $O(\lambda n) = O((nd)^{(r+1)/(r+2)})$ , which shows that (1.3) would be optimal for all odd  $r$ . Regarding this problem, Zeng and Hou [24] showed that  $\text{sp}(m, C_r) \geq m^{(r+1)/(r+3)+o(1)}$  for all odd  $r$ , and very recently Fox, Himwich and Mani [14] improved the surplus to  $\Omega_r(m^{(r+5)/(r+7)})$ . We settle this problem completely by proving the following tight result, which confirms the conjecture of Alon, Krivelevich and Sudakov [5].

**Theorem 1.1.** *For odd  $r \geq 3$  there is a constant  $\alpha_r > 0$  such that any  $C_r$ -free graph with  $m$  edges has a cut of size at least  $m/2 + \alpha_r \cdot m^{(r+1)/(r+2)}$ . This is tight up to the value of  $\alpha_r$ .*

## 1.2 Few triangles

Shearer's bound (1.2) implies that any triangle-free  $d$ -degenerate graph with  $m$  edges has surplus  $\Omega(m/\sqrt{d})$ . In his proof of  $\text{sp}(m, K_3) = \Omega(m^{4/5})$ , Alon [2] combined this estimate with an additional argument which shows that a triangle-free graph with average degree  $d$  has surplus  $\Omega(d^2)$ . This supersedes Shearer's result when  $d$  is larger than roughly  $m^{2/5}$ . It is natural to ask what happens with both bounds if we do not insist that the graph is triangle-free, but only has few triangles.

Alon, Krivelevich and Sudakov [5] proved an extension of Shearer's bound (1.2) to graphs with few triangles. They showed that if every vertex  $v \in V(G)$  is contained in  $o(d_G(v)^{3/2})$  triangles, then (1.2) remains true. Their motivation was to use this as a tool to study the surplus of  $H$ -free graphs when  $H$  is an even cycle or a small complete bipartite graph, since then neighbourhoods are sparse and hence every vertex is contained in few triangles. Recently, Carlson, Kolla, Li, Mani, Sudakov and Trevisan [6] showed that the local assumption can be relaxed to a global condition, namely, any  $d$ -degenerate graph with  $m$  edges and  $O(m\sqrt{d})$  triangles has surplus  $\Omega(m/\sqrt{d})$ . Their proof is based on a probabilistic rounding of the solution of the semidefinite programming relaxation of MaxCut. This is one of our main tools too, and we will discuss it in more detail in Section 2.

While the above extends Shearer's bound to graphs with few triangles, it is also natural to ask how Alon's second bound, the surplus  $\Omega(d^2)$ , is affected when we allow triangles. We answer this question, proving that if  $G$  has less triangles than a random graph of the same average degree, it has large surplus.

**Theorem 1.2.** *Let  $G$  be a graph with average degree  $d$  and less than  $(1 - \epsilon)d^3/6$  triangles, for some  $\epsilon > 0$ . Then  $G$  has surplus  $\Omega_\epsilon(d^2)$ .*

In the case of regular graphs, we obtain even a much sharper result, which expresses the surplus as a function of the order, degree, and number of triangles of the graph.

**Theorem 1.3.** *Let  $G$  be a  $d$ -regular graph with  $n$  vertices and  $d^3/6 + s$  triangles, where  $d \leq n/2$ .*

- If  $s < -nd^{3/2}$ , then  $G$  has surplus  $\Omega(\frac{|s|}{d})$ .
- If  $-nd^{3/2} \leq s \leq nd^{3/2}$ , then  $G$  has surplus  $\Omega(nd^{1/2})$ .
- If  $s > nd^{3/2}$ , then  $G$  has surplus  $\Omega(\frac{n^2 d^2}{s})$ .

The bounds given here are tight for all strongly regular graphs that attain the respective parameters (see Theorem 2.4 for the exact statement). Also, after a simple modification of the proof of

Theorem 1.3, the condition  $d \leq n/2$  can be replaced by  $d \leq (1 - \epsilon)n$  for any positive constant  $\epsilon$ . We remark that the conclusion of Theorem 1.3 is no longer true if we only assume that the average degree of  $G$  is  $d$ , rather than that it is  $d$ -regular. To see this, take for example  $G$  to be the graph which is the union of a random graph with  $n^{4/5}$  vertices and density  $n^{-2/5}$  and an empty graph on  $n - n^{4/5}$  vertices. It is not hard to see that with high probability,  $G$  has average degree  $d = \Theta(n^{1/5})$ , the number of triangles in  $G$  is  $t(G) = \Theta(n^{6/5})$ , but the surplus is only  $\Theta(n)$  rather than  $\Omega(nd^{1/2})$ .

### 1.3 Cliques

Recall the conjecture of Alon, Bollobás, Krivelevich and Sudakov [3] that for every graph  $H$ , there exist constants  $\epsilon = \epsilon(H) > 0$  and  $c = c(H) > 0$  such that  $\text{sp}(m, H) \geq cm^{3/4+\epsilon}$  for all  $m$ . Clearly, it suffices to prove this conjecture for cliques. Carlson et al. [6] speculated that the following analogue of Shearer's bound should hold for  $K_r$ -free graphs, and proved that this would imply the conjecture of Alon, Bollobás, Krivelevich and Sudakov.

**Conjecture 1.4** ([6]). For every fixed  $r \geq 3$ , any  $K_r$ -free  $d$ -degenerate graph with  $m$  edges has surplus  $\Omega_r(m/\sqrt{d})$ .

They also proved a weakening of this conjecture, namely the lower bound  $\Omega_r(m/d^{1-1/(2r-4)})$ . We improve this bound slightly. It would be very interesting to obtain a version of this result where the exponent of  $d$  is bounded by a constant smaller than 1 uniformly for all  $r$ .

**Theorem 1.5.** For every fixed  $r \geq 3$ , any  $K_r$ -free  $d$ -degenerate graph with  $m$  edges has surplus  $\Omega_r(m/d^{1-1/(r-1)})$ .

As discussed in Section 1.1, essentially all known extremal constructions for MaxCut in  $H$ -free graphs are dense pseudorandom graphs. Motivated by this observation, one can even make a more precise guess for  $\text{sp}(m, K_r)$ . However, the construction of dense pseudorandom  $K_r$ -free graphs is a major open problem in its own right. Sudakov, Szabó and Vu [23] raised the question whether, for fixed  $r \geq 3$ , there exist  $K_r$ -free  $(n, d, \lambda)$ -graphs with  $d = \Omega(n^{1-1/(2r-3)})$  and  $\lambda = O(\sqrt{d})$ . It has been proved that this would have striking consequences for off-diagonal Ramsey numbers [18]. Speculating that such graphs do exist, this would via (1.4) imply that

$$\text{sp}(m, K_r) = O\left(m^{1-\frac{1}{4+1/(r-2)}}\right). \quad (1.5)$$

Perhaps this gives the correct answer. Note that Conjecture 1.4 would provide the matching lower bound as long as  $d = O(n^{1-1/(2r-3)})$ , that is, in the ‘‘sparse’’ case, when the density is smaller than that of the speculative extremal examples. We prove a result that complements this in the ‘‘dense’’ regime, analogous to Alon's  $\Omega(d^2)$  surplus for triangle-free graphs.

**Theorem 1.6.** For every fixed  $r \geq 3$ , any  $K_r$ -free  $n$ -vertex graph with average degree  $d$  has surplus  $\Omega_r(d^{r-1}/n^{r-3})$ .

For  $d = \Omega(n^{1-1/(2r-3)})$ , the implied surplus is at least  $\Omega_r\left((nd)^{1-\frac{1}{4+1/(r-2)}}\right)$ . Hence, if Conjecture 1.4 is true, then combining it with Theorem 1.6 similar as in Alon's result, we could deduce that indeed  $\text{sp}(m, K_r) = \Omega_r\left(m^{1-\frac{1}{4+1/(r-2)}}\right)$ , matching the speculative upper bound in (1.5). Using Theorem 1.5 instead of Conjecture 1.4, we still improve the current best known lower bound  $\text{sp}(m, K_r) = \Omega_r(m^{\frac{1}{2}+\frac{1}{4r-8}})$ , due to Carlson et al. [6].

**Theorem 1.7.** For every  $r \geq 3$ , any  $K_r$ -free graph with  $m$  edges has surplus  $\Omega_r(m^{\frac{1}{2}+\frac{3}{4r-2}})$ .

In fact, we will establish Theorem 1.6 in the stronger form that it is sufficient that the number of  $r$ -cliques is smaller than in a random graph of the same density. This generalizes Theorem 1.2.

**Theorem 1.8.** For any integer  $r \geq 3$  and any  $\epsilon > 0$ , there exists a positive constant  $c = c(r, \epsilon)$  such that the following holds. If  $G$  is a graph on  $n$  vertices with average degree  $d$  which has at most  $(1 - \epsilon)\frac{n^r}{r!}(d/n)^{\binom{r}{2}}$  copies of  $K_r$ , then  $G$  has surplus at least  $cd^{r-1}/n^{r-3}$ .

Finally, we remark that Theorems 1.2, 1.3 and 1.6 generalize results of Nikiforov [19, 20]. In [19], he proved that the smallest eigenvalue of a  $K_r$ -free  $n$ -vertex graph with average degree  $d$  has absolute value  $\Omega_r(d^{r-1}/n^{r-2})$ . This also follows from Theorem 1.6 and (1.4). However, a lower bound on MaxCut is stronger than a bound on  $\lambda_{\min}$  since the converse of (1.4) is not true. Theorem 1.8 shows that the conclusion of Nikiforov’s theorem remains true even if we only require that there are “few” copies of  $K_r$ . Nikiforov also proved [20] that an  $n$ -vertex graph with average degree  $d$  and at most  $(1 - \epsilon)d^3/6$  triangles has  $|\lambda_{\min}| = \Omega(\epsilon d^2/n)$  provided that  $d/n$  is bounded away from 1. This follows from our Theorem 1.2 when  $\epsilon > 0$  is constant. For an arbitrary positive  $\epsilon$  (possibly tending to 0), his result follows from our Theorem 1.3 provided that the graph is regular.

**Notation:** We use standard notation. All graphs considered are finite, simple and undirected. We write  $|G|$  for the order of  $G$ ,  $e(G)$  for the number of edges, and  $t(G)$  for the number of triangles in  $G$ . The minimum and maximum degree of  $G$  are denoted by  $\delta(G)$  and  $\Delta(G)$ , respectively. For a set  $S \subset V(G)$ , we write  $e_G(S)$  for the number of edges in  $G[S]$ , and for disjoint sets  $S, T \subset V(G)$ , we write  $e_G(S, T)$  for the number of edges in  $G$  with one endpoint in  $S$  and one in  $T$ .  $N_G(v)$  and  $d_G(v)$  stand for the neighbourhood and degree of a vertex  $v$ . The number of common neighbours of vertices  $v_1, \dots, v_r$  is denoted  $d_G(v_1, \dots, v_r)$ . We occasionally omit subscripts if they are clear from the context. We let  $P_n$  and  $C_n$  denote the path and cycle, respectively, with  $n$  edges.

When using Landau symbols,  $\Omega_r(\cdot)$ ,  $\Theta_r(\cdot)$ ,  $O_r(\cdot)$  mean that the implicit constants may depend on  $r$ . We denote by  $\|\cdot\|$  the Euclidean norm. As customary, we tacitly treat large numbers like integers whenever this has no effect on the argument.

The rest of this paper is organized as follows. In Section 2, we introduce some of the main tools utilised in our proofs, and use them to prove a few of our less technical results. In particular, we first prove Theorem 1.3 in the special case where  $G$  is strongly regular. Also, we present the proof of Theorem 1.2. We conclude the section with the proof of Theorem 1.1 for the simplest new case  $C_5$  (the case  $r = 3$  is already covered by Alon’s theorem) under the assumption that  $G$  is regular. In Section 3, we prove Theorem 1.1 in the general case. We give the full proof of Theorem 1.3 in Section 4. Finally, we prove Theorems 1.5, 1.7 and 1.8 in Section 5.

## 2 Main ideas and first applications

### 2.1 Preliminaries

We start with a well known observation that we will repeatedly use in our proofs. In order to show that a graph has large surplus, it suffices to find many disjoint induced subgraphs with relatively large surplus.

**Lemma 2.1.** *Let  $G$  be a graph and let  $S_1, S_2, \dots, S_k$  be pairwise disjoint subsets of  $V(G)$ . Then  $\text{sp}(G) \geq \sum_{i=1}^k \text{sp}(G[S_i])$ .*

Variants of this have been used in many preceding papers (see e.g. [5, Lemma 3.2]). The proof is similar to the standard  $m/2$  bound for  $\text{mc}(G)$ . Note that by adding singletons, one can assume that the  $S_i$ ’s partition  $V(G)$ . Take an optimal cut of each of the induced subgraphs, and form a cut of  $G$  by putting the sets of the smaller cuts randomly on either side. The result then follows by inspecting the expected size of the obtained cut. We omit the details.

Note that in the simplest case, the lemma says that the surplus of a graph is at least as large as the surplus of any of its induced subgraphs. We will use this fact naturally in our proofs without necessarily referring to the lemma.

### 2.2 Semidefinite programming

Semidefinite programming in the algorithmic context of MaxCut was first used by Goemans and Williamson [15]. The method was exploited by Carlson et al. [6] to give lower bounds on the MaxCut for  $H$ -free graphs. The approach can be summarised as follows. Given a graph  $G$ , we try to assign a unit vector  $\mathbf{x}^v \in \mathbb{R}^N$  to each vertex  $v \in V(G)$  in a way that the inner products  $\langle \mathbf{x}^u, \mathbf{x}^v \rangle$  are negative

on average over all adjacent pairs of vertices  $u, v$ . Then we take a random unit vector  $\mathbf{z} \in \mathbb{R}^N$  and define  $A = \{v \in V(G) : \langle \mathbf{x}^v, \mathbf{z} \rangle \geq 0\}$ ,  $B = V(G) \setminus A$ . Consider the random cut  $(A, B)$ . If vertices  $u$  and  $v$  have  $\langle \mathbf{x}^u, \mathbf{x}^v \rangle < 0$ , then with probability more than  $1/2$ ,  $u$  and  $v$  will end up in different parts of the cut. Hence, intuitively, we expect our cut to contain more than half of the edges. This is made precise by the following lemma, essentially due to Goemans and Williamson [15].

**Lemma 2.2.** *Let  $G$  be a graph and let  $N$  be a positive integer. Then, for any set of non-zero vectors  $\{\mathbf{x}^v : v \in V(G)\} \subset \mathbb{R}^N$ , we have  $\text{sp}(G) \geq -\frac{1}{\pi} \sum_{uv \in E(G)} \arcsin\left(\frac{\langle \mathbf{x}^u, \mathbf{x}^v \rangle}{\|\mathbf{x}^u\| \|\mathbf{x}^v\|}\right)$ .*

**Proof.** Let  $\mathbf{z}$  be a uniformly random unit vector in  $\mathbb{R}^N$ , let  $A = \{v \in V(G) : \langle \mathbf{x}^v, \mathbf{z} \rangle \geq 0\}$  and let  $B = V(G) \setminus A$ . The probability that  $uv$  belongs to the cut  $(A, B)$  is  $\frac{\alpha}{\pi}$ , where  $\alpha = \arccos\left(\frac{\langle \mathbf{x}^u, \mathbf{x}^v \rangle}{\|\mathbf{x}^u\| \|\mathbf{x}^v\|}\right)$  is the angle between  $\mathbf{x}^u$  and  $\mathbf{x}^v$ . Hence,

$$\mathbb{P}(uv \text{ belongs to the cut}) = \frac{1}{\pi} \arccos\left(\frac{\langle \mathbf{x}^u, \mathbf{x}^v \rangle}{\|\mathbf{x}^u\| \|\mathbf{x}^v\|}\right) = \frac{1}{2} - \frac{1}{\pi} \arcsin\left(\frac{\langle \mathbf{x}^u, \mathbf{x}^v \rangle}{\|\mathbf{x}^u\| \|\mathbf{x}^v\|}\right).$$

Summing the equality over all edges  $uv \in E(G)$  shows that the expected number of edges in the cut  $(A, B)$  is  $\frac{e(G)}{2} - \frac{1}{\pi} \sum_{uv \in E(G)} \arcsin\left(\frac{\langle \mathbf{x}^u, \mathbf{x}^v \rangle}{\|\mathbf{x}^u\| \|\mathbf{x}^v\|}\right)$ , from which the result follows.  $\square$

In some cases, we can use a more convenient form of the above, which is implicit in the work of Carlson et al. [6].

**Corollary 2.3.** *Let  $G$  be a graph and let  $N$  be a positive integer. Let  $\{\mathbf{x}^v : v \in V(G)\} \subset \mathbb{R}^N$  be vectors with  $\|\mathbf{x}^v\| = O(1)$ , and suppose that for every edge  $uv \in E(G)$ , we have  $\langle \mathbf{x}^u, \mathbf{x}^v \rangle \leq -a_{uv} + b_{uv}$  for some  $a_{uv}, b_{uv} \geq 0$ . Then*

$$\text{sp}(G) \geq \Omega\left(\sum_{uv \in E(G)} a_{uv}\right) - O\left(\sum_{uv \in E(G)} b_{uv}\right).$$

**Proof.** We add extra coordinates, one for each vertex  $v \in V(G)$ . The vector  $\mathbf{y}^v$  is obtained from  $\mathbf{x}^v$  by setting the new coordinate for  $v$  to 1 and all other new coordinates to 0. Hence, we have  $\|\mathbf{y}^v\|^2 = \|\mathbf{x}^v\|^2 + 1 = \Theta(1)$  for all  $v \in V(G)$ , and  $\langle \mathbf{y}^u, \mathbf{y}^v \rangle = \langle \mathbf{x}^u, \mathbf{x}^v \rangle \leq -a_{uv} + b_{uv}$  for all  $uv \in E(G)$ .

For any  $x \in [-1, 1]$  with  $x \leq b - a$  for some  $a, b \geq 0$ , we have  $\arcsin(x) \leq \frac{\pi}{2}b - a$ . This is because when  $x < 0$ , then  $\arcsin(x) \leq x \leq b - a \leq \frac{\pi}{2}b - a$ , and if  $x \geq 0$ , then  $\arcsin(x) \leq \frac{\pi}{2}x \leq \frac{\pi}{2}(b - a) \leq \frac{\pi}{2}b - a$ . Hence, for every edge  $uv \in E(G)$ , we have

$$\arcsin\left(\frac{\langle \mathbf{y}^u, \mathbf{y}^v \rangle}{\|\mathbf{y}^u\| \|\mathbf{y}^v\|}\right) \leq \frac{\pi}{2} \frac{b_{uv}}{\|\mathbf{y}^u\| \|\mathbf{y}^v\|} - \frac{a_{uv}}{\|\mathbf{y}^u\| \|\mathbf{y}^v\|} = -\Omega(a_{uv}) + O(b_{uv}).$$

Summing over all edges and using Lemma 2.2 gives the result.  $\square$

### 2.3 Surplus in strongly regular graphs

As a first illustration of the use of semidefinite programming, we determine the order of magnitude of the surplus in any strongly regular graph. Recall that a graph is called *strongly regular* if in addition to being regular, there exist  $\eta, \mu$  such that any two adjacent vertices have exactly  $\eta$  common neighbours and any two distinct non-adjacent vertices have exactly  $\mu$  common neighbours.

