How unproportional must a graph be?

Humberto Naves* Oleg Pikhurko† Alex Scott ‡

Abstract

Let $u_k(G,p)$ be the maximum over all k-vertex graphs F of by how much the number of induced copies of F in G differs from its expectation in the Erdős-Rényi random graph with the same number of vertices as G and with edge probability p. This may be viewed as a measure of how close G is to being p-quasirandom. For a positive integer n and $0 < p < 1$, let $D(n,p)$ be the distance from $p\binom{n}{2}$ to the nearest integer. Our main result is that, for fixed $k \geq 4$ and for n large, the minimum of $u_k(G,p)$ over n-vertex graphs has order of magnitude $\Theta\big(\max\{D(n,p), p(1-p)\}n^{k-2}\big)$ provided that $p(1-p)n^{1/2} \to \infty$.

1 Introduction

An important result of Erdős and Spencer [8] states that every graph G of order n contains a set $S \subseteq V(G)$ such that $e(G[S])$, the number of edges in the subgraph induced by S, differs from $\frac{1}{2}\binom{|S|}{2}$ by at least $\Omega(n^{3/2})$; an earlier observation of Erdős [6] shows that this lower bound is tight up to the constant. More generally, it was shown in [7] that for graphs with density $p \in (\frac{2}{n-1}, 1-\frac{2}{n-1})$, there is some subset where the number of edges differs from expectation by at least $c\sqrt{p(1-p)}n^{3/2}$ (see [3, 4] for further results and discussion).

When p is constant, the above results can be equivalently reformulated in the language of graph limits as that the smallest cut-distance from the constant-p graphon to an order-n graph G is $\Theta(n^{-1/2})$. Instead of defining all terms here (which can be found in Lovász' book [18]), we observe that the cut-distance in this special case is equal, within some multiplicative constant, to the maximum over $S \subseteq V(G)$ of $\frac{1}{n^2}\left|2e(G[S]) - p|S|^2\right|$.

There are other measures of how close a graph G is to the constant-p graphon, which means measuring how close G is to being p-quasirandom. Here we consider two possibilities, subgraph statistics and graph norms, as follows.

For graphs G and H, we denote by $N(H,G)$ the number of induced subgraphs of G that are isomorphic to H. For example, if $v(H) = k \leq n$, then the expected number of H-subgraphs in the Erdős-Rényi random graph $\mathbb{G}_{n,p}$ is

$$\mathbf{E}[N(H,\mathbb{G}_{n,p})] = \frac{n(n-1)\dots(n-k+1)}{|\mathrm{Aut}(H)|}\, p^{e(H)}(1-p)^{\binom{k}{2}-e(H)},$$

where $\mathrm{Aut}(H)$ is the group of automorphisms of H.

*Institute for Mathematics and its Applications, University of Minnesota, Minneapolis, MN 55455, USA. Email: hnaves@ima.umn.edu. This research was supported in part by the Institute for Mathematics and its Applications with funds provided by the National Science Foundation.

†Mathematics Institute and DIMAP, University of Warwick, Coventry CV4 7AL, UK. Supported by ERC grant 306493 and EPSRC grant EP/K012045/1. Email: O.Pikhurko@warwick.ac.uk

‡Mathematical Institute, University of Oxford, Andrew Wiles Building, Radcliffe Observatory Quarter, Woodstock Road, Oxford OX2 6GG, UK; Email: scott@maths.ox.ac.uk.

Let $k \geq 2$ be a fixed integer parameter. For any graph G on n vertices, let

$$u_k(G,p) := \max\Big\{ N(F,G) - \mathbf{E}[N(F,\mathbb{G}_{n,p})] \,:\, v(F) = k \Big\}, \tag{1.1}$$

where the maximum is taken over all (non-isomorphic) graphs F on k vertices and $0 < p < 1$. The quantity $u_k(G,p)$ measures how far the graph G is away from the random graph $\mathbb{G}_{n,p}$ in terms of k-vertex induced subgraph counts. This is within a constant factor (that depends on k only) from the variational distance between $\mathbb{G}_{k,p}$ and a random k-vertex subgraph of G.

We are interested in estimating

$$u_k(n,p) := \min\{u_k(G,p) : v(G) = n\}, \tag{1.2}$$

the minimum value of $u_k(G,p)$ that a graph G of order n can have. Informally speaking, we ask how p-quasirandom a graph of order n can be.

Clearly, $u_2(n,p) < 1$ and $u_2(n,p) = 0$ if $p\binom{n}{2}$ is integer. Indeed, if we denote by $D(n,p)$ the distance from $p\binom{n}{2}$ to the nearest integer, then $u_2(n,p) = D(n,p)$. The problem of constructing pairs (F,p) with $u_3(F,p) = 0$ (such graphs F were called *p-proportional*) received some attention because the Central Limit Theorem fails for the random variable $N(F,\mathbb{G}_{n,p})$ for such F, see [2, 11, 15]. Apart from sporadic examples, infinitely many such pairs were constructed by Janson and Kratochvil [14] for $p = 1/2$ and by Janson and Spencer [16] for every fixed rational p; see Kärrman [17] for a different proof of the last result.

The main contribution of this paper is the following.

Theorem 1.1. *(a) Let $k \geq 3$ be fixed and $p = p(n) \in (0,1)$ with $\frac{1}{p(1-p)} = o(n^{1/2})$. Then*

$$u_k(n,p) = O\big(\max\{D(n,p),\, p(1-p)\}n^{k-2}\big).$$

(b) Let $k \geq 4$ be fixed and $p = p(n) \in (0,1)$. Then

$$u_k(n,p) = \Omega\big(\max\{D(n,p),\, p(1-p)\}n^{k-2}\big).$$

Another measure of graph similarity is the $2k$-th Shatten norm. Lemma 8.12 in [18] shows that the 4-th Shatten norm defines the same topology as the cut-norm. Again, we define it only for the special case when we want to measure how p-quasirandom a graph (possibly with loops) G of order n is. Here, we take the ℓ_{2k}-norm of the eigenvalues of $M = A - pJ$, where A is the adjacency matrix of G and J is the all-1 matrix. We remark that when G has a loop, the corresponding diagonal entry in the matrix A is one. An equivalent and more combinatorial definition of the $2k$-th Shatten norm is to take $\|G - p\|_{C_{2k}} := t(C_{2k}, M)^{1/2k}$, where C_{2k} is the $2k$-cycle and $t(F