In Section 4, we will use a more involved argument to generalize the lower bound to every regular graph, proving Theorem 1.3.

**Theorem 2.4.** *Let  $G$  be a strongly regular graph with  $n$  vertices, degree  $d$  and  $d^3/6 + s$  triangles, where  $d \leq 0.99n$ .*

- If  $s < -nd^{3/2}$ , then  $\text{sp}(G) = \Theta\left(\frac{|s|}{d}\right)$ .
- If  $-nd^{3/2} \leq s \leq nd^{3/2}$ , then  $\text{sp}(G) = \Theta(nd^{1/2})$ .

- If  $s > nd^{3/2}$ , then  $\text{sp}(G) = \Theta(\frac{n^2 d^2}{s})$ .

**Proof.** Let us start with the lower bounds. Let  $a = \frac{\sqrt{d}}{n-d}$  and let  $0 < \gamma \leq 1$ . For every  $v \in V(G)$ , we define a vector  $\mathbf{x}^v$  whose coordinates are labelled by  $V(G)$ , i.e.  $\mathbf{x}^v \in \mathbb{R}^{V(G)}$ . Take

$$\mathbf{x}_u^v = \begin{cases} 1 + \gamma a & \text{if } u = v, \\ -\gamma \frac{1}{\sqrt{d}} & \text{if } u \in N(v), \\ \gamma a & \text{otherwise.} \end{cases}$$

Then for any  $uv \in E(G)$ ,

$$\begin{aligned} \langle \mathbf{x}^u, \mathbf{x}^v \rangle &= -2(1 + \gamma a)\gamma \frac{1}{\sqrt{d}} + |N(u) \cap N(v)|\gamma^2 \frac{1}{d} + |V(G) \setminus (N(u) \cup N(v))|\gamma^2 a^2 \\ &\quad + \left( |N(u) \setminus (N(v) \cup \{v\})| + |N(v) \setminus (N(u) \cup \{u\})| \right) \left( -\gamma^2 \frac{a}{\sqrt{d}} \right). \end{aligned}$$

Observe that  $|N(u) \cap N(v)| = d(u, v)$ ,  $|V(G) \setminus (N(u) \cup N(v))| = n - 2d + d(u, v)$  and  $|N(u) \setminus (N(v) \cup \{v\})| = |N(v) \setminus (N(u) \cup \{u\})| = d - d(u, v) - 1$ , so

$$\langle \mathbf{x}^u, \mathbf{x}^v \rangle = -2\gamma \frac{1}{\sqrt{d}} + \gamma^2 \left( \left( \frac{1}{d} + 2\frac{a}{\sqrt{d}} + a^2 \right) d(u, v) + (n - 2d)a^2 - 2d\frac{a}{\sqrt{d}} \right).$$

Let  $\delta(u, v) = d(u, v) - \frac{d^2}{n}$ . Note that

$$\left( \frac{1}{d} + 2\frac{a}{\sqrt{d}} + a^2 \right) \frac{d^2}{n} + (n - 2d)a^2 - 2d\frac{a}{\sqrt{d}} = \frac{1}{n}(d - 2a\sqrt{d}(n - d) + a^2(n - d)^2) = \frac{1}{n}(d - 2d + d) = 0,$$

so

$$\langle \mathbf{x}^u, \mathbf{x}^v \rangle = -2\gamma \frac{1}{\sqrt{d}} + \gamma^2 \left( \frac{1}{d} + 2\frac{a}{\sqrt{d}} + a^2 \right) \delta(u, v). \quad (2.1)$$

Note that so far we have not needed the assumption that  $G$  is strongly regular, only that it is regular. We will in fact use (2.1) also in the proof of Theorem 1.3.

Since  $G$  is strongly regular, every pair of adjacent vertices has the same number  $\eta$  of common neighbours. Note that  $e(G)\eta = 3t(G)$ , so

$$\eta = \frac{3t(G)}{e(G)} = \frac{3(d^3/6 + s)}{nd/2} = \frac{d^2}{n} + \frac{6s}{nd}, \quad (2.2)$$

hence  $\delta(u, v) = \frac{6s}{nd}$  for every  $uv \in E(G)$ . Thus,

$$\langle \mathbf{x}^u, \mathbf{x}^v \rangle = -2\gamma \frac{1}{\sqrt{d}} + \gamma^2 \left( \frac{1}{d} + 2\frac{a}{\sqrt{d}} + a^2 \right) \frac{6s}{nd}.$$

Observe that since  $d \leq 0.99n$ , we have  $a \leq \frac{100}{\sqrt{d}}$ , so  $\frac{1}{d} \leq \frac{1}{d} + 2\frac{a}{\sqrt{d}} + a^2 \leq \frac{10^5}{d}$ .

If  $s < -nd^{3/2}$ , then take  $\gamma = 1$  to get

$$\langle \mathbf{x}^u, \mathbf{x}^v \rangle \leq \frac{1}{d} \cdot \frac{6s}{nd} = \frac{6s}{nd^2}.$$

If  $-nd^{3/2} \leq s \leq nd^{3/2}$ , take  $\gamma = 1/10^6$  and note that  $\gamma^2 \left( \frac{1}{d} + 2\frac{a}{\sqrt{d}} + a^2 \right) \frac{6s}{nd} \leq \gamma \frac{1}{\sqrt{d}}$ , so

$$\langle \mathbf{x}^u, \mathbf{x}^v \rangle \leq -\gamma \frac{1}{\sqrt{d}} = -\frac{1}{10^6 \sqrt{d}}.$$

Finally, if  $s > nd^{3/2}$ , take  $\gamma = \frac{nd^{3/2}}{10^6 s}$  and note that  $\gamma^2 \left( \frac{1}{d} + 2\frac{a}{\sqrt{d}} + a^2 \right) \frac{6s}{nd} \leq \gamma \frac{1}{\sqrt{d}}$ , so

$$\langle \mathbf{x}^u, \mathbf{x}^v \rangle \leq -\gamma \frac{1}{\sqrt{d}} = -\frac{nd}{10^6 s}.$$

In all three cases,  $\langle \mathbf{x}^u, \mathbf{x}^v \rangle < 0$ , and moreover it is easy to see that  $\|\mathbf{x}^v\| = O(1)$  for every  $v \in V(G)$ , so by Corollary 2.3, we have

$$\text{sp}(G) \geq \Omega \left( \sum_{uv \in E(G)} -\langle \mathbf{x}^u, \mathbf{x}^v \rangle \right).$$

Summing the respective upper bounds for  $\langle \mathbf{x}^u, \mathbf{x}^v \rangle$  in the three cases, we obtain the desired lower bounds for  $\text{sp}(G)$ .

It remains to prove the upper bounds for  $\text{sp}(G)$ . For this, we will use the eigenvalue bound (1.4). It is well known (see [16], for example) that the smallest eigenvalue of a strongly regular graph is

$$\lambda_{\min} = \frac{1}{2} \left( \eta - \mu - \sqrt{(\eta - \mu)^2 + 4(d - \mu)} \right),$$

where  $\eta$  is the number of common neighbours of adjacent pairs and  $\mu$  is the number of common neighbours of non-adjacent pairs. It is a straightforward (but somewhat tedious) exercise to check that this, using (2.2), implies the desired bounds. For the details, see Lemma A.1 in the appendix.  $\square$

## 2.4 Finding good almost regular induced subgraphs

In this subsection we establish an important lemma that allows us to reduce many of our statements to the special case where  $G$  is almost regular. For example, in the next subsection we already use it to study the surplus of graphs with few triangles. We believe that this lemma will be a useful tool for future works on this subject as well.

**Lemma 2.5.** *Let  $\alpha$  and  $\beta$  be real numbers such that  $\beta > 0$ ,  $\alpha < \beta$  and  $\alpha + \beta \leq 2$ . Then there exist positive constants  $c_1$ ,  $c_2$  and  $C$  (depending on  $\alpha, \beta$ ), such that the following holds.*

*Let  $G$  be a graph on  $n$  vertices with average degree  $d$ . Then either  $G$  has surplus at least  $c_1 n^\alpha d^\beta$  or  $G$  has an induced subgraph  $\tilde{G}$  with  $\tilde{n}$  vertices and average degree  $\tilde{d}$  where  $\tilde{n}^\alpha \tilde{d}^\beta \geq c_2 n^\alpha d^\beta$ ,  $\delta(\tilde{G}) \geq \tilde{d}/C$  and  $\Delta(\tilde{G}) \leq C\tilde{d}$ .*

A typical use of this lemma will be as follows. Suppose that we want to prove that an  $n$ -vertex  $H$ -free graph with average degree  $d$  has surplus  $\Omega(n^\alpha d^\beta)$ . Applying Lemma 2.5, we can pass to the induced subgraph  $\tilde{G}$  if necessary, so it suffices to prove the bound  $\Omega(n^\alpha d^\beta)$  under the assumption that  $\delta(G) \geq d/C$  and  $\Delta(G) \leq Cd$ .

For a few of our results, we will need the following somewhat stronger version of Lemma 2.5. Similarly to the previous lemma, one could also impose an additional minimum degree condition  $\delta(\tilde{G}) \geq \tilde{d}/C$ , but we will not need that.

**Lemma 2.6.** *Let  $\epsilon$ ,  $\alpha$  and  $\beta$  be real numbers such that  $\epsilon, \beta > 0$ ,  $\alpha < \beta$  and  $\alpha + \beta \leq 2$ . Then there exist positive constants  $c$  and  $C$  (depending on  $\epsilon, \alpha, \beta$ ), such that the following holds.*

*Let  $G$  be a graph on  $n$  vertices with average degree  $d$ . Then either  $G$  has surplus at least  $cn^\alpha d^\beta$  or  $G$  has an induced subgraph  $\tilde{G}$  with  $\tilde{n}$  vertices and average degree  $\tilde{d}$  where  $\tilde{n}^\alpha \tilde{d}^\beta \geq (1 - \epsilon)n^\alpha d^\beta$  and  $\Delta(\tilde{G}) \leq C\tilde{d}$ .*

**Proof.** Choose  $\theta$  such that  $(1 - \theta)^\beta = 1 - \epsilon$ . Let  $c = \frac{\theta^2}{320}$ , let  $C_0$  be chosen so that  $(\theta^2/20)^\beta C_0^{\beta-\alpha} = 1$  and let  $C = C_0/(1 - \theta)$ .

*Claim 1.* *If  $G$  is a graph with  $n$  vertices,  $m$  edges and average degree  $d$ , then at least one of the following must hold.*

- (i)  $G$  has an induced subgraph  $G'$  with  $n' < n$  vertices and average degree  $d'$ , where  $(n')^\alpha (d')^\beta \geq n^\alpha d^\beta$ .
- (ii)  $G$  has surplus at least  $\frac{\theta^2}{160}m$ .
- (iii)  $G$  has an induced subgraph  $\tilde{G}$  with  $\tilde{n}$  vertices and average degree  $\tilde{d}$ , where  $\tilde{n}^\alpha \tilde{d}^\beta \geq (1 - \epsilon)n^\alpha d^\beta$  and  $\Delta(\tilde{G}) \leq C\tilde{d}$ .

Before proving the claim, let us see how the lemma follows. Given a graph  $G$ , apply the claim repeatedly for the obtained induced subgraphs  $G'$ ,  $(G')'$ , etc. as long as (i) holds. Eventually, (ii) or (iii) must hold, so  $G$  has an induced subgraph  $F$  with  $N$  vertices and average degree  $D$  such that  $N^\alpha D^\beta \geq n^\alpha d^\beta$ , and either  $F$  has surplus at least  $\frac{\theta^2}{320}ND$ , or  $F$  has an induced subgraph  $\tilde{G}$  with  $\tilde{n}$  vertices and average degree  $\tilde{d}$ , where  $\tilde{n}^\alpha \tilde{d}^\beta \geq (1 - \epsilon)N^\alpha D^\beta$  and  $\Delta(\tilde{G}) \leq C\tilde{d}$ . In the latter case, we are done. In the former case, note that since  $\alpha + \beta \leq 2$ ,  $\alpha < \beta$  and  $N > D$ , we have  $ND \geq (ND)^{\frac{1}{2}(\alpha+\beta)} \geq N^\alpha D^\beta$ , so  $F$  (and, consequently,  $G$ ) has surplus at least  $\frac{\theta^2}{320}N^\alpha D^\beta \geq \frac{\theta^2}{320}n^\alpha d^\beta = cn^\alpha d^\beta$ . Thus, it suffices to prove the claim.

*Proof of claim:* Let  $S$  be the set of vertices in  $G$  with degree at most  $C_0d$  and let  $T = V(G) \setminus S$ . We consider the following three cases (observe that at least one of these must occur).

**Case 1.**  $e_G(T) \geq \frac{\theta^2}{20}m$ .

Let  $G' = G[T]$  and let  $n'$  and  $d'$  be the number of vertices and average degree of  $G'$ , respectively. Then  $n'd' = 2e(G') \geq \theta^2 m/10 = \theta^2 nd/20$ , so  $\frac{d'}{d} \geq \frac{\theta^2 n}{20 n'}$  and hence

$$\frac{(n')^\alpha (d')^\beta}{n^\alpha d^\beta} \geq \left(\frac{n'}{n}\right)^\alpha \left(\frac{\theta^2 n}{20 n'}\right)^\beta = (\theta^2/20)^\beta \left(\frac{n}{n'}\right)^{\beta-\alpha}.$$

Note that  $G$  has average degree  $d$  but every vertex in  $T$  has degree at least  $C_0d$  in  $G$ , so  $n' = |T| \leq n/C_0$ . Hence,

$$(\theta^2/20)^\beta \left(\frac{n}{n'}\right)^{\beta-\alpha} \geq (\theta^2/20)^\beta C_0^{\beta-\alpha} = 1.$$

Thus,  $G'$  satisfies (i).

**Case 2.**  $e_G(T) < \frac{\theta^2}{20}m$ , but  $e_G(S, T) \geq \frac{\theta}{2}m$ .

Let  $S'$  be a random subset of  $S$  obtained by keeping each vertex of  $S$  with probability  $\theta/4$ . Then  $\mathbb{E}[e_G(S')] \leq \frac{\theta^2}{16}m$  and  $\mathbb{E}[e_G(S', T)] \geq \frac{\theta^2}{8}m$ . Using  $e_G(T) \leq \frac{\theta^2}{20}m$ , it follows that there exists a cut in  $G$  with surplus at least  $\frac{1}{2}(\frac{\theta^2}{8}m - \frac{\theta^2}{16}m - \frac{\theta^2}{20}m) = \frac{\theta^2}{160}m$ , so (ii) holds.

**Case 3.**  $e_G(S) \geq (1 - \theta)m$ .

Let  $\tilde{G} = G[S]$  and let  $\tilde{n}$  and  $\tilde{d}$  be the number of vertices and average degree of  $\tilde{G}$ , respectively. Clearly,  $\tilde{d} \geq (1 - \theta)d$ , so  $\Delta(\tilde{G}) \leq C_0d = (1 - \theta)C_0d \leq C\tilde{d}$ . Also,  $\tilde{n}\tilde{d} \geq (1 - \theta)nd$ , so

$$\frac{\tilde{n}^\alpha \tilde{d}^\beta}{n^\alpha d^\beta} \geq \left(\frac{\tilde{n}}{n}\right)^\alpha \left((1 - \theta)\frac{n}{\tilde{n}}\right)^\beta = (1 - \epsilon) \left(\frac{n}{\tilde{n}}\right)^{\beta-\alpha} \geq 1 - \epsilon.$$

Hence,  $\tilde{G}$  satisfies (iii). □

It is not hard to deduce Lemma 2.5.

**Proof of Lemma 2.5.** By Lemma 2.6, applied with  $\epsilon = 1/2$ , there exist positive constants  $c_1$  and  $C'$  such that either  $G$  has surplus at least  $c_1 n^\alpha d^\beta$  or  $G$  has an induced subgraph  $G'$  with  $n'$  vertices and average degree  $d'$  where  $(n')^\alpha (d')^\beta \geq \frac{1}{2}n^\alpha d^\beta$  and  $\Delta(G') \leq C'd'$ . In the former case we are done. In the latter case, note that  $G'$  has an induced subgraph  $\tilde{G}$  with at least  $e(G')/2$  edges and minimum degree at least  $d'/4$ . Then, writing  $\tilde{n}$  for the number of vertices and  $\tilde{d}$  for the average degree of  $\tilde{G}$ , we have that  $\tilde{d} \geq d'/2$  and  $\tilde{n}\tilde{d} \geq n'd'/2$ . Therefore  $\tilde{n}^\alpha \tilde{d}^\beta \geq 2^{-\beta}(n')^\alpha (d')^\beta$ , so we can choose  $c_2 = 2^{-\beta-1}$  and  $C = 4C'$ . □

## 2.5 Fewer triangles than in the random graph

In this subsection, we prove Theorem 1.2. Using Lemma 2.6 (as we show in the end of the subsection), it suffices to consider graphs whose maximum degree is not much larger than the average degree. Under this assumption, we prove the following slightly stronger result, which we will use to prove Theorem 1.8.

**Theorem 2.7.** *Let  $G$  be a graph with  $n$  vertices, average degree  $d$ , maximum degree at most  $Cd$  and less than  $(1 - \epsilon) \frac{d}{6n} \sum_{u \in V(G)} d(u)^2$  triangles, for some  $\epsilon > 0$ . Then  $G$  has surplus  $\Omega(\frac{\epsilon^2}{C} d^2)$ .*

**Proof.** Let  $\mu = \frac{\epsilon}{4C}$  and  $k = \lceil \mu n / d \rceil$ .

First, we give every vertex a random label from  $\{0, 1, \dots, k+1\}$ , where 0 is chosen with probability  $1/3$ , each of  $1, \dots, k$  is chosen with probability  $d/3n$ , and the remaining probability falls on  $k+1$ . This is feasible since  $k \cdot \frac{d}{3n} \leq (\mu n / d + 1) \frac{d}{3n} \leq \frac{\mu}{3} + \frac{d}{3n} \leq 2/3$ . For every  $j \in [k]$ , let  $B_j$  be the set of vertices with label  $j$ .

Now, pick uniformly at random with repetition vertices  $v_1, \dots, v_k$ . For  $j \in [k]$ , let  $A_j$  be the set of neighbours of  $v_j$  with label 0 which are not neighbours of any other  $v_i$ .

Let  $X$  be the number of edges which go between  $A_j$  and  $B_j$  for some  $j \in [k]$  (that is,  $X$  is the sum of the number of edges between  $A_j$  and  $B_j$  over all  $j \in [k]$ ), let  $Y$  be the number of edges inside some  $A_j$  and let  $Z$  be the number of edges inside some  $B_j$ . Our goal will be to show that  $\mathbb{E}[X - Y - Z] = \Omega(\frac{\epsilon^2}{C} d^2)$ . This is achieved in three separate claims.

*Claim 1.*  $\mathbb{E}[Y] \leq (1 - \epsilon) \frac{kd}{18n^2} \sum_{u \in V(G)} d(u)^2$ .

*Proof of claim:* Consider an edge  $uv$ . Note that  $uv$  is only going to be an internal edge of some  $A_j$  if some common neighbour is among the  $v_j$ 's. The probability of this is at most  $k \frac{d(u,v)}{n}$ . Moreover, both  $u$  and  $v$  need to be labelled 0. Now, summing over all edges, we see that  $\mathbb{E}[Y] \leq \frac{k}{9n} \sum_{uv \in E(G)} d(u,v)$ . The sum counts every triangle exactly three times. Using the assumption, we get the claim.  $\diamond$

*Claim 2.*  $\mathbb{E}[Z] \leq \frac{kd^3}{18n}$ .

*Proof of claim:* Consider an edge  $uv$ . The probability that it is contained in  $B_j$  for some fixed  $j$  is  $(d/3n)^2$ . Multiplying this with  $k$  for the number of possibilities for  $j$  and  $nd/2$  for the total number of edges yields the claim.  $\diamond$

*Claim 3.*  $\mathbb{E}[X] \geq (1 - \epsilon/4) \frac{kd}{9n^2} \sum_{u \in V(G)} d(u)^2$ .

*Proof of claim:* Consider an edge  $uv$ . With probability  $kd/9n$ ,  $u$  is labelled 0 and  $v$  gets a label from  $[k]$ . Condition on such a label  $j \in [k]$ . Now, the probability that  $v_j$  is a neighbour of  $u$ , but no other  $v_i$  is a neighbour of  $u$ , is  $\frac{d(u)}{n} (1 - d(u)/n)^{k-1} \geq \frac{d(u)}{n} (1 - (k-1)d(u)/n) \geq \frac{d(u)}{n} (1 - (k-1)Cd/n) \geq \frac{d(u)}{n} (1 - C\mu) = \frac{d(u)}{n} (1 - \epsilon/4)$ . The same applies with the roles of  $u$  and  $v$  swapped. Hence, the probability of  $uv$  contributing to  $X$  is at least  $\frac{kd}{9n^2} (d(u) + d(v)) (1 - \epsilon/4)$ . Summing over all edges, we obtain the desired inequality.  $\diamond$

By convexity,  $d^2 \leq \frac{1}{n} \sum_{u \in V(G)} d(u)^2$ , so combining the three claims, we get

$$\mathbb{E}[X - Y - Z] \geq \frac{kd}{9n^2} \left( \sum_{u \in V(G)} d(u)^2 \right) ((1 - \epsilon/4) - (1 - \epsilon)/2 - 1/2) = \frac{\epsilon kd}{36n^2} \sum_{u \in V(G)} d(u)^2 \geq \frac{\epsilon \mu}{36} d^2.$$

We infer that there exists an outcome for which  $X - Y - Z \geq \Omega(\epsilon \mu d^2)$ . By Lemma 2.1, we conclude that

$$\text{sp}(G) \geq \sum_{j=1}^k \text{sp}(G[A_j \cup B_j]) \geq \frac{1}{2} \sum_{j=1}^k (e(A_j, B_j) - e(A_j) - e(B_j)) = \frac{1}{2} (X - Y - Z) \geq \Omega(\epsilon \mu d^2),$$

which completes the proof by using  $\mu = \frac{\epsilon}{4C}$ .  $\square$

It is easy to deduce Theorem 1.2.

**Proof of Theorem 1.2.** By Lemma 2.6 (with  $\epsilon/3$  in place of  $\epsilon$ ,  $\alpha = 0$  and  $\beta = 2$ ), either  $G$  has surplus  $\Omega_\epsilon(d^2)$ , or it has an induced subgraph  $\tilde{G}$  with  $\tilde{n}$  vertices and average degree  $\tilde{d}$ , where  $\tilde{d}^2 \geq (1 - \epsilon/3)d^2$  and  $\Delta(\tilde{G}) \leq C\tilde{d}$  for some  $C = C(\epsilon)$ . In the latter case, we have  $\tilde{d}^3 \geq (1 - \epsilon/3)^{3/2} d^3 \geq (1 - \epsilon/2)d^3$ , so the number of triangles in  $\tilde{G}$  is less than  $(1 - \epsilon)d^3/6 \leq \frac{1-\epsilon}{1-\epsilon/2} \tilde{d}^3/6 \leq (1 - \epsilon/2)\tilde{d}^3/6$ . Clearly  $\sum_{u \in V(\tilde{G})} d_{\tilde{G}}(u)^2 \geq \tilde{n}\tilde{d}^2$ , so we can apply Theorem 2.7 with  $\epsilon/2$  in place of  $\epsilon$  and deduce that  $\tilde{G}$  has surplus  $\Omega_\epsilon(\tilde{d}^2) \geq \Omega_\epsilon(d^2)$ . It follows that  $G$  has the desired surplus.  $\square$

## 2.6 $C_5$ -free graphs

In this subsection, we prove the following result, which is a special case of Theorem 1.1. The proof already contains many of the ideas needed to prove the more general result, but is less technical.

**Theorem 2.8.** *Let  $G$  be a regular  $C_5$ -free graph with  $m$  edges. Then  $G$  has surplus  $\Omega(m^{6/7})$ .*

The following lemma will be convenient to use. (Here we only need it for  $r = 5$ , but we will use the general version later.)

**Lemma 2.9.** *Let  $r \geq 3$  and let  $G$  be a  $d$ -degenerate  $C_r$ -free graph with  $m$  edges. Then  $\text{sp}(G) = \Omega_r(\frac{m}{\sqrt{d}})$ .*

Since the proof is very standard, we just briefly sketch it. As  $G$  is  $C_r$ -free, the neighbourhood of any vertex induces a  $P_{r-2}$ -free graph. In particular, the number of edges in  $G[N(v)]$  is  $O_r(d(v))$ , so the number of triangles in  $G$  is at most  $O_r(m)$ . Using Corollary 1.2 from [6] or Lemma 3.3 from [5] (mentioned in Section 1.2), we conclude that  $\text{sp}(G) = \Omega_r(\frac{m}{\sqrt{d}})$ .

**Proof of Theorem 2.8.** Let  $d$  be the degree of  $G$  and let  $n$  be the number of vertices in  $G$ . First note that we may assume that  $d$  is arbitrarily large (when  $d$  is constant, the theorem follows from Lemma 2.9). Then clearly, the number of paths of length 2 in  $G$  is  $\Omega(nd^2)$ .

Here comes one of the key ideas of the proof. By dyadic pigeonholing, there exists a positive integer  $s$  such that for at least  $\Omega(\frac{nd^2}{s^{1/10}})$  paths  $uvw$  in  $G$ , we have  $s \leq d(u, w) < 2s$ . (The choice of  $1/10$  for the exponent is unimportant; it could be replaced by something smaller.) We now prove the following two lower bounds:

$$\text{sp}(G) = \Omega(nd^{1/2}s^{2/5}) \tag{2.3}$$

and

$$\text{sp}(G) = \Omega\left(\frac{d^3}{s^{6/5}}\right). \tag{2.4}$$

Combining these bounds, we can conclude that

$$\text{sp}(G) \geq \Omega\left((nd^{1/2}s^{2/5})^{6/7}\left(\frac{d^3}{s^{6/5}}\right)^{1/7}\right) = \Omega(n^{6/7}d^{6/7}s^{6/35}) \geq \Omega(m^{6/7}).$$

Before proving (2.3) and (2.4), let us briefly discuss the intuition behind them. We know that there are many paths  $uvw$  with  $d(u, w) \approx s$ . Assume for the moment that most paths of length 2 are like that. Then the second neighbourhood of a typical vertex has size roughly  $d^2/s$ . Since  $G$  is  $C_5$ -free, these neighbourhoods induce very sparse subgraphs. We can use these very sparse subgraphs to construct a large cut, just like in Theorem 2.7. Note that we can take roughly  $k = \frac{ns}{d^2}$  more-or-less disjoint second neighbourhoods of size roughly  $d^2/s$ . Each of them is very sparse whereas in a random graph with average degree  $d$  these sets would induce roughly  $(d^2/s)^2(d/n) = d^5/(s^2n)$  edges. This results in a surplus roughly  $k \cdot d^5/(s^2n) = d^3/s$ , which is what (2.4) gives, up to a small error term.

The inequality (2.3) is harder to explain, but observe that a natural type of graph with the property that paths  $uvw$  have  $d(u, w) \geq s$  is an  $s$ -blowup of another graph. However, when  $G$  is the  $s$ -blowup of  $H$ , then  $G$  has larger surplus than what is directly provided by Lemma 2.9. Indeed,  $H$  has  $m/s^2$  edges and degeneracy at most  $d/s$ , so by Lemma 2.9,  $\text{sp}(H) = \Omega(\frac{m}{d^{1/2}s^{3/2}})$ , and hence  $\text{sp}(G) = s^2\text{sp}(H) = \Omega(\frac{ms^{1/2}}{d^{1/2}}) = \Omega(nd^{1/2}s^{1/2})$ , coinciding with (2.3), up to a small error term. We now proceed with the proofs of (2.3) and (2.4).

**The lower bound  $\Omega(nd^{1/2}s^{2/5})$ .** We use the method inspired by the semidefinite relaxation of the MaxCut problem which we presented in Section 2.2. For a vertex  $v \in V(G)$ , let  $S(v)$  be the set of vertices  $u \neq v$  with  $s \leq d(u, v) < 2s$ . Now for every  $v \in V(G)$ , define  $\mathbf{x}^v \in \mathbb{R}^{V(G)}$  by

$$\mathbf{x}_u^v = \begin{cases} -\frac{1}{\sqrt{d}} & \text{if } u \in N(v) \setminus S(v), \\ \frac{\sqrt{s}}{d} & \text{if } u \in S(v) \setminus N(v), \\ -\frac{1}{\sqrt{d}} + \frac{\sqrt{s}}{d} & \text{if } u \in N(v) \cap S(v), \\ 0 & \text{otherwise.} \end{cases}$$

Then for any  $u, v \in V(G)$ ,

$$\langle \mathbf{x}^u, \mathbf{x}^v \rangle = |N(u) \cap N(v)| \frac{1}{d} + |S(u) \cap S(v)| \frac{s}{d^2} - (|N(u) \cap S(v)| + |S(u) \cap N(v)|) \frac{s^{1/2}}{d^{3/2}}. \quad (2.5)$$

In particular, for any  $u \in V(G)$ ,  $\|\mathbf{x}^u\|^2 \leq |N(u)|/d + |S(u)|s/d^2$ . Clearly  $|N(u)| \leq d$ . Moreover,  $|S(u)|s$  is a lower bound for the number of paths of the form  $uvw$  since for any  $w \in S(u)$  there are at least  $s$  such paths. Therefore,

$$|S(u)|s \leq d^2, \quad (2.6)$$

so we get  $\|\mathbf{x}^u\|^2 \leq 2$ . Hence, using (2.5), and Corollary 2.3,

$$\begin{aligned} \text{sp}(G) &\geq \Omega \left( \sum_{uv \in E(G)} (|N(u) \cap S(v)| + |S(u) \cap N(v)|) \frac{s^{1/2}}{d^{3/2}} \right) \\ &\quad - O \left( \sum_{uv \in E(G)} \left( |N(u) \cap N(v)| \frac{1}{d} + |S(u) \cap S(v)| \frac{s}{d^2} \right) \right). \end{aligned} \quad (2.7)$$

Observe that

$$\sum_{uv \in E(G)} |N(u) \cap N(v)| = 3t(G) \leq O(nd), \quad (2.8)$$

where in the inequality we used that for any  $v \in V(G)$ , the graph  $G[N(v)]$  is  $P_3$ -free.

Write  $\text{hom}(C_5, G)$  for the number of graph homomorphisms  $C_5 \rightarrow G$ . In other words,  $\text{hom}(C_5, G)$  is the number of 5-tuples  $(y_1, y_2, y_3, y_4, y_5) \in V(G)^5$  with  $y_5 y_1 \in E(G)$  and  $y_i y_{i+1} \in E(G)$  for every  $1 \leq i \leq 4$ . Note that if  $w \in S(u) \cap S(v)$  and  $uv \in E(G)$ , then there are at least  $s^2$  such 5-tuples of the form  $(u, v, y_3, w, y_5)$ . Thus,  $\sum_{uv \in E(G)} |S(u) \cap S(v)| \leq \frac{\text{hom}(C_5, G)}{s^2}$ . On the other hand,  $G$  does not contain a 5-cycle. Hence, in any such 5-tuple, we must have  $y_i = y_j$  for some  $i \neq j$ . Without loss of generality, assume that  $y_1 = y_4$ . Then  $y_1, y_2$  and  $y_3$  form a triangle in  $G$ . Since  $y_5$  is a neighbour of  $y_1$  and  $y_4 = y_1$ , the number of homomorphic  $C_5$ 's in  $G$  is at most  $O(t(G) \cdot \Delta(G))$ . As we noted in (2.8),  $t(G) = O(nd)$ , so  $\text{hom}(C_5, G) = O(nd^2)$  and hence

$$\sum_{uv \in E(G)} |S(u) \cap S(v)| \leq O \left( \frac{nd^2}{s^2} \right). \quad (2.9)$$

Finally, recall that the number of paths  $uvw$  in  $G$  with  $s \leq d(u, w) < 2s$  is  $\Omega(\frac{nd^2}{s^{1/10}})$ . Any such path has  $w \in S(u) \cap N(v)$ . Thus,

$$\sum_{uv \in E(G)} (|S(u) \cap N(v)| + |S(v) \cap N(u)|) \geq \Omega \left( \frac{nd^2}{s^{1/10}} \right). \quad (2.10)$$

Plugging in inequalities (2.8), (2.9) and (2.10) to (2.7), we get

$$\text{sp}(G) \geq \Omega(nd^{1/2}s^{2/5}) - O(n) - O(n/s) \geq \Omega(nd^{1/2}s^{2/5}).$$

**The lower bound  $\Omega(d^3/s^{6/5})$ .** This part of the proof is somewhat similar to the proof of Theorem 2.7. Let  $\nu nd^2$  be the number of paths  $uvw$  in  $G$  with  $s \leq d(u, w) < 2s$ . Recall that  $\nu = \Omega(s^{-1/10})$ . Let  $k = \lceil \frac{sn}{2d^2} \rceil$ . Note that  $\binom{n}{2} 2s$  is an upper bound for the number of paths  $uvw$  with  $d(u, w) < 2s$ , so  $\nu nd^2 \leq n^2 s$  and  $\nu d^2 / (sn) \leq 1$ . Give every vertex a random label from  $\{0, 1, \dots, k+1\}$ , where 0 is chosen with probability  $1/3$ , each of  $1, \dots, k$  is chosen with probability  $p = \frac{\nu d^2}{10sn}$ , and the remaining probability falls on  $k+1$ . This is feasible since  $kp \leq (\frac{sn}{2d^2} + 1) \frac{\nu d^2}{10sn} \leq \frac{\nu}{20} + \frac{1}{10} \leq 2/3$ . For every  $j \in [k]$ , let  $B_j$  be the set of vertices with label  $j$ .

Now, pick uniformly at random with repetition vertices  $v_1, \dots, v_k$ . For  $j \in [k]$ , let  $A_j$  be the set of vertices in  $S(v_j)$  with label 0 which do not belong to any other  $S(v_i)$ .

Let  $X$  be the number of edges which go between  $A_j$  and  $B_j$  for some  $j \in [k]$ , let  $Y$  be the number of edges inside some  $A_j$  and let  $Z$  be the number of edges inside some  $B_j$ .

*Claim 1.*  $\mathbb{E}[Y] \leq O\left(\frac{kd^2}{s^2}\right)$ .

*Proof of claim:* If an edge  $uv \in E(G)$  is in some  $A_j$ , then  $v_j \in S(u) \cap S(v)$ . Hence, the probability that  $uv$  contributes to  $Y$  is at most  $k \frac{|S(u) \cap S(v)|}{n}$ . Thus,

$$\mathbb{E}[Y] \leq \sum_{uv \in E(G)} k \frac{|S(u) \cap S(v)|}{n} \leq O\left(\frac{kd^2}{s^2}\right),$$

where for the second inequality we used (2.9).  $\diamond$

*Claim 2.*  $\mathbb{E}[Z] \leq \frac{kp^2nd}{2}$ .

*Proof of claim:* For any  $j$ , the expected number of edges in  $G[B_j]$  is  $p^2e(G) = p^2\frac{nd}{2}$ . Summing over the  $k$  possibilities for  $j$ , the claim follows.  $\diamond$

*Claim 3.*  $\mathbb{E}[X] \geq \frac{kpvd^3}{12s}$ .

*Proof of claim:* Let  $uv \in E(G)$ . For any  $j$ , the probability that  $v_j \in S(u)$ , but  $v_i \notin S(u)$  for every  $j \neq i$  is

$$\frac{|S(u)|}{n} \left(1 - \frac{|S(u)|}{n}\right)^{k-1} \geq \frac{|S(u)|}{n} \left(1 - \frac{(k-1)|S(u)|}{n}\right) \geq \frac{|S(u)|}{n} \left(1 - \frac{(k-1)d^2}{sn}\right) \geq \frac{|S(u)|}{2n},$$

where the second inequality uses (2.6). Thus, the probability that  $uv$  contributes to  $X$  is at least  $k \frac{p}{6n} (|S(u)| + |S(v)|)$ . Summing over all edges,

$$\mathbb{E}[X] \geq \frac{kp d}{6n} \sum_{u \in V(G)} |S(u)|.$$

Since the number of paths  $uvw$  with  $s \leq d(u, w) < 2s$  is  $\nu nd^2$ , there are at least  $\nu nd^2/(2s)$  pairs  $u, w$  with  $w \in S(u)$ , so  $\sum_{u \in V(G)} |S(u)| \geq \nu nd^2/(2s)$ , from which the claim follows.  $\diamond$

Combining the three claims, we get

$$\mathbb{E}[X - Y - Z] \geq kd \left( \frac{p\nu d^2}{12s} - O(d/s^2) - p^2n/2 \right) = kd \left( \frac{\nu^2 d^4}{120s^2n} - O(d/s^2) - \frac{\nu^2 d^4}{200s^2n} \right).$$

If  $\frac{\nu^2 d^4}{s^2n} = O(d/s^2)$ , then  $d^3 = O(n/\nu^2) \leq O(ns^{1/5})$ , so the bound  $\text{sp}(G) \geq \Omega(n\sqrt{d}) \geq \Omega(d^3/s^{6/5})$  follows from Lemma 2.9. Otherwise,

$$\mathbb{E}[X - Y - Z] \geq \Omega\left(\frac{k\nu^2 d^5}{s^2n}\right) \geq \Omega\left(\frac{\nu^2 d^3}{s}\right) \geq \Omega\left(\frac{d^3}{s^{6/5}}\right).$$

Therefore there exists an outcome in which  $X - Y - Z \geq \Omega(d^3/s^{6/5})$ , and then

$$\text{sp}(G) \geq \sum_{j=1}^k \text{sp}(G[A_j \cup B_j]) \geq \frac{1}{2} \sum_{j=1}^k (e(A_j, B_j) - e(A_j) - e(B_j)) = \frac{1}{2}(X - Y - Z) \geq \Omega(d^3/s^{6/5}),$$

completing the proof.  $\square$

### 3 Odd cycles

In this section, we prove Theorem 1.1. As discussed in Section 1.1, the tightness of the result follows from the known construction of dense pseudorandom  $C_r$ -free graphs due to Alon and Kahale [4]. It remains to show that, for fixed odd  $r \geq 3$ , any  $C_r$ -free graph with  $m$  edges has surplus  $\Omega_r(m^{(r+1)/(r+2)})$ . Since the case  $r = 3$  is covered by Alon's theorem, we may assume  $r \geq 5$ . The proof is similar to the one for  $C_5$  which we presented in Section 2, but more technical, and some new ideas are needed.

In the "sparse case", when the graph is  $d$ -degenerate, we have the bound  $\Omega(m/\sqrt{d})$  on the surplus. The rest of this section is devoted to the proof of the following lemma which we use in the "dense" case. We set

$$\alpha = \frac{r}{r+1} \text{ and } \beta = \frac{2r+1}{2r+2}. \quad (3.1)$$

Note that  $0 < \alpha < \beta$  and  $\alpha + \beta \leq 2$ .

**Lemma 3.1.** *Let  $G$  be a  $C_r$ -free graph with  $n$  vertices and average degree  $d$ . Then  $\text{sp}(G) = \Omega_r(n^\alpha d^\beta)$ .*

Before proving it, let us see how it implies Theorem 1.1. We will need the following lemma.

**Lemma 3.2.** *Let  $G$  be a graph and let  $d$  be a non-negative real number. Then there exists a bipartition  $(S, T)$  of  $V(G)$  such that  $G[S]$  is  $d$ -degenerate and  $G[T]$  has minimum degree at least  $d$ .*

**Proof.** Start with  $S_0 = \emptyset$  and  $T_0 = V(G)$ . If there exists a vertex  $v \in T_0$  with less than  $d$  neighbours in  $T_0$ , then set  $S_1 = S_0 \cup \{v\}$  and  $T_1 = T_0 \setminus \{v\}$ . More generally, for  $i \geq 0$ , if there exists  $v \in T_i$  with less than  $d$  neighbours in  $T_i$ , then let  $S_{i+1} = S_i \cup \{v\}$  and let  $T_{i+1} = T_i \setminus \{v\}$ . Eventually the process stops and we are left with a partition  $(S_k, T_k)$ . Take  $S = S_k$ ,  $T = T_k$ . By definition,  $G[S]$  is  $d$ -degenerate and  $G[T]$  has minimum degree at least  $d$ .  $\square$

**Proof of Theorem 1.1.** Let  $G$  be a  $C_r$ -free graph with  $m$  edges. Let  $d = m^{2/(r+2)}$  and take a partition  $(S, T)$  of  $V(G)$  as in Lemma 3.2. If  $e(S, T) \geq 2m/3$ , then clearly  $\text{sp}(G) \geq m/6$ . If  $e(S) \geq m/6$ , then by Lemma 2.9,

$$\text{sp}(G) \geq \text{sp}(G[S]) \geq \Omega_r\left(\frac{m/6}{\sqrt{d}}\right) = \Omega_r(m^{1-1/(r+2)})$$

as desired. Finally, suppose that  $e(T) \geq m/6$  and write  $n$  for the number of vertices in  $G[T]$ . Clearly,  $nd \geq 2e(T) \geq m/3$ . Therefore, by Lemma 3.1,

$$\text{sp}(G) \geq \text{sp}(G[T]) \geq \Omega_r(n^\alpha d^\beta) = \Omega_r\left(m^\alpha d^{\beta-\alpha}\right) = \Omega_r(m^{\alpha+2(\beta-\alpha)/(r+2)}) = \Omega_r(m^{1-1/(r+2)}),$$

since  $\alpha + 2(\beta - \alpha)/(r + 2) = 1 - 1/(r + 2)$ .  $\square$

It remains to prove Lemma 3.1. By Lemma 2.5 (applied with  $\alpha, \beta$  as defined in (3.1)), it suffices to prove Lemma 3.1 for  $C_r$ -free graphs with  $n$  vertices, average degree  $d$ , minimum degree  $\Omega_r(d)$  and maximum degree  $O_r(d)$ . Assume throughout that  $G$  is such a graph. We may of course also assume that  $d$  is sufficiently large.

Let  $\epsilon > 0$  be chosen sufficiently small, depending only on  $r$ . (Hence, implicit constants depending on  $\epsilon$  will also be indicated by a subscript  $r$ .) Set

$$\ell = \frac{r-1}{2}.$$

Recall from Section 2.6 that it was crucial to find many 2-paths  $uvw$  for which the codegree of  $u$  and  $w$ , that is, the number of 2-paths from  $u$  to  $w$ , was roughly the same. Here, we need to generalize this concept to paths of length  $\ell$ . For some steps of our argument, it will actually be more convenient to work with walks rather than paths. For two vertices  $u, v \in V(G)$  and an integer  $j$ , we write  $h_j(u, v)$  for the number of walks in  $G$  of length  $j$  from  $u$  to  $v$ .

Using the minimum degree condition, we easily observe that there are  $\Omega_r(nd^\ell)$  paths of length  $\ell$  in  $G$ . Now, the first crucial step in our argument is to select a large subset of  $\ell$ -paths which exhibit a certain regularity condition.

**Proposition 3.3.** *There exist  $c = c(r) > 0$ , positive integers  $(s_{i,j})_{0 \leq i < j \leq \ell}$  and at least*

$$c \left( \prod_{i < j} s_{i,j} \right)^{-\epsilon} nd^\ell$$

*paths  $u_0 u_1 \dots u_\ell$  in  $G$  with the property that*

$$s_{i,j} \leq h_{j-i}(u_i, u_j) < 2s_{i,j} \quad (3.2)$$

*for all  $0 \leq i < j \leq \ell$ .*

**Proof.** We apply dyadic pigeonholing. For each given path  $u_0 u_1 \dots u_\ell$ , there are unique non-negative integers  $(b_{i,j})_{0 \leq i < j \leq \ell}$  such that  $2^{b_{i,j}} \leq h_{j-i}(u_i, u_j) < 2^{b_{i,j}+1}$ .

Suppose, for a contradiction, that for every choice of  $(b_{i,j})_{0 \leq i < j \leq \ell}$ , there are at most

$$c \left( \prod_{i < j} 2^{b_{i,j}} \right)^{-\epsilon} nd^\ell$$

corresponding paths. Summing over all choices for the  $b_{i,j}$ , we deduce that the total number of  $\ell$ -paths is at most  $cnd^\ell \left( \sum_{b=0}^{\infty} 2^{-cb} \right)^{\binom{\ell+1}{2}} = cnd^\ell O_r(1)$ . Choosing  $c$  small enough, we reach a contradiction.  $\square$

From now on, we fix  $c$  and  $(s_{i,j})_{0 \leq i < j \leq \ell}$  which satisfy the conclusion of Proposition 3.3, and call the paths  $u_0 u_1 \dots u_\ell$  in  $G$  which satisfy (3.2) *good*. Hence, there are at least  $\Omega_r(\nu nd^\ell)$  good  $\ell$ -paths in  $G$ , where we set

$$\nu = \left( \prod_{i < j} s_{i,j} \right)^{-\epsilon}.$$

Note that we clearly have  $s_{i,j} \leq \Delta(G)^{j-i-1} = d^{O(\ell)}$  for all  $0 \leq i < j \leq \ell$ . Thus, for small enough  $\epsilon$ , we have, say,

$$\nu \geq d^{-1/10}. \quad (3.3)$$

We extend the notion of a good path to smaller lengths in the obvious way. That is, for  $q \leq \ell$ , a path  $u_0 u_1 \dots u_q$  is called *good* if (3.2) holds for all  $0 \leq i < j \leq q$ . Since  $G$  has maximum degree  $O_r(d)$ , every  $q$ -path can be extended into at most  $O_r(d^{\ell-q})$   $\ell$ -paths. Moreover, every subpath of a good  $\ell$ -path (with the same start vertex) is also good. Hence, for each  $q \leq \ell$ , there are at least  $\Omega_r(\nu nd^q)$  good  $q$ -paths in  $G$ .

We now state the key lemmas which provide lower bounds on the surplus. Using the SDP (semi-definite programming) method, we will prove the following.

**Lemma 3.4.** *For each  $2 \leq q \leq \ell$ , we have*

$$\text{sp}(G) = \Omega_r \left( \nu nd^{1/2} (s_{0,q}/s_{0,q-1})^{1/2} \right).$$

**Corollary 3.5.** *We have*

$$\text{sp}(G) = \Omega_r \left( \nu nd^{1/2} s_{0,\ell}^{1/2(\ell-1)} \right).$$

**Proof.** By the AM–GM inequality,

$$\frac{1}{\ell-1} \sum_{q=2}^{\ell} (s_{0,q}/s_{0,q-1})^{1/2} \geq \left( \prod_{q=2}^{\ell} (s_{0,q}/s_{0,q-1})^{1/2} \right)^{1/(\ell-1)} = s_{0,\ell}^{1/2(\ell-1)}$$

since  $s_{0,1} = 1$ . We infer that there exists  $2 \leq q \leq \ell$  with  $(s_{0,q}/s_{0,q-1})^{1/2} \geq s_{0,\ell}^{1/2(\ell-1)}$ . Lemma 3.4 then implies the claim.  $\square$

Moreover, using random neighbourhood sampling, we will prove the following bound on the surplus. We actually only need this for  $q = \ell$ .

**Lemma 3.6.** *For each  $2 \leq q \leq \ell$ , we have*

$$\text{sp}(G) = \Omega_r(\nu^2 d^{q+1} / s_{0,q}).$$

Before proving the two key lemmas, let us complete the proof of Lemma 3.1. We need one final simple observation.

**Proposition 3.7.** *For any  $0 \leq i \leq i' < j' \leq j \leq \ell$ , we have  $s_{i',j'} \leq 2s_{i,j}$ . In particular,  $\nu = (2s_{0,\ell})^{-O(\ell^2\epsilon)}$ .*

**Proof.** Consider any good  $\ell$ -path  $u_0 u_1 \dots u_\ell$ . Then we know  $h_{j-i}(u_i, u_j) \leq 2s_{i,j}$ . On the other hand, we have  $h_{j-i}(u_i, u_j) \geq h_{j'-i'}(u_{i'}, u_{j'}) \geq s_{i',j'}$  since we can replace  $u_{i'} \dots u_{j'}$  with any walk of length  $j' - i'$  from  $u_{i'}$  to  $u_{j'}$ . Hence,  $s_{i',j'} \leq 2s_{i,j}$ , proving the first claim. The second claim easily follows.  $\square$

**Proof of Lemma 3.1.** Write  $s = s_{0,\ell}$ . We have

$$\max\{d^{\ell+1}/s, nd^{1/2}s^{1/2(\ell-1)}\} \geq \left(\frac{d^{\ell+1}}{s} \cdot (nd^{1/2}s^{1/2(\ell-1)})^{2\ell+1}\right)^{\frac{1}{2\ell+2}} = n^\alpha d^\beta s^{\frac{3}{(r-3)(r+1)}}.$$

Thus, by Corollary 3.5 and Lemma 3.6, we have  $\text{sp}(G) = \Omega_r(\nu^2 n^\alpha d^\beta s^{\frac{3}{(r-3)(r+1)}})$ . Finally, by Proposition 3.7, we can ensure that  $\nu^2 s^{\frac{3}{(r-3)(r+1)}} = \Omega(1)$  when  $\epsilon$  is sufficiently small, which completes the proof.  $\square$

**Remark 3.8.** Observe that the same proof, using  $\max\{a, b\} \geq (ab^{2\ell+2})^{1/(2\ell+3)}$  instead of  $\max\{a, b\} \geq (ab^{2\ell+1})^{1/(2\ell+2)}$ , gives us  $\text{sp}(G) = \Omega_r((nd)^{(2\ell+2)/(2\ell+3)})$ , which proves Theorem 1.1 directly when  $G$  is (almost) regular. We used the slightly biased weighting here in order to be able to apply our regularization lemma.

It remains to prove the key Lemmas 3.4 and 3.6. Since some steps in the proofs are similar, we treat them in a unified setup. For the rest of this section, fix  $2 \leq q \leq \ell$ . To enhance readability, we also set  $s = s_{0,q}$  and  $s' = s_{0,q-1}$ .

**Partition of  $V(G)$ .** We will be interested in good paths of length  $q$ . In order to maintain better control on how different such paths can intersect, we consider ‘‘partite’’ paths where the vertices are chosen from disjoint vertex subsets. To this end, partition  $V(G)$  randomly as  $U_0 \cup U_1 \cup \dots \cup U_q$  by placing, each vertex  $u$  independently, in  $U_i$  with probability  $1/(q+1)$ , for every  $0 \leq i \leq q$ . Write  $\mathcal{A}$  for the set of tuples  $(u_0, u_1, \dots, u_q) \in U_0 \times U_1 \times \dots \times U_q$  such that  $u_0 u_1 \dots u_q$  is a good path. Write  $\mathcal{B}$  for the set of tuples  $(u_0, u_1, \dots, u_{q-1}) \in U_0 \times U_1 \times \dots \times U_{q-1}$  such that  $u_0 u_1 \dots u_{q-1}$  is a good path.

Since there are  $\Omega_r(\nu n d^q)$  good paths of length  $q$  in  $G$ , each of which contributes to  $\mathcal{A}$  with probability  $(q+1)^{-(q+1)} = \Omega_r(1)$ , we have  $\mathbb{E}[|\mathcal{A}|] = \Omega_r(\nu n d^q)$ . Hence, there exists a partition  $U_0 \cup U_1 \cup \dots \cup U_q$ , which we will fix from now on, for which  $|\mathcal{A}| = \Omega_r(\nu n d^q)$ . (Note that, since any tuple in  $\mathcal{B}$  can be extended to at most  $\Delta(G)$  tuples in  $\mathcal{A}$ , we also know that  $|\mathcal{B}| = \Omega_r(\nu n d^{q-1})$ , although we will not explicitly need this.)

**The sets  $S(u), T(u)$ .** For every vertex  $u \in V(G)$ , define sets  $T(u)$  and  $S(u)$  as follows. If  $u \in U_0 \cup U_1 \cup \dots \cup U_{q-2}$ , then let  $T(u) = S(u) = \emptyset$ . If  $u \in U_{q-1}$ , then let  $S(u) = \emptyset$  and let  $T(u)$  be the set of those  $u_0 \in U_0$  for which there exist at least  $\frac{|\mathcal{A}|s'}{4n\Delta(G)^q} = \Omega_r(\nu s')$  tuples of the form  $(u_0, u_1, u_2, \dots, u_{q-2}, u) \in \mathcal{B}$ . Finally, if  $u \in U_q$ , then let  $T(u) = \emptyset$  and let  $S(u)$  be the set of those  $u_0 \in U_0$  for which there exist at least  $\frac{|\mathcal{A}|s}{4n\Delta(G)^q} = \Omega_r(\nu s)$  tuples of the form  $(u_0, u_1, \dots, u_{q-1}, u) \in \mathcal{A}$ .

Very roughly, we can think of  $S(u)$  as the  $q$ th neighbourhood of  $u$ , but we only use this for  $u \in U_q$  and only consider “ $q$ th neighbours” in  $U_0$  which can be reached by a substantial amount of good paths.

The following lemma is the crucial point where we use the assumption that  $G$  is  $C_r$ -free. We will use it in the proofs of both key lemmas in order to bound terms that would otherwise drive the surplus to be small.

**Lemma 3.9.** *We have*

$$\sum_{uv \in E(G)} |S(u) \cap S(v)| = O_r \left( \frac{nd^q}{\nu s} \right) \quad \text{and} \quad \sum_{uv \in E(G)} |T(u) \cap T(v)| = O_r \left( \frac{nd^{q-1}}{\nu s'} \right).$$

The intuition is that if  $u_0 \in S(u) \cap S(v)$ , then we must have  $u, v \in U_q$ , and two of the good paths in  $\mathcal{A}$ , one from  $u_0$  to  $u$  and one from  $u_0$  to  $v$ , will give us together with the edge  $uv$  an odd cycle. In the extreme case that  $q = \ell$  and the two paths are internally vertex-disjoint, this would be a cycle of length  $2\ell + 1 = r$ , a contradiction to our assumption that  $G$  is  $C_r$ -free. In order to prove the more general case when  $q$  can be smaller and the paths might not be internally disjoint, we need a counting lemma which we will establish in the next subsection.

### 3.1 Counting good paths and short cycles

As discussed above, when  $u_0 \in S(u) \cap S(v)$ , then we find an odd cycle of the form  $u_i \dots u_{q-1} uvv_{q-1} \dots v_i$ , where  $u_i = v_i$  and  $u_j, v_j \in U_j$  for all  $i \leq j \leq q - 1$ . The following lemma shows that we can utilize the  $C_r$ -freeness of  $G$  to control such short odd cycles. In its statement, one can think of  $z = u_i = v_i$  and  $R \subset U_q$ .

**Lemma 3.10.** *Let  $z \in V(G)$  and let  $R \subset V(G) \setminus \{z\}$ . Let  $j \leq \ell$  be a positive integer. Define the auxiliary graph  $H$  with vertex set  $R$  such that  $uv \in E(H)$  if there is a  $C_{2j+1}$  in  $G$  of the form  $zu_1u_2 \dots u_{j-1}uvv_{j-1}v_{j-2} \dots v_1z$  with  $u_1, \dots, u_{j-1}, v_1, \dots, v_{j-1} \notin R$ . Then  $H$  has average degree at most  $2^{2j+1}(r - 2j) = O_r(1)$ .*

**Proof.** Define a random partition  $V(G) \setminus \{z\} = A \cup B$  by placing each vertex on either side with probability  $1/2$ , independently of the other vertices. Let  $H'$  be the subgraph of  $H$  on the same vertex set  $R$  in which  $ab$  is an edge if  $a \in A, b \in B$  and there is a  $C_{2j+1}$  in  $G$  of the form  $zu_1u_2 \dots u_{j-1}abv_{j-1}v_{j-2} \dots v_1z$  such that  $u_1, \dots, u_{j-1} \in A \setminus R$  and  $v_1, \dots, v_{j-1} \in B \setminus R$ . Clearly, each edge in  $H$  is an edge in  $H'$  with probability at least  $2^{-2j}$ . Hence,  $\mathbb{E}[e(H')] \geq 2^{-2j}e(H)$ , and there exists a partition  $A, B$  for which  $e(H') \geq 2^{-2j}e(H)$ .

Crucially,  $H'$  cannot contain a path of length  $r - 2j$ . Indeed, suppose there is such a path  $a_1b_1a_2b_2 \dots a_ib_i$  where  $r - 2j = 2i - 1$ . Without loss of generality,  $a_1 \in A$  and  $b_i \in B$ . Since  $a_1b_1 \in E(H')$ , there exists a path  $zu_1u_2 \dots u_{j-1}a_1$  in  $G$  with  $u_1, u_2, \dots, u_{j-1} \in A \setminus R$ . Similarly, there exists a path  $zv_1v_2 \dots v_{j-1}b_i$  in  $G$  with  $v_1, v_2, \dots, v_{j-1} \in B \setminus R$ . Then

$$zu_1u_2 \dots u_{j-1}a_1b_1a_2b_2 \dots a_ib_iv_{j-1} \dots v_1z$$

is a  $C_r$  in  $G$ , which is a contradiction. Thus,  $H'$  does not contain a path of length  $r - 2j$ . It follows that  $H'$  has average degree at most  $2(r - 2j)$ . Hence,  $H$  has average degree at most  $2^{2j+1}(r - 2j) = O_r(1)$ .  $\square$

**Lemma 3.11.** *Let  $z \in V(G)$  and let  $R \subset V(G) \setminus \{z\}$ . Let  $j \leq \ell$  be a positive integer. Suppose there exists a real number  $s'' \geq 1$  such that for any  $u \in R$ , we have  $s'' \leq h_j(z, u) < 2s''$ . Then there are at most  $O_r(d^j)$  paths  $zu_1 \dots u_{j-1}uv$  in  $G$  with  $u_1, \dots, u_{j-1} \notin R, u, v \in R$  which can be extended to a  $C_{2j+1}$  of the form  $zu_1 \dots u_{j-1}uvv_{j-1} \dots v_1z$  with  $v_1, \dots, v_{j-1} \notin R$ .*

**Proof.** Since every  $u \in R$  has  $h_j(z, u) \geq s''$  and the total number of walks of length  $j$  starting from  $z$  is at most  $\Delta(G)^j = O_r(d^j)$ , we have  $|R| = O_r(d^j/s'')$ . Define the auxiliary graph  $H$  like in Lemma 3.10. Then we have  $e(H) = O_r(d^j/s'')$ . Since  $h_j(z, u) < 2s''$  for every  $u \in R$ , there are at most  $2e(H) \cdot 2s'' = O_r(d^j)$  paths  $zu_1 \dots u_{j-1}uv$  in  $G$  with  $u_1, \dots, u_{j-1} \notin R$  and  $u, v \in R$  which can be extended to a  $C_{2j+1}$  of the form  $zu_1 \dots u_{j-1}uvv_{j-1} \dots v_1z$  with  $v_1, \dots, v_{j-1} \notin R$ .  $\square$

**Lemma 3.12.** *The number of tuples  $(u_0, \dots, u_q, v)$  with  $(u_0, \dots, u_q) \in \mathcal{A}$ ,  $v \in N(u_q) \cap U_q$  and  $u_0 \in S(v)$  is at most  $O_r(nd^q)$ . Similarly, the number of tuples  $(u_0, \dots, u_{q-1}, v)$  with  $(u_0, \dots, u_{q-1}) \in \mathcal{B}$ ,  $v \in N(u_{q-1}) \cap U_{q-1}$  and  $u_0 \in T(v)$  is at most  $O_r(nd^{q-1})$ .*

**Proof.** We prove the first claim. The proof of the second is almost verbatim the same. Take a tuple  $(u_0, \dots, u_q, v)$  with  $(u_0, \dots, u_q) \in \mathcal{A}$  and  $v \in N(u_q) \cap U_q$  for which  $u_0 \in S(v)$ . The last condition implies that there exists  $(v_0, v_1, \dots, v_{q-1}, v) \in \mathcal{A}$  with  $v_0 = u_0$ . Let  $i$  be the largest index for which  $u_i = v_i$ . Then  $u_i \dots u_{q-1} u_q v v_{q-1} \dots v_i$  is a  $C_{2j+1}$  in  $G$ , where  $1 \leq j = q - i \leq q \leq \ell$ . We now count the number of choices for  $(u_i, u_{i+1}, \dots, u_{q-1}, u_q, v)$ .

There are at most  $n$  possibilities for  $u_i$ . Fix such a choice  $u_i$  (which also determines  $i$ ). Set  $R = \{u' \in U_q : s_{i,q} \leq h_{q-i}(u_i, u') < 2s_{i,q}\}$ . By Lemma 3.11 (with  $z = u_i = v_i$  and  $s'' = s_{i,q}$ ), there are at most  $O_r(d^j)$  choices for  $(u_{i+1}, \dots, u_{q-1}, u_q, v)$ . Given any such choice, there are clearly at most  $\Delta(G)^i = O_r(d^{q-j})$  possibilities to choose  $u_{i-1}, \dots, u_0$ . In total, this gives  $O_r(nd^q)$  possibilities.  $\square$

**Proof of Lemma 3.9.** We prove the first bound, the proof of the second is verbatim the same. Note that  $\sum_{uv \in E(G)} |S(u) \cap S(v)|$  is half the number of triples  $(u_0, u, v)$  with  $u, v \in U_q$ ,  $uv \in E(G)$  and  $u_0 \in S(u) \cap S(v)$ . Each such triple, spelling out  $u_0 \in S(u)$ , gives us  $\Omega_r(\nu s)$  tuples  $(u_0, \dots, u_{q-1}, u, v)$  with  $(u_0, \dots, u_{q-1}, u) \in \mathcal{A}$ ,  $v \in N(u) \cap U_q$  and  $u_0 \in S(v)$ .

Thus, it suffices to show that the total number of tuples  $(u_0, \dots, u_{q-1}, u, v)$  with  $(u_0, \dots, u_{q-1}, u) \in \mathcal{A}$ ,  $v \in N(u) \cap U_q$  and  $u_0 \in S(v)$  is at most  $O_r(nd^q)$ . This follows from Lemma 3.12.  $\square$

One last tool is the following, which will help us in both key lemmas to give good bounds on terms that influence the surplus in our favour.

**Proposition 3.13.** *At least half of the tuples  $(u_0, \dots, u_{q-1}, u_q) \in \mathcal{A}$  satisfy  $u_0 \in T(u_{q-1}) \cap S(u_q)$ .*

**Proof.** We count the number of tuples  $(u_0, \dots, u_{q-1}, u_q) \in \mathcal{A}$  with  $u_0 \notin S(u_q)$  as follows. First, we have at most  $n$  choices for  $u_0$ . Then, the number of possibilities for  $u_q$  such that there exists at least one tuple  $(u_0, u_1, \dots, u_q) \in \mathcal{A}$  is at most  $\Delta(G)^q/s$ , since there are obviously at most  $\Delta(G)^q$  walks of length  $q$  starting from  $u_0$ , and if the final vertex  $u_q$  is such that there exists a good  $q$ -path from  $u_0$  to  $u_q$ , then we have  $h_q(u_0, u_q) \geq s$ . Now, given  $u_0$  and  $u_q$ , if  $u_0 \notin S(u_q)$ , then there are at most  $\frac{|\mathcal{A}|s}{4n\Delta(G)^q}$  tuples of the form  $(u_0, \dots, u_{q-1}, u_q) \in \mathcal{A}$  by definition of  $S(u_q)$ . Hence, the total number of such tuples is at most  $|\mathcal{A}|/4$ .

The same proof shows that the number of  $(u_0, \dots, u_{q-1}) \in \mathcal{B}$  with  $u_0 \notin T(u_{q-1})$  is at most

$$n \cdot \frac{\Delta(G)^{q-1}}{s'} \cdot \frac{|\mathcal{A}|s'}{4n\Delta(G)^q} = \frac{|\mathcal{A}|}{4\Delta(G)}.$$

Since any tuple in  $\mathcal{B}$  can be extended to at most  $\Delta(G)$  tuples in  $\mathcal{A}$ , we see that the number of tuples  $(u_0, \dots, u_{q-1}, u_q) \in \mathcal{A}$  with  $u_0 \notin T(u_{q-1})$  is also at most  $|\mathcal{A}|/4$ , which completes the proof.  $\square$

### 3.2 Lower bound using the SDP method

Our goal in this subsection is to prove Lemma 3.4. We need one more tool which controls the gain that we will achieve with the SDP method.

**Proposition 3.14.** *We have*

$$\sum_{uv \in E(G)} |S(u) \cap T(v)| = \Omega_r \left( \frac{\nu nd^q}{s'} \right),$$

where the sum here is over all ordered pairs  $(u, v)$  with  $uv \in E(G)$ .

**Proof.** By Proposition 3.13, there are  $\Omega_r(\nu nd^q)$  tuples  $(u_0, \dots, u_{q-1}, u_q) \in \mathcal{A}$  which satisfy  $u_0 \in T(u_{q-1}) \cap S(u_q)$ . Each of these gives a triple  $(u_0, u_{q-1}, u_q)$  which is counted in  $\sum_{uv \in E(G)} |S(u) \cap T(v)|$ . Moreover, any such triple  $(u_0, u_{q-1}, u_q)$  arises in this way at most  $2s'$  times since, if some  $(u_0, u_1, \dots, u_{q-1}, u_q) \in \mathcal{A}$  exists, then the number of choices for  $u_1, \dots, u_{q-2}$  is at most  $h_{q-1}(u_0, u_{q-1}) < 2s'$ .  $\square$

**Proof of Lemma 3.4.** For every  $u \in V(G)$ , define  $\mathbf{x}^u \in \mathbb{R}^{V(G)}$  by

$$\mathbf{x}_w^u = \begin{cases} -(s'/d^{q-1})^{1/2} & \text{if } w \in T(u), \\ (s/d^q)^{1/2} & \text{if } w \in S(u), \\ 0 & \text{otherwise.} \end{cases}$$

This is well-defined since for any  $u \in V(G)$ , we have  $T(u) \cap S(u) = \emptyset$ .

Our goal is to apply Corollary 2.3. Clearly, we have  $\|\mathbf{x}^u\|^2 \leq |T(u)|s'/d^{q-1} + |S(u)|s/d^q$ . For each  $u_0 \in T(u)$ , we know that  $h_{q-1}(u_0, u) \geq s'$ . Hence, the number of walks of length  $q-1$  starting from  $u$  is at least  $|T(u)|s'$ . On the other hand, this quantity is trivially bounded from above by  $\Delta(G)^{q-1}$ . Hence, we have  $|T(u)|s'/d^{q-1} = O_r(1)$ . Similarly, the number of walks of length  $q$  starting from  $u$  is at least  $|S(u)|s$  and at most  $\Delta(G)^q$ , implying that  $|S(u)|s/d^q = O_r(1)$ , too. Hence,  $\|\mathbf{x}^u\| = O_r(1)$  for all  $u \in V(G)$ .

Now, consider  $uv \in E(G)$ . We have  $\langle \mathbf{x}^u, \mathbf{x}^v \rangle = -a_{uv} + b_{uv}$ , where

$$a_{uv} = (|T(u) \cap S(v)| + |S(u) \cap T(v)|) \frac{(ss')^{1/2}}{d^{q-1/2}},$$

$$b_{uv} = |T(u) \cap T(v)| \frac{s'}{d^{q-1}} + |S(u) \cap S(v)| \frac{s}{d^q}.$$

Invoking Proposition 3.14, we have a ‘‘gain’’ of

$$\sum_{uv \in E(G)} a_{uv} = \Omega_r \left( \frac{\nu nd^q (ss')^{1/2}}{s' d^{q-1/2}} \right) = \Omega_r \left( \nu nd^{1/2} (s/s')^{1/2} \right).$$

On the other hand, by Lemma 3.9, we have a ‘‘loss’’ of

$$\sum_{uv \in E(G)} b_{uv} = O_r \left( \frac{nd^{q-1} s'}{\nu s' d^{q-1}} + \frac{nd^q s}{\nu s d^q} \right) = O_r(n/\nu).$$

Since  $s' \leq 2s$  by Proposition 3.7 and  $\nu \geq d^{-1/10}$  by (3.3), the second sum is negligible compared to the first, and so by Corollary 2.3, we conclude  $\text{sp}(G) = \Omega_r(\nu nd^{1/2}(s/s')^{1/2})$ . This completes the proof.  $\square$

Note that in the above proof,  $\mathbf{x}^u$  is the zero vector unless  $u \in U_{q-1} \cup U_q$ . This essentially means that we find a cut with large surplus in  $G[U_{q-1} \cup U_q]$ .

### 3.3 Random neighbourhood sampling

Our goal in this subsection is to prove Lemma 3.6. While in the proof of Lemma 3.4, the dominant term for the surplus was given by the sum over all edges  $uv$  of the intersection size  $|S(u) \cap T(v)|$ , here, the surplus will be mainly given by  $\sum_{u \in V(G)} |S(u)|$ . Recall that  $S(u) = \emptyset$  unless  $u \in U_q$ , so the important vertices here are the ones in  $U_q$ . Roughly speaking, we find subsets  $A_j$  inside  $U_q$  which contain few edges, and use this to find a large cut.

**Proposition 3.15.** *We have  $\sum_{u \in V(G)} |S(u)| = \Omega_r(\nu nd^q/s)$ .*

**Proof.** The sum  $\sum_{u \in V(G)} |S(u)|$  counts the number of pairs  $(u_0, u_q) \in U_0 \times U_q$  with  $u_0 \in S(u_q)$ . By Proposition 3.13 and  $\mathcal{A} \geq \Omega_r(\nu nd^q)$ , and since each pair  $(u_0, u_q) \in U_0 \times U_q$  appears in at most  $2s$  tuples in  $\mathcal{A}$ , we have  $\sum_{u \in V(G)} |S(u)| = \Omega_r(\nu nd^q/s)$ .  $\square$

**Proof of Lemma 3.6.** Let  $\mu$  be a sufficiently small constant depending only on  $r$ . Note that  $|\mathcal{A}| \leq n^2(2s)$  since for any choice of  $(u_0, u_q) \in U_0 \times U_q$  that appears in some tuple in  $\mathcal{A}$ , there are at most  $2s$  tuples containing it. Moreover, recall  $|\mathcal{A}| = \Omega_r(\nu n d^q)$ . Hence,  $\frac{\nu d^q}{sn} = O_r(1)$ . Let

$$k = \left\lceil \frac{sn}{2\Delta(G)^q} \right\rceil.$$

Give every vertex a random label from  $\{0, 1, \dots, k+1\}$ , where 0 is chosen with probability  $1/3$ , each of  $1, \dots, k$  is chosen with probability  $p = \mu \frac{\nu d^q}{sn}$ , and the remaining probability falls on  $k+1$ . This is feasible since  $kp \leq \mu/2 + p \leq 2/3$ , provided  $\mu$  is sufficiently small. For every  $j \in [k]$ , let  $B_j$  be the set of vertices with label  $j$ .

Now, pick uniformly at random with repetition vertices  $v_1, \dots, v_k \in V(G)$ . For  $j \in [k]$ , let  $A_j$  be the set of vertices  $u$  with label 0 such that  $v_j \in S(u)$ , but  $v_i \notin S(u)$  for all  $i \neq j$ .

Let  $X$  be the number of edges which go between  $A_j$  and  $B_j$  for some  $j \in [k]$ , let  $Y$  be the number of edges inside some  $A_j$  and let  $Z$  be the number of edges inside some  $B_j$ .

*Claim 1.*  $\mathbb{E}[Y] = O_r\left(\frac{kd^q}{\nu s}\right)$ .

*Proof of claim:* If an edge  $uv \in E(G)$  is in some  $A_j$ , then  $v_j \in S(u) \cap S(v)$ . Hence, the probability that  $uv$  contributes to  $Y$  is at most  $k \frac{|S(u) \cap S(v)|}{n}$ . Thus,

$$\mathbb{E}[Y] \leq \sum_{uv \in E(G)} k \frac{|S(u) \cap S(v)|}{n} = O_r\left(\frac{kd^q}{\nu s}\right)$$

by Lemma 3.9. ◇

*Claim 2.*  $\mathbb{E}[Z] = O(kp^2nd)$ .

*Proof of claim:* For any  $j$ , the expected number of edges in  $G[B_j]$  is  $p^2 e(G) = p^2 \frac{nd}{2}$ . Summing over the  $k$  possibilities for  $j$ , the claim follows. ◇

*Claim 3.*  $\mathbb{E}[X] = \Omega_r(kp\nu d^{q+1}/s)$ .

*Proof of claim:* Let  $uv \in E(G)$ . For any  $j \in [k]$ , the probability that  $v_j \in S(u)$ , but  $v_i \notin S(u)$  for every  $i \neq j$ , is

$$\frac{|S(u)|}{n} \left(1 - \frac{|S(u)|}{n}\right)^{k-1} \geq \frac{|S(u)|}{n} \left(1 - \frac{(k-1)|S(u)|}{n}\right) \geq \frac{|S(u)|}{n} \left(1 - \frac{(k-1)\Delta(G)^q}{sn}\right) \geq \frac{|S(u)|}{2n},$$

where we used the bound  $|S(u)|s \leq \Delta(G)^q$  (explained in the proof of Lemma 3.4) and the definition of  $k$ . Thus, the probability that  $uv$  contributes to  $X$  is at least  $k \frac{p}{6n} (|S(u)| + |S(v)|)$ . Hence,

$$\mathbb{E}[X] \geq \frac{kp}{6n} \sum_{u \in V(G)} d(u)|S(u)| \geq \frac{kp\delta(G)}{6n} \sum_{u \in V(G)} |S(u)| = \Omega_r(kp\nu d^{q+1}/s),$$

where we have used Proposition 3.15 in the last step. ◇

Combining the three claims, and plugging in  $p$  with a sufficiently small  $\mu$ , we get

$$\mathbb{E}[X - Y - Z] \geq kd \left( \Omega_r(p\nu d^q/s) - O_r\left(\frac{d^{q-1}}{\nu s}\right) - O(p^2n) \right) \geq kd \left( \Omega_r\left(\frac{\nu^2 d^{2q}}{s^2 n}\right) - O_r\left(\frac{d^{q-1}}{\nu s}\right) \right).$$

If  $\frac{\nu^2 d^{2q}}{s^2 n} = O_r\left(\frac{d^{q-1}}{\nu s}\right)$ , then  $\frac{\nu^2 d^{q+1}}{s} = O_r(n/\nu) = O_r(nd^{1/10})$  by (3.3), so the surplus we are aiming for already follows from Lemma 2.9. Otherwise,

$$\mathbb{E}[X - Y - Z] = \Omega_r\left(\frac{k\nu^2 d^{2q+1}}{s^2 n}\right) = \Omega_r\left(\frac{\nu^2 d^{q+1}}{s}\right).$$

Then there exists an outcome in which  $X - Y - Z = \Omega_r(\nu^2 d^{q+1}/s)$ , and therefore

$$\text{sp}(G) \geq \sum_{j=1}^k \text{sp}(G[A_j \cup B_j]) \geq \frac{1}{2}(X - Y - Z) = \Omega_r(\nu^2 d^{q+1}/s)$$

by Lemma 2.1, completing the proof. □

## 4 Few triangles

In this section, we prove Theorem 1.3. Similarly to the proof of Theorem 2.4, we let  $a = \frac{\sqrt{d}}{n-d}$ ,  $0 < \gamma \leq 1$  and for every  $v \in V(G)$ , we define  $\mathbf{x}^v \in \mathbb{R}^{V(G)}$  by

$$\mathbf{x}_u^v = \begin{cases} 1 + \gamma a & \text{if } u = v, \\ -\gamma \frac{1}{\sqrt{d}} & \text{if } u \in N(v), \\ \gamma a & \text{otherwise.} \end{cases}$$

Just like in the strongly regular case, the sum  $\sum_{uv \in E(G)} \frac{\langle \mathbf{x}^u, \mathbf{x}^v \rangle}{\|\mathbf{x}^u\| \|\mathbf{x}^v\|}$  is negative for a suitable choice of  $\gamma$ . However, this does not directly lead to a large surplus because the inner product is not negative for every  $uv \in E(G)$ , and for positive  $x$ , we have  $\arcsin(x) > x$ . Nevertheless, for small positive  $x$ ,  $\arcsin(x)$  is very close to  $x$ , so the main discrepancy is caused by edges  $uv$  for which  $\langle \mathbf{x}^u, \mathbf{x}^v \rangle$  is very large. In view of (2.1), this happens when  $d(u, v)$  is much larger than  $d^2/n$ . To deal with these edges, we will introduce a slightly modified version of  $\mathbf{x}^v$ . For  $v \in V(G)$ , define  $\mathbf{y}^v \in \mathbb{R}^{V(G)}$  randomly by

$$\mathbf{y}_u^v = \begin{cases} 1 + \gamma a & \text{if } u = v, \\ -\gamma \frac{1}{\sqrt{d}} & \text{if } u \in N(v) \text{ and } d(u, v) \leq 20d^2/n, \\ \pm \gamma \frac{1}{\sqrt{d}} & \text{if } u \in N(v) \text{ and } d(u, v) > 20d^2/n, \\ \gamma a & \text{otherwise,} \end{cases}$$

where, for any  $uv \in E(G)$  with  $d(u, v) > 20d^2/n$ , the sign of  $\mathbf{y}_u^v$  is  $+$  with probability  $1/2$ , independently of all other coordinates and vectors.

Our first lemma states that if there are many edges with large codegree, then, when applying Lemma 2.2, the vectors  $\mathbf{y}^v$  provide significantly better surplus than the vectors  $\mathbf{x}^v$ .

**Lemma 4.1.** *Let  $0 < \gamma \leq 1/10$  and  $d \leq n/2$ . If  $G$  is an  $n$ -vertex  $d$ -regular graph with at most  $d^3/3$  triangles and the vectors  $\mathbf{x}^v, \mathbf{y}^v$  are defined as above, then*

$$\mathbb{E} \left[ \sum_{uv \in E(G)} \left( \arcsin \left( \frac{\langle \mathbf{y}^u, \mathbf{y}^v \rangle}{\|\mathbf{y}^u\| \|\mathbf{y}^v\|} \right) - \arcsin \left( \frac{\langle \mathbf{x}^u, \mathbf{x}^v \rangle}{\|\mathbf{x}^u\| \|\mathbf{x}^v\|} \right) \right) \right] \leq \frac{\gamma n \sqrt{d}}{4} - \gamma^2 \left( \sum_{\substack{uv \in E(G): \\ d(u, v) > 20d^2/n}} \frac{d(u, v)}{10d} \right).$$

The proof of this lemma is a tedious calculation, so it is given in the appendix. Choose signs in a way that

$$\sum_{uv \in E(G)} \left( \arcsin \left( \frac{\langle \mathbf{y}^u, \mathbf{y}^v \rangle}{\|\mathbf{y}^u\| \|\mathbf{y}^v\|} \right) - \arcsin \left( \frac{\langle \mathbf{x}^u, \mathbf{x}^v \rangle}{\|\mathbf{x}^u\| \|\mathbf{x}^v\|} \right) \right) \leq \frac{\gamma n \sqrt{d}}{4} - \gamma^2 \left( \sum_{\substack{uv \in E(G): \\ d(u, v) > 20d^2/n}} \frac{d(u, v)}{10d} \right).$$

By Lemma 2.2, we get

$$\begin{aligned} \text{sp}(G) &\geq -\frac{1}{\pi} \sum_{uv \in E(G)} \arcsin \left( \frac{\langle \mathbf{y}^u, \mathbf{y}^v \rangle}{\|\mathbf{y}^u\| \|\mathbf{y}^v\|} \right) \\ &\geq -\frac{1}{\pi} \left( \sum_{uv \in E(G)} \arcsin \left( \frac{\langle \mathbf{x}^u, \mathbf{x}^v \rangle}{\|\mathbf{x}^u\| \|\mathbf{x}^v\|} \right) + \frac{\gamma n \sqrt{d}}{4} - \gamma^2 \left( \sum_{\substack{uv \in E(G): \\ d(u, v) > 20d^2/n}} \frac{d(u, v)}{10d} \right) \right). \end{aligned} \quad (4.1)$$

We can use this inequality to prove the following lemma, which is a lower bound on the surplus of a graph in terms of the number of triangles and the codegrees. Recall that for vertices  $u, v$ , we write  $\delta(u, v) = d(u, v) - d^2/n$ . Let  $\delta_+(u, v)$  be equal to  $\delta(u, v)$  when  $\delta(u, v) > 0$ , and otherwise let  $\delta_+(u, v) = 0$ .

**Lemma 4.2.** *Let  $0 < \gamma \leq 1/10$  and  $d \leq n/2$ . If  $G$  is an  $n$ -vertex  $d$ -regular graph with  $d^3/6 + s$  triangles where  $s \leq d^3/6$ , then*

$$\text{sp}(G) \geq \Theta(\gamma n \sqrt{d}) - \Theta\left(\frac{\gamma^2 s}{d}\right) - \Theta\left(\gamma^2 \sum_{\substack{uv \in E(G): \\ d(u,v) \leq 20d^2/n}} \frac{\delta_+(u,v)^3}{d^3}\right).$$

The proof is again given in the appendix. Our goal is now to remove the ‘‘error term’’ in the above lemma and prove the following stronger result.

**Lemma 4.3.** *Let  $0 < \gamma \leq 1/10$  and  $d \leq n/2$ . If  $G$  is an  $n$ -vertex  $d$ -regular graph with  $d^3/6 + s$  triangles where  $s \leq d^3/6$ , then  $\text{sp}(G) \geq \Theta(\gamma n \sqrt{d}) - \Theta\left(\frac{\gamma^2 s}{d}\right)$ .*

This is a significant sharpening (in the regular case) of Corollary 1.2 from [6], which states that if  $G$  is a  $d$ -degenerate graph with  $m$  edges and  $t$  triangles, then for any  $\gamma \leq 1$ , we have  $\text{sp}(G) \geq \frac{\gamma m}{2\pi\sqrt{d}} - \frac{\gamma^2 t}{2d}$ . Note that  $s$  is much smaller than  $t$ , and can be negative. Given Lemma 4.3, it is easy to deduce Theorem 1.3.

**Proof of Theorem 1.3.** We claim that for any  $0 < \gamma \leq 1/10$ ,

$$\text{sp}(G) \geq \Theta(\gamma n \sqrt{d}) - \Theta\left(\frac{\gamma^2 s}{d}\right). \quad (4.2)$$

When  $s \leq d^3/6$ , this follows from Lemma 4.3. Otherwise,  $s = \Theta(t(G))$  and we can use the aforementioned Corollary 1.2 from [6] to obtain the desired inequality.

When  $s < -nd^{3/2}$ , take  $\gamma = 1/10$  and use (4.2) to get  $\text{sp}(G) = \Omega(|s|/d)$ .

When  $-nd^{3/2} \leq s \leq nd^{3/2}$ , take  $\gamma$  to be a sufficiently small constant and use (4.2) to get  $\text{sp}(G) = \Omega(n\sqrt{d})$ .

Finally, when  $s > nd^{3/2}$ , take  $\gamma = c\frac{nd^{3/2}}{s}$  for a sufficiently small constant  $c$  and use (4.2) to get  $\text{sp}(G) = \Omega(n^2 d^2/s)$ .  $\square$

The rest of this section is devoted to the proof of Lemma 4.3.

**Lemma 4.4.** *Let  $H$  be a graph on  $N$  vertices. Let  $D > 0$  be a real number. For  $z \in V(H)$ , let  $\Delta_+(z) = d_H(z) - D$  if  $d_H(z) > D$ , and otherwise let  $\Delta_+(z) = 0$ . Then there exists a positive absolute constant  $c$  with the following property. If*

$$e(H) \leq \frac{ND}{2} + \frac{c}{D^2} \sum_{\substack{z \in V(H): \\ d_H(z) \leq 20D}} \Delta_+(z)^3, \quad (4.3)$$

then there exists some  $T \subset V(H)$  such that

$$e_H(T) \leq \frac{|T|^2 D}{2N} - \frac{c}{D^2} \sum_{\substack{z \in V(H): \\ d_H(z) \leq 20D}} \Delta_+(z)^3.$$

**Proof.** Let  $c = \frac{1}{12 \cdot 40^2}$  and let  $H$  be a graph on  $N$  vertices satisfying (4.3). Write

$$q = \sum_{\substack{z \in V(H): \\ d_H(z) \leq 20D}} \Delta_+(z)^3.$$

Then

$$e(H) \leq \frac{ND}{2} + \frac{c}{D^2} q. \quad (4.4)$$

Observe that there exists a positive integer  $i$  such that

$$\sum_{z: 2^{-i}20D < \Delta_+(z) \leq 2^{-(i-1)}20D} \Delta_+(z)^3 \geq q/2^i.$$

Letting  $S$  be the set of  $z \in V(H)$  with  $2^{-i}20D < \Delta_+(z) \leq 2^{-(i-1)}20D$ , we get

$$|S| \geq \frac{q/2^i}{(2^{-(i-1)}20D)^3} = \frac{2^{2i}q}{40^3 D^3}. \quad (4.5)$$

Let  $p = 2^{-i}$  and let  $S'$  be the random subset of  $S$  obtained by keeping each vertex of  $S$  with probability  $p$ , independently of the other vertices. Let  $T = V(H) \setminus S'$ .

*Claim 1.*

$$\mathbb{E}[e_H(T)] \leq \frac{\mathbb{E}[|T|^2] D}{2} \frac{1}{N} - \frac{c}{D^2} q. \quad (4.6)$$

This suffices since then we can find  $T \subset V(H)$  such that  $e_H(T) \leq \frac{|T|^2 D}{2} \frac{1}{N} - \frac{c}{D^2} q$ .

*Proof of claim:* For a set  $X \subset V(H)$ , write  $e(X)$  for the number of edges in  $H[X]$ . For disjoint sets  $X$  and  $Y$ , write  $e(X, Y)$  for the number of edges between  $X$  and  $Y$  in  $H$ .

We start by giving an upper bound for  $\mathbb{E}[e(T)]$ . Let  $R = V(H) \setminus S$ . Since  $T = V(H) \setminus S'$ , we have  $e(T) = e(H) - e(T, S') - e(S') = e(H) - e(R, S') - e(S \setminus S', S') - e(S')$ . Taking expectations, we get

$$\begin{aligned} \mathbb{E}[e(T)] &= e(H) - \mathbb{E}[e(R, S')] - \mathbb{E}[e(S \setminus S', S')] - \mathbb{E}[e(S')] \\ &= e(H) - pe(R, S) - 2(1-p)pe(S) - p^2e(S) \\ &= e(H) - p(e(R, S) + 2e(S)) + p^2e(S) \\ &= e(H) - p \left( \sum_{z \in S} d_H(z) \right) + p^2e(S) \\ &\leq e(H) - p|S|(D + 2^{-i}20D) + p^2e(S) \\ &\leq e(H) - p|S|D - p2^{-i}|S|20D + p^2 \frac{|S|21D}{2} \\ &\leq e(H) - p|S|D - \frac{1}{3}2^{-2i}|S|20D, \\ &\leq \frac{ND}{2} + \frac{c}{D^2}q - p|S|D - \frac{1}{3}2^{-2i}|S|20D, \end{aligned} \quad (4.7)$$

where the first inequality follows from the fact every  $z \in S$  has  $\Delta_+(z) > 2^{-i}20D$ , the second inequality holds since  $\Delta_+(z) \leq 2^{-(i-1)}20D \leq 20D$  for every  $z \in S$ , and the last inequality is true because of (4.4). By (4.5), we have  $\frac{1}{3}2^{-2i}|S|20D \geq \frac{q}{6 \cdot 40^2 D^2} = \frac{2c}{D^2}q$ . Hence, by (4.7) we have

$$\mathbb{E}[e(T)] \leq \frac{ND}{2} - p|S|D - \frac{c}{D^2}q.$$

On the other hand,  $\mathbb{E}[|T|^2] \geq \mathbb{E}[|T|]^2 = (N - p|S|)^2 \geq N^2 - 2p|S|N$ , so

$$\frac{\mathbb{E}[|T|^2] D}{2} \frac{1}{N} \geq \frac{ND}{2} - p|S|D,$$

which completes the proof of the claim.  $\square$

Finally, we need the following lemma.

**Lemma 4.5.** *Let  $G$  be a  $d$ -regular graph on  $n$  vertices and let  $S \subset V(G)$  be a subset. Then  $\text{sp}(G) \geq \frac{1}{2} \left( \frac{|S|^2 d}{2n} - e_G(S) \right)$ .*

**Proof.** Let  $T$  be a random subset of  $V(G)$  such that each  $v \in V(G)$  belongs to  $T$  with probability  $|S|/n$ , independently of the other vertices. Write  $e_G(S, T)$  for the number of pairs  $(u, v) \in S \times T$  with  $uv \in E(G)$ . Clearly,  $\mathbb{E}[e_G(S, T)] = |S|d \cdot \frac{|S|}{n}$ . Moreover,  $\mathbb{E}[e_G(T)] = \frac{nd}{2} \cdot \frac{|S|^2}{n^2}$ . Hence,

$$\mathbb{E}[e_G(S, T) - e_G(S) - e_G(T)] = \frac{|S|^2 d}{2n} - e_G(S).$$

So we may choose  $T$  in a way that  $e_G(S, T) - e_G(S) - e_G(T) \geq \frac{|S|^2 d}{2n} - e_G(S)$ . If  $S$  and  $T$  are disjoint, then we get the desired surplus in  $G[S \cup T]$ . Otherwise, one by one for each vertex  $v$  which appears both in  $S$  and  $T$ , we can remove  $v$  from one of the two remaining sets (to be precise, from the one in which it has more neighbours) and not decrease the surplus corresponding to the cut.  $\square$

**Proof of Lemma 4.3.** Let  $c$  be the constant provided by Lemma 4.4. If

$$\sum_{\substack{uv \in E(G): \\ d_G(u, v) \leq 20d^2/n}} \delta_+(u, v)^3 < \frac{3}{2c} sd^2,$$

then the result follows from Lemma 4.2. Assume that

$$\sum_{\substack{uv \in E(G): \\ d_G(u, v) \leq 20d^2/n}} \delta_+(u, v)^3 \geq \frac{3}{2c} sd^2. \quad (4.8)$$

Define an auxiliary graph  $H$  as follows. Let  $V(H) = \{(u, v) \in V(G)^2 : uv \in E(G)\}$  and take an edge between  $(u, v)$  and  $(u, w)$  if  $vw \in E(G)$ . Let all other vertex pairs be non-edges. Observe that  $e(H) = 3t(G) = d^3/2 + 3s$ . Let  $N = |V(H)| = nd$  and  $D = \frac{d^2}{n}$ . Note that for any  $z = (u, v) \in V(H)$ , we have  $d_H(z) = d_G(u, v)$ . Moreover, if  $\Delta_+(z)$  is defined as in Lemma 4.4, then for  $z = (u, v)$ , we have  $\Delta_+(z) = \delta_+(u, v)$ . Hence, (4.8) translates to

$$\sum_{\substack{z \in V(H): \\ d_H(z) \leq 20D}} \Delta_+(z)^3 \geq \frac{3}{c} sd^2,$$

since an edge  $uv \in E(G)$  corresponds to both  $(u, v) \in V(H)$  and  $(v, u) \in V(H)$ . Thus,

$$\frac{c}{D^2} \sum_{\substack{z \in V(H): \\ d_H(z) \leq 20D}} \Delta_+(z)^3 \geq \frac{3sd^2}{D^2} \geq 3s,$$

and hence

$$e(H) = d^3/2 + 3s \leq \frac{ND}{2} + \frac{c}{D^2} \sum_{\substack{z \in V(H): \\ d_H(z) \leq 20D}} \Delta_+(z)^3.$$

Therefore, by Lemma 4.4, there exists  $T \subset V(H)$  such that

$$e_H(T) \leq \frac{|T|^2 D}{2N} - \frac{c}{D^2} \sum_{\substack{z \in V(H): \\ d_H(z) \leq 20D}} \Delta_+(z)^3.$$

For every  $u \in V(G)$ , let  $T_u = \{(u, v) \in T\}$ . Note that  $\sum_{u \in V(G)} |T_u|^2 \geq \frac{(\sum_{u \in V(G)} |T_u|)^2}{n} = \frac{|T|^2}{n}$ . Moreover, there are no edges in  $H$  between  $T_u$  and  $T_v$  when  $u \neq v$ , so  $e_H(T) = \sum_{u \in V(G)} e_H(T_u)$ . Thus,

$$\sum_{u \in V(G)} e_H(T_u) \leq \sum_{u \in V(G)} \frac{|T_u|^2 nD}{2N} - \frac{c}{D^2} \sum_{\substack{z \in V(H): \\ d_H(z) \leq 20D}} \Delta_+(z)^3.$$

Therefore there exists  $w \in V(G)$  such that

$$e_H(T_w) \leq \frac{|T_w|^2 nD}{2N} - \frac{c}{nD^2} \sum_{\substack{z \in V(H): \\ d_H(z) \leq 20D}} \Delta_+(z)^3 = \frac{|T_w|^2 d}{2n} - \frac{2cn}{d^4} \sum_{\substack{uv \in E(G): \\ d_G(u,v) \leq 20d^2/n}} \delta_+(u, v)^3.$$

This means that there exists a set  $S \subset N_G(w)$  such that

$$e_G(S) \leq \frac{|S|^2 d}{2n} - \frac{2cn}{d^4} \sum_{\substack{uv \in E(G): \\ d_G(u,v) \leq 20d^2/n}} \delta_+(u, v)^3.$$

Hence, by Lemma 4.5,

$$\text{sp}(G) \geq \frac{cn}{d^4} \left( \sum_{\substack{uv \in E(G): \\ d_G(u,v) \leq 20d^2/n}} \delta_+(u, v)^3 \right) \geq \Omega \left( \gamma^2 \sum_{\substack{uv \in E(G): \\ d(u,v) \leq 20d^2/n}} \frac{\delta_+(u, v)^3}{d^3} \right).$$

This, combined with Lemma 4.2, implies the result.  $\square$

## 5 Cliques

In this section, we prove Theorems 1.5, 1.7 and 1.8. We start with the proof of Theorem 1.8, as the proof of the other two results will use its weaker version Theorem 1.6.

**Proof of Theorem 1.8.** We prove the theorem by induction on  $r$ . The case  $r = 3$  is Theorem 1.2. Assume that  $r \geq 4$ . First observe that we may assume that  $G$  has maximum degree at most  $Cd$  for some  $C$ , depending only on  $\epsilon$  and  $r$ . Indeed, we may apply Lemma 2.6 with  $\epsilon/r$  in place of  $\epsilon$ ,  $\alpha = -(r-3)$  and  $\beta = r-1$  to find an induced subgraph  $\tilde{G}$  with  $\tilde{n}$  vertices and average degree  $\tilde{d}$  such that  $\frac{\tilde{d}^{r-1}}{\tilde{n}^{r-3}} \geq (1-\epsilon/r) \frac{d^{r-1}}{n^{r-3}}$  and  $\Delta(\tilde{G}) \leq C\tilde{d}$  (if such subgraph does not exist, then  $G$  has surplus  $\Omega_{\epsilon,r}(d^{r-1}/n^{r-3})$  by the lemma). Now

$$\tilde{n}^r (\tilde{d}/\tilde{n})^{\binom{r}{2}} = (\tilde{d}^{r-1}/\tilde{n}^{r-3})^{r/2} \geq ((1-\epsilon/r)d^{r-1}/n^{r-3})^{r/2} = (1-\epsilon/r)^{r/2} n^r (d/n)^{\binom{r}{2}} \geq (1-\epsilon/2)n^r (d/n)^{\binom{r}{2}},$$

so the number of  $K_r$ 's in  $\tilde{G}$  is at most  $(1-\epsilon/2) \frac{\tilde{n}^r}{r!} (\tilde{d}/\tilde{n})^{\binom{r}{2}}$ .

Thus, by replacing  $G$  by  $\tilde{G}$  and  $\epsilon$  by  $\epsilon/2$  if necessary, we may assume that  $\Delta(G) \leq Cd$  for some  $C = C(\epsilon, r)$ .

If  $G$  has at most  $(1-\frac{\epsilon}{4r^2}) \frac{d}{6n} \sum_{v \in V(G)} d(v)^2$  triangles, then by Theorem 2.7,  $G$  has surplus  $\Omega_{\epsilon,r,C}(d^2)$  and we are done since  $d^2 \geq d^{r-1}/n^{r-3}$ . So assume that

$$t(G) \geq \left(1 - \frac{\epsilon}{4r^2}\right) \frac{d}{6n} \sum_{v \in V(G)} d(v)^2. \quad (5.1)$$

Let  $k = \lceil \frac{\epsilon}{8Cr^2} \frac{n}{d} \rceil$  and let  $v_1, v_2, \dots, v_k$  be random vertices chosen from  $V(G)$  with replacement. For every  $1 \leq i \leq k$ , let  $A_i = N(v_i) \setminus (\cup_{j \neq i} N(v_j))$ .

Fix  $1 \leq i \leq k$ . Let  $X$  be the number of edges in  $G[A_i]$  and let  $Y$  be the number of copies of  $K_{r-1}$  in  $G[A_i]$ .

*Claim 1.*  $\mathbb{E}[X] \geq (1 - \frac{\epsilon}{4r^2}) \frac{3t(G)}{n}$ .

*Proof of claim:* Let  $xy \in E(G)$ . We have  $xy \in E(G[A_i])$  precisely if  $x, y \in N(v_i) \setminus (\cup_{j \neq i} N(v_j))$ , i.e. if  $v_i \in N(x) \cap N(y)$ , but  $v_j \notin N(x) \cup N(y)$  for every  $j \neq i$ . Since  $|N(x) \cup N(y)| \leq 2Cd$ , we get

$$\mathbb{P}(xy \in E(G[A_i])) \geq \frac{d(x, y)}{n} \left(1 - \frac{2Cd}{n}\right)^{k-1} \geq \frac{d(x, y)}{n} \left(1 - (k-1) \frac{2Cd}{n}\right) \geq \left(1 - \frac{\epsilon}{4r^2}\right) \frac{d(x, y)}{n}.$$

Hence,

$$\mathbb{E}[X] \geq \sum_{xy \in E(G)} \left(1 - \frac{\epsilon}{4r^2}\right) \frac{d(x, y)}{n} = \left(1 - \frac{\epsilon}{4r^2}\right) \frac{3 \cdot t(G)}{n}.$$

◇

*Claim 2.*  $\mathbb{E}[Y] \leq (1 - \epsilon) \frac{n^{r-1}}{(r-1)!} (d/n)^{\binom{r}{2}}$ .

*Proof of claim:* Let  $x_1 x_2 \dots x_{r-1}$  be a copy of  $K_{r-1}$  in  $G$ . Then, writing  $d(x_1, \dots, x_{r-1})$  for the number of common neighbours of  $x_1, \dots, x_{r-1}$ , we have  $\mathbb{P}(x_1, \dots, x_{r-1} \in A_i) \leq \frac{d(x_1, \dots, x_{r-1})}{n}$ . Since the sum of  $d(x_1, \dots, x_{r-1})$  over all  $(r-1)$ -cliques  $x_1 \dots x_{r-1}$  is  $r$  times the number of  $K_r$ 's, and the number of  $K_r$ 's in  $G$  is at most  $(1 - \epsilon) \frac{n^r}{r!} (d/n)^{\binom{r}{2}}$ , we get the desired bound. ◇

For simplicity, write  $A = A_i$ . Note that  $G[A]$  has average degree  $\frac{2X}{|A|}$ . By the induction hypothesis, there exists a positive  $c' = c'(r, \epsilon)$  such that if  $Y \leq (1 - \epsilon/2) \frac{|A|^{r-1}}{(r-1)!} (2X/|A|^2)^{\binom{r-1}{2}}$ , then  $G[A]$  has surplus at least  $c'(2X)^{r-2}/|A|^{2r-6}$ . On the other hand, since  $|A|^{r-1} (2X/|A|^2)^{\binom{r-1}{2}} = ((2X)^{r-2}/|A|^{2r-6})^{\frac{r-1}{2}}$ , if

$$Y \geq (1 - \epsilon/2) \frac{|A|^{r-1}}{(r-1)!} (2X/|A|^2)^{\binom{r-1}{2}},$$

then

$$c' \left( \frac{(r-1)! \cdot Y}{1 - \epsilon/2} \right)^{\frac{2}{r-1}} \geq c'(2X)^{r-2}/|A|^{2r-6}.$$

Hence, in both cases we have

$$\text{sp}(G[A]) \geq c'(2X)^{r-2}/|A|^{2r-6} - c' \left( \frac{(r-1)! \cdot Y}{1 - \epsilon/2} \right)^{\frac{2}{r-1}}.$$

Taking expectations, we get

$$\mathbb{E}[\text{sp}(G[A])] \geq c' \left( 2^{r-2} \mathbb{E}[X^{r-2}/|A|^{2r-6}] - \left( \frac{(r-1)!}{1 - \epsilon/2} \right)^{\frac{2}{r-1}} \mathbb{E}[Y^{\frac{2}{r-1}}] \right). \quad (5.2)$$

By Hölder's inequality, we have

$$\mathbb{E} \left[ \frac{X^{r-2}}{|A|^{2r-6}} \right]^{\frac{1}{r-2}} \mathbb{E} [|A|^2]^{\frac{r-3}{r-2}} \geq \mathbb{E} \left[ \left( \frac{X^{r-2}}{|A|^{2r-6}} \right)^{\frac{1}{r-2}} |A|^{\frac{2(r-3)}{r-2}} \right] = \mathbb{E}[X],$$

so

$$\mathbb{E} \left[ \frac{X^{r-2}}{|A|^{2r-6}} \right] \geq \frac{\mathbb{E}[X]^{r-2}}{\mathbb{E}[|A|^2]^{r-3}}.$$

Observe that  $\mathbb{E}[|A|^2] \leq \frac{1}{n} \sum_{v \in V(G)} d(v)^2$ . This inequality is the reason why we used Theorem 2.7 instead of Theorem 1.2 in our argument. By Claim 1 and (5.1),

$$\mathbb{E}[X] \geq \left(1 - \frac{\epsilon}{4r^2}\right) \frac{3t(G)}{n} \geq \left(1 - \frac{\epsilon}{4r^2}\right)^2 \frac{d}{2n^2} \sum_{v \in V(G)} d(v)^2 \geq \left(1 - \frac{\epsilon}{4r^2}\right)^2 \frac{d}{2n} \mathbb{E}[|A|^2].$$

Thus, using Claim 1, (5.1) and  $\sum_v d(v)^2 \geq nd^2$ ,

$$\begin{aligned} \frac{\mathbb{E}[X]^{r-2}}{\mathbb{E}[|A|^2]^{r-3}} &\geq \left( \left(1 - \frac{\epsilon}{4r^2}\right)^2 \frac{d}{2n} \right)^{r-3} \mathbb{E}[X] \geq \left( \left(1 - \frac{\epsilon}{4r^2}\right)^2 \frac{d}{2n} \right)^{r-3} \left(1 - \frac{\epsilon}{4r^2}\right) \frac{3t(G)}{n} \\ &\geq \left( \left(1 - \frac{\epsilon}{4r^2}\right)^2 \frac{d}{2n} \right)^{r-3} \left(1 - \frac{\epsilon}{4r^2}\right)^2 \frac{d^3}{2n} \geq \left(1 - \frac{\epsilon}{2r}\right) \frac{d^r}{(2n)^{r-2}}, \end{aligned}$$

so

$$\mathbb{E} \left[ \frac{X^{r-2}}{|A|^{2r-6}} \right] \geq \left(1 - \frac{\epsilon}{2r}\right) \frac{d^r}{(2n)^{r-2}}. \quad (5.3)$$

On the other hand, by Claim 2, we have

$$\mathbb{E} \left[ Y^{\frac{2}{r-1}} \right] \leq \mathbb{E} [Y]^{\frac{2}{r-1}} \leq \left( \frac{1-\epsilon}{(r-1)!} \right)^{\frac{2}{r-1}} \frac{d^r}{n^{r-2}}.$$

Substituting this and equation (5.3) into equation (5.2), we get

$$\mathbb{E} [\text{sp}(G[A])] \geq c' \left( \left(1 - \frac{\epsilon}{2r}\right) \frac{d^r}{n^{r-2}} - \left( \frac{1-\epsilon}{1-\epsilon/2} \right)^{\frac{2}{r-1}} \frac{d^r}{n^{r-2}} \right).$$

Note that  $\left( \frac{1-\epsilon}{1-\epsilon/2} \right)^{\frac{2}{r-1}} \leq (1-\epsilon/2)^{\frac{2}{r-1}} \leq 1 - \epsilon/(r-1) \leq 1 - \epsilon/r$ , so we conclude that

$$\mathbb{E} [\text{sp}(G[A])] \geq c' \frac{\epsilon}{2r} \frac{d^r}{n^{r-2}}.$$

Since  $\text{sp}(G) \geq \sum_{i=1}^k \text{sp}(G[A_i])$  by Lemma 2.1, we get

$$\text{sp}(G) \geq k \mathbb{E} [\text{sp}(G[A])] \geq \frac{\epsilon}{8Cr^2} c' \frac{\epsilon}{2r} \frac{d^{r-1}}{n^{r-3}},$$

which completes the proof.  $\square$

It remains to prove Theorems 1.5 and 1.7. We prove them simultaneously using induction on  $r$ . Carlson et al. [6] provided a general tool that allows to convert results on the surplus formulated purely in terms of the number of edges (as in Theorem 1.7) into results that hold for degenerate graphs (as in Theorem 1.5). In particular, a special case of their Lemma 3.5 is the following: If  $\text{sp}(m, K_{r-1}) = \Omega_r(m^a)$  for some  $a = a(r) \in [\frac{1}{2}, 1]$ , then any  $d$ -degenerate graph  $G$  with  $m \geq 1$  edges has surplus

$$\Omega_r(m/d^{(2-a)/(1+a)}). \quad (5.4)$$

**Proof of Theorems 1.5 and 1.7.** We know that both theorems hold for  $r = 3$ . Assume now that  $r \geq 4$  and that both theorems hold for  $r-1$ . It follows from (5.4) (applied with  $a = \frac{1}{2} + \frac{3}{4(r-1)-2} = \frac{r}{2r-3}$ ) that Theorem 1.5 holds for  $r$ . We now use this additional information to prove Theorem 1.7 for  $r$ .

Let  $G$  be any  $K_r$ -free graph with  $m$  edges. Let  $d = m^{\frac{r-1}{2r-1}}$ . Assume first that  $G$  is  $d$ -degenerate. Then by Theorem 1.5,  $G$  has surplus  $\Omega_r(md^{-\frac{r-2}{r-1}}) = \Omega_r(m^{1-\frac{r-2}{2r-1}}) = \Omega_r(m^{\frac{r+1}{2r-1}})$ , which is the desired bound. Otherwise,  $G$  has a non-empty induced subgraph  $H$  with minimum degree at least  $d$ . If the number of vertices in  $H$  is  $n$ , then clearly  $nd \leq 2e(H) \leq 2e(G) = 2m$ , so  $n \leq \frac{2m}{d} = 2m^{\frac{r}{2r-1}}$ . Thus, by Theorem 1.6,  $H$  has surplus

$$\Omega_r(d^{r-1}/n^{r-3}) \geq \Omega_r \left( m^{\frac{(r-1)^2}{2r-1} - \frac{r(r-3)}{2r-1}} \right) = \Omega_r(m^{\frac{r+1}{2r-1}}),$$

which is again the desired bound.  $\square$

## 6 Concluding remarks

- We proved that for any odd  $r \geq 3$ ,  $\text{sp}(m, C_r) = \Theta_r(m^{(r+1)/(r+2)})$ . The same function has been shown to be a lower bound when  $r$  is even, but a matching upper bound is only known for  $r \in \{4, 6, 10\}$  (see [5]). A related open problem is to construct graphs with  $m$  edges and girth at least  $r + 1$  which have surplus  $O_r(m^{(r+1)/(r+2)})$ . Such examples are only known for  $r = 3$  and  $r = 4$  (see [2] and [3]).
- We obtained sharp bounds for the surplus of regular graphs as a function of the order, degree and number of triangles. As we remarked after Theorem 1.3, it is easy to find examples for which the statement of the theorem is not true when  $G$  is not  $d$ -regular, but has average degree  $d$ . However, it seems plausible that the first bound in Theorem 1.3 still holds in this more general setting; that is, if  $G$  is an  $n$ -vertex graph with average degree  $d \leq n/2$  and at most  $d^3/6 - s$  triangles for some  $s > 0$ , then  $\text{sp}(G) = \Omega(s/d)$ . This would strengthen Theorem 1.2 in the case where  $\epsilon$  is not constant. By a result of Nikiforov [20] mentioned at the end of Section 1.3, such graphs have  $|\lambda_{\min}| = \Omega(\frac{s}{nd})$ , providing some evidence towards  $\text{sp}(G) = \Omega(s/d)$ .
- We also proved that any  $K_r$ -free graph with  $m$  edges has surplus  $\Omega_r(m^{\frac{1}{2} + \frac{3}{4r-2}})$ . Arguably, the main open problem is to decide whether there exists a positive absolute constant  $\epsilon$  such that any  $K_r$ -free graph with  $m$  edges has surplus  $\Omega_r(m^{1/2+\epsilon})$ . It seems that the bound  $\Omega_r(m^{\frac{1}{2} + \frac{3}{4r-2}})$  is the best our methods can give. We think that any improvement on it, even just beating the exponent  $5/7$  in the  $r = 4$  case, would be interesting.
- Another natural infinite sequence of graphs to consider is the set of complete bipartite graphs  $K_{r,r}$ . Forbidding bipartite subgraphs puts more restrictions on the host graph, so it is potentially easier to show that there exists an absolute constant  $\epsilon > 0$  such that any  $K_{r,r}$ -free graph with  $m$  edges has surplus  $\Omega_r(m^{1/2+\epsilon})$ . Nevertheless, it is likely that this requires new ideas. See [5] for the correct exponent for  $K_{r,s}$  when  $r \in \{2, 3\}$ , and a conjecture for the general case.
- Conlon, Fox, Kwan and Sudakov [7] proved analogues of Edwards's result for hypergraphs. It would be interesting to further investigate the MaxCut problem in this setting and to obtain the tight bounds for the surplus.
- Finally, we remark that if Conjecture 1.4 and the conjecture of Alon, Bollobás, Krivelevich and Sudakov stating  $\text{sp}(m, K_r) \geq c_r m^{3/4+\epsilon_r}$  are not true, then a possible way of disproving them is by finding a  $K_r$ -free strongly regular graph with many triangles. To be more precise, by Lemma A.1, a strongly regular graph with degree  $d$  and  $d^3/6 + \omega(nd^{3/2})$  triangles has smallest eigenvalue  $|\lambda_{\min}| = o(\sqrt{d})$ , which implies that it has surplus  $o(n\sqrt{d})$ . If a  $K_r$ -free graph exists (for some fixed  $r$ ) with these properties, it disproves Conjecture 1.4.

## References

- [1] N. Alon, *Explicit Ramsey graphs and orthonormal labelings*, Electron. J. Combin. **1** (1994), Research Paper 12, 8 pages.
- [2] ———, *Bipartite subgraphs*, Combinatorica **16** (1996), 301–311.
- [3] N. Alon, B. Bollobás, M. Krivelevich, and B. Sudakov, *Maximum cuts and judicious partitions in graphs without short cycles*, J. Combin. Theory Ser. B **88** (2003), 329–346.
- [4] N. Alon and N. Kahale, *Approximating the independence number via the  $\theta$ -function*, Math. Programming **80** (1998), 253–264.
- [5] N. Alon, M. Krivelevich, and B. Sudakov, *MaxCut in  $H$ -free graphs*, Combin. Probab. Comput. **14** (2005), 629–647.

- [6] C. Carlson, A. Kolla, R. Li, N. Mani, B. Sudakov, and L. Trevisan, *Lower bounds for Max-Cut in  $H$ -free graphs via semidefinite programming*, SIAM J. Discrete Math. **35** (2021), 1557–1568.
- [7] D. Conlon, J. Fox, M. Kwan and B. Sudakov, *Hypergraph cuts above the average*, Israel J. Math. **233** (2019), 67–111.
- [8] C. Delorme and S. Poljak, *Laplacian eigenvalues and the maximum cut problem*, Math. Programming **62** (1993), 557–574.
- [9] C. S. Edwards, *Some extremal properties of bipartite subgraphs*, Canadian J. Math. **25** (1973), 475–485.
- [10] ———, *An improved lower bound for the number of edges in a largest bipartite subgraph*, Recent advances in graph theory (Proc. Second Czechoslovak Sympos., Prague, 1974), 1975, pp. 167–181.
- [11] P. Erdős, *On even subgraphs of graphs*, Mat. Lapok **18** (1967), 283–288.
- [12] ———, *Problems and results in graph theory and combinatorial analysis*, Graph theory and related topics (Proc. Conf., Univ. Waterloo, Waterloo, 1977), Academic Press, 1979, pp. 153–163.
- [13] P. Erdős and A. Rényi, *On a problem in the theory of graphs*, Publ. Math. Inst. Hungar. Acad. Sci. **7** (1962), 215–235.
- [14] J. Fox, Z. Himwich, and N. Mani, *Making an  $H$ -free graph  $k$ -colorable*, J. Graph Theory **102** (2023), 234–261.
- [15] M. X. Goemans and D. P. Williamson, *Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming*, J. ACM **42** (1995), 1115–1145.
- [16] M. Krivelevich and B. Sudakov, *Pseudo-random graphs*, More sets, graphs and numbers, Bolyai Soc. Math. Stud. 15, Springer, Berlin, 2006, pp. 199–262.
- [17] B. Mohar and S. Poljak, *Eigenvalues and the max-cut problem*, Czechoslovak Math. J. **40** (1990), 343–352.
- [18] D. Mubayi and J. Verstraëte, *A note on pseudorandom Ramsey graphs*, arXiv:1909.01461.
- [19] V. Nikiforov, *The smallest eigenvalue of  $K_r$ -free graphs*, Discrete Math. **306** (2006), 612–616.
- [20] ———, *Eigenvalues and forbidden subgraphs I*, Linear Algebra Appl. **422** (2007), 284–290.
- [21] S. Poljak and Zs. Tuza, *Bipartite subgraphs of triangle-free graphs*, SIAM J. Discrete Math. **7** (1994), 307–313.
- [22] J. B. Shearer, *A note on bipartite subgraphs of triangle-free graphs*, Random Structures Algorithms **3** (1992), 223–226.
- [23] B. Sudakov, T. Szabó, and V. H. Vu, *A generalization of Turán’s theorem*, J. Graph Theory **49** (2005), 187–195.
- [24] Q. Zeng and J. Hou, *Maximum cuts of graphs with forbidden cycles*, Ars Math. Contemp. **15** (2018), 147–160.

## A The smallest eigenvalue of strongly regular graphs

**Lemma A.1.** *Let  $G$  be a strongly regular graph with  $n$  vertices, degree  $d \leq 0.99n$  and  $d^3/6 + s$  triangles. Let  $\lambda_{\min}$  be the smallest eigenvalue of  $G$ .*

- If  $s < -nd^{3/2}$ , then  $|\lambda_{\min}| = O(\frac{|s|}{nd})$ .
- If  $-nd^{3/2} \leq s \leq nd^{3/2}$ , then  $|\lambda_{\min}| = O(d^{1/2})$ .
- If  $s > nd^{3/2}$ , then  $|\lambda_{\min}| = O(\frac{nd^2}{s})$ .

**Proof.** We can assume that  $n$  is sufficiently large, otherwise the statement is clear. Recall that

$$\lambda_{\min} = \frac{1}{2} \left( \eta - \mu - \sqrt{(\eta - \mu)^2 + 4(d - \mu)} \right), \quad (\text{A.1})$$

where  $\eta$  is the number of common neighbours of adjacent pairs and  $\mu$  is the number of common neighbours of non-adjacent pairs. Double-counting the number of (not necessarily induced) copies of  $K_{1,2}$  in  $G$ , we get

$$n \binom{d}{2} = \frac{nd}{2} \eta + \left( \binom{n}{2} - \frac{nd}{2} \right) \mu. \quad (\text{A.2})$$

Let  $\delta = \eta - \frac{d^2}{n}$  and  $\beta = \mu - \frac{d^2}{n}$ . After a straightforward algebraic manipulation of (A.2), we get

$$\beta = -\frac{d}{n-d-1} \delta - \frac{d(n-d)}{n(n-d-1)}.$$

Using  $d \leq 0.99n$ , we see that if  $|\delta| \leq 6d^{1/2}$ , then  $|\beta| = O(d^{1/2})$ ; if  $\delta < -6d^{1/2}$ , then  $\delta - \beta = -\Theta(|\delta|)$ ; and if  $\delta > 6d^{1/2}$ , then  $\delta - \beta = \Theta(\delta)$ .

Now if  $s < -nd^{3/2}$ , then by (2.2), we have  $\delta = \frac{6s}{nd} < -6d^{1/2}$ , so  $\eta - \mu = \delta - \beta = -\Theta(|\delta|) = -\Theta(\frac{|s|}{nd})$ . Thus, (A.1) gives  $\lambda_{\min} = -\Theta(\frac{|s|}{nd})$ .

If  $-nd^{3/2} \leq s \leq nd^{3/2}$ , then  $|\delta| = \frac{|6s|}{nd} \leq 6d^{1/2}$ , so  $|\eta - \mu| = |\delta - \beta| = O(d^{1/2})$ . Thus, (A.1) gives that  $|\lambda_{\min}| = O(d^{1/2})$ .

Finally, if  $s > nd^{3/2}$ , then  $\delta = \frac{6s}{nd} > 6d^{1/2}$ , so  $\eta - \mu = \delta - \beta = \Theta(\delta)$ . Note that

$$\sqrt{(\eta - \mu)^2 + 4(d - \mu)} \leq \eta - \mu + \frac{2(d - \mu)}{\eta - \mu},$$

so by (A.1) we have  $\lambda_{\min} \geq -\frac{d-\mu}{\eta-\mu} \geq -\frac{d}{\eta-\mu}$ . Thus,  $|\lambda_{\min}| = O(\frac{d}{\delta}) = O(\frac{nd^2}{s})$ .  $\square$

## B Proofs for Section 4

Before proving Lemma 4.1 and Lemma 4.2, we make a few computations which will be used in the proofs. Assume that  $d \leq n/2$  and  $0 < \gamma \leq 1/10$ , like in the setting of these lemmas. Recall that  $a = \frac{\sqrt{d}}{n-d}$ , which is at most  $\frac{2\sqrt{d}}{n}$  since  $d \leq n/2$ . Again using  $d \leq n/2$ , we have  $a \leq \frac{1}{\sqrt{d}}$ . In particular,  $\frac{1}{d} + \frac{2a}{\sqrt{d}} + a^2 \leq \frac{4}{d}$ .

Note also that if vectors  $\mathbf{x}^v$  and  $\mathbf{y}^v$  are defined as at the beginning of Section 4, then for every  $v \in V(G)$ ,  $\|\mathbf{x}^v\|^2 = \|\mathbf{y}^v\|^2 = (1 + \gamma a)^2 + d \frac{\gamma^2}{d} + (n-d-1)\gamma^2 a^2$ . Write  $q$  for this value and note that  $1 \leq q \leq (1 + \gamma \frac{1}{\sqrt{d}})^2 + \gamma^2 + (n-d-1)\gamma^2 \frac{d}{(n-d)^2} \leq 2$ .

**Proof of Lemma 4.1.** For a vertex  $v \in V(G)$ , write  $S(v)$  for the set of vertices  $u \in N(v)$  for which  $d(u, v) > 20d^2/n$ . Also write  $F(v) = V(G) \setminus N(v)$ . First note that for any  $uv \in E(G)$ , we have

$$\begin{aligned} |\langle \mathbf{x}^u, \mathbf{x}^v \rangle| &\leq \sum_{w \in V(G)} |\mathbf{x}_w^u| |\mathbf{x}_w^v| \leq 2(1 + \gamma a) \frac{\gamma}{\sqrt{d}} + d \frac{\gamma^2}{d} + (n-d)\gamma^2 a^2 + 2d\gamma^2 \frac{a}{\sqrt{d}} \\ &\leq 2 \frac{\gamma}{\sqrt{d}} + \gamma^2 \left( \frac{2}{n-d} + 1 + \frac{d}{n-d} + 2 \frac{d}{n-d} \right) \leq 2\gamma + 6\gamma^2 \leq 1/2. \end{aligned}$$

The same estimate holds for  $|\langle \mathbf{y}^u, \mathbf{y}^v \rangle|$ . On the other hand, we have  $\|\mathbf{x}^v\|, \|\mathbf{y}^v\| \geq 1$  for all  $v$ . Hence, for any  $uv \in E(G)$ ,  $\left| \frac{\langle \mathbf{x}^u, \mathbf{x}^v \rangle}{\|\mathbf{x}^u\| \|\mathbf{x}^v\|} \right| \leq 1/2$  and  $\left| \frac{\langle \mathbf{y}^u, \mathbf{y}^v \rangle}{\|\mathbf{y}^u\| \|\mathbf{y}^v\|} \right| \leq 1/2$ .

The derivative of  $\arcsin(x)$  is  $\frac{1}{\sqrt{1-x^2}}$  which is always at least 1, and for  $-1/2 \leq x \leq 1/2$  it is at most 2. Thus, by the mean value theorem, we have

$$\arcsin\left(\frac{\langle \mathbf{y}^u, \mathbf{y}^v \rangle}{\|\mathbf{y}^u\| \|\mathbf{y}^v\|}\right) - \arcsin\left(\frac{\langle \mathbf{x}^u, \mathbf{x}^v \rangle}{\|\mathbf{x}^u\| \|\mathbf{x}^v\|}\right) = \alpha \left( \frac{\langle \mathbf{y}^u, \mathbf{y}^v \rangle}{\|\mathbf{y}^u\| \|\mathbf{y}^v\|} - \frac{\langle \mathbf{x}^u, \mathbf{x}^v \rangle}{\|\mathbf{x}^u\| \|\mathbf{x}^v\|} \right) \quad (\text{B.1})$$

for some random variable  $1 \leq \alpha \leq 2$  ( $\alpha$  depends on the choices of signs in  $\mathbf{y}^u$  and  $\mathbf{y}^v$ ).

Write  $S(v)_+$  for the set of those  $u \in S(v)$  for which  $\mathbf{y}_u^v = \gamma/\sqrt{d}$ . Now if  $uv \in E(G)$ , then

$$\begin{aligned} \langle \mathbf{y}^u, \mathbf{y}^v \rangle - \langle \mathbf{x}^u, \mathbf{x}^v \rangle &= \mathbb{1}_{u \in S(v)_+} \frac{2\gamma}{\sqrt{d}} + \mathbb{1}_{v \in S(u)_+} \frac{2\gamma}{\sqrt{d}} + |(S(u)_+ \cap F(v)) \cup (S(v)_+ \cap F(u))| \frac{2\gamma^2 a}{\sqrt{d}} \\ &\quad - |(S(u)_+ \cap N(v)) \cup (S(v)_+ \cap N(u)) \setminus (S(u)_+ \cap S(v)_+)| \frac{2\gamma^2}{d}. \end{aligned}$$

Hence, using that  $1 \leq \alpha \leq 2$ ,

$$\begin{aligned} \alpha (\langle \mathbf{y}^u, \mathbf{y}^v \rangle - \langle \mathbf{x}^u, \mathbf{x}^v \rangle) &\leq \mathbb{1}_{u \in S(v)_+} \frac{4\gamma}{\sqrt{d}} + \mathbb{1}_{v \in S(u)_+} \frac{4\gamma}{\sqrt{d}} + |(S(u)_+ \cap F(v)) \cup (S(v)_+ \cap F(u))| \frac{4\gamma^2 a}{\sqrt{d}} \\ &\quad - |(S(u)_+ \cap N(v)) \cup (S(v)_+ \cap N(u)) \setminus (S(u)_+ \cap S(v)_+)| \frac{2\gamma^2}{d}. \end{aligned}$$

If  $w \in S(u)$ , then  $\mathbb{P}(w \in S(u)_+) = 1/2$ , so

$$\begin{aligned} \mathbb{E}[\alpha (\langle \mathbf{y}^u, \mathbf{y}^v \rangle - \langle \mathbf{x}^u, \mathbf{x}^v \rangle)] &\leq \mathbb{1}_{u \in S(v)} \frac{2\gamma}{\sqrt{d}} + \mathbb{1}_{v \in S(u)} \frac{2\gamma}{\sqrt{d}} + |(S(u) \cap F(v)) \cup (S(v) \cap F(u))| \frac{2\gamma^2 a}{\sqrt{d}} \\ &\quad - |(S(u) \cap N(v)) \cup (S(v) \cap N(u))| \frac{\gamma^2}{d} \\ &\leq \frac{4\gamma}{\sqrt{d}} \mathbb{1}_{d(u,v) > 20d^2/n} + \frac{2\gamma^2 a}{\sqrt{d}} (|S(u)| + |S(v)|) \\ &\quad - \frac{\gamma^2}{2d} (|S(u) \cap N(v)| + |S(v) \cap N(u)|). \end{aligned} \quad (\text{B.2})$$

Recall that  $\|\mathbf{x}^v\|^2 = \|\mathbf{y}^v\|^2 = q$  for every  $v$ , where  $1 \leq q \leq 2$ . Therefore, by equations (B.1) and (B.2), we have

$$\begin{aligned} (*) &:= \mathbb{E} \left[ \sum_{uv \in E(G)} \left( \arcsin \frac{\langle \mathbf{y}^u, \mathbf{y}^v \rangle}{\|\mathbf{y}^u\| \|\mathbf{y}^v\|} - \arcsin \frac{\langle \mathbf{x}^u, \mathbf{x}^v \rangle}{\|\mathbf{x}^u\| \|\mathbf{x}^v\|} \right) \right] \\ &\leq \frac{4\gamma}{\sqrt{d}} \sum_{uv \in E(G)} \mathbb{1}_{d(u,v) > 20d^2/n} + \frac{2\gamma^2 a}{\sqrt{d}} \sum_{uv \in E(G)} (|S(u)| + |S(v)|) \\ &\quad - \frac{1}{2} \cdot \frac{\gamma^2}{2d} \sum_{uv \in E(G)} (|S(u) \cap N(v)| + |S(v) \cap N(u)|). \end{aligned} \quad (\text{B.3})$$

Now note that

$$\begin{aligned} \sum_{uv \in E(G)} (|S(u) \cap N(v)| + |S(v) \cap N(u)|) &= |\{(u, v, w) \in V(G)^3 : uv, vw, wu \in E(G), d(u, w) > 20d^2/n\}| \\ &= 2 \sum_{\substack{uv \in E(G): \\ d(u,w) > 20d^2/n}} d(u, w). \end{aligned} \quad (\text{B.4})$$

On the other hand,

$$\sum_{uv \in E(G)} (|S(u)| + |S(v)|) = \sum_{u \in V(G)} d|S(u)| = \sum_{\substack{uv \in E(G): \\ d(u,v) > 20d^2/n}} 2d. \quad (\text{B.5})$$

Finally, as the number of triangles in  $G$  is at most  $d^3/3$ , the number of  $uv \in E(G)$  with  $d(u, v) > 20d^2/n$  is at most  $\frac{nd}{20}$ , so

$$\sum_{uv \in E(G)} \mathbb{1}_{d(u,v) > 20d^2/n} \leq \frac{nd}{20}. \quad (\text{B.6})$$

Thus, plugging (B.4), (B.5) and (B.6) into (B.3), we get

$$(*) \leq \gamma \frac{4n\sqrt{d}}{20} - \gamma^2 \left( \sum_{\substack{uv \in E(G): \\ d(u,v) > 20d^2/n}} \left( \frac{d(u,v)}{2d} - 4\sqrt{da} \right) \right) \leq \frac{\gamma n\sqrt{d}}{4} - \gamma^2 \left( \sum_{\substack{uv \in E(G): \\ d(u,v) > 20d^2/n}} \frac{d(u,v)}{10d} \right),$$

where in the second inequality we used that  $a \leq \frac{2\sqrt{d}}{n}$ .  $\square$

**Proof of Lemma 4.2.** Define the vectors  $\mathbf{x}^v$  as before. Looking at the Taylor series of  $\arcsin(x)$ , one can see that  $\arcsin(x) \leq x + 10x^3$  holds for all  $0 \leq x \leq 1$ . On the other hand,  $\arcsin(x) \leq x$  for  $-1 \leq x < 0$ . Thus, by (2.1),

$$\begin{aligned} \arcsin \frac{\langle \mathbf{x}^u, \mathbf{x}^v \rangle}{\|\mathbf{x}^u\| \|\mathbf{x}^v\|} &\leq \frac{\langle \mathbf{x}^u, \mathbf{x}^v \rangle}{\|\mathbf{x}^u\| \|\mathbf{x}^v\|} + 10 \left( \frac{1}{\|\mathbf{x}^u\| \|\mathbf{x}^v\|} \gamma^2 \left( \frac{1}{d} + \frac{2a}{\sqrt{d}} + a^2 \right) \delta_+(u, v) \right)^3 \\ &\leq \frac{\langle \mathbf{x}^u, \mathbf{x}^v \rangle}{\|\mathbf{x}^u\| \|\mathbf{x}^v\|} + \gamma^2 \frac{\delta_+(u, v)^3}{10d^3}, \end{aligned}$$

where in the second inequality we used  $\|\mathbf{x}^u\|, \|\mathbf{x}^v\| \geq 1$ ,  $\frac{1}{d} + \frac{2a}{\sqrt{d}} + a^2 \leq \frac{4}{d}$  and  $\gamma \leq 1/10$ . Summing over all  $uv \in E(G)$  and using that  $q = \|\mathbf{x}^v\|^2$  for every  $v \in V(G)$ , we get

$$\sum_{uv \in E(G)} \arcsin \frac{\langle \mathbf{x}^u, \mathbf{x}^v \rangle}{\|\mathbf{x}^u\| \|\mathbf{x}^v\|} \leq \frac{1}{q} \sum_{uv \in E(G)} \langle \mathbf{x}^u, \mathbf{x}^v \rangle + \gamma^2 \sum_{uv \in E(G)} \frac{\delta_+(u, v)^3}{10d^3}.$$

By (2.1),

$$\begin{aligned} \sum_{uv \in E(G)} \langle \mathbf{x}^u, \mathbf{x}^v \rangle &= -\gamma n\sqrt{d} + \gamma^2 \left( \frac{1}{d} + \frac{2a}{\sqrt{d}} + a^2 \right) \sum_{uv \in E(G)} \delta(u, v) \\ &= -\gamma n\sqrt{d} + \gamma^2 \left( \frac{1}{d} + \frac{2a}{\sqrt{d}} + a^2 \right) 3s = -\gamma n\sqrt{d} + \Theta \left( \frac{\gamma^2 s}{d} \right) \end{aligned}$$

since  $\sum_{uv \in E(G)} \delta(u, v) = \sum_{uv \in E(G)} d(u, v) - \frac{nd}{2} = 3t(G) - d^3/2 = 3s$ . Using  $1 \leq q \leq 2$ , we get

$$\sum_{uv \in E(G)} \arcsin \frac{\langle \mathbf{x}^u, \mathbf{x}^v \rangle}{\|\mathbf{x}^u\| \|\mathbf{x}^v\|} \leq -\frac{\gamma n\sqrt{d}}{2} + \Theta \left( \frac{\gamma^2 s}{d} \right) + \gamma^2 \sum_{uv \in E(G)} \frac{\delta_+(u, v)^3}{10d^3}.$$

Substituting this into (4.1) and using that if  $d(u, v) > d^2/n$ , then  $\delta_+(u, v) \leq d(u, v) \leq d(u) = d$ ,

$$\begin{aligned} \text{sp}(G) &\geq -\frac{1}{\pi} \left( -\frac{\gamma n\sqrt{d}}{4} + \Theta \left( \frac{\gamma^2 s}{d} \right) + \gamma^2 \sum_{uv \in E(G)} \frac{\delta_+(u, v)^3}{10d^3} - \gamma^2 \sum_{\substack{uv \in E(G): \\ d(u,v) > 20d^2/n}} \frac{d(u, v)}{10d} \right) \\ &\geq -\frac{1}{\pi} \left( -\frac{\gamma n\sqrt{d}}{4} + \Theta \left( \frac{\gamma^2 s}{d} \right) + \gamma^2 \sum_{\substack{uv \in E(G): \\ d(u,v) \leq 20d^2/n}} \frac{\delta_+(u, v)^3}{10d^3} \right), \end{aligned}$$

which completes the proof of the lemma.  $\square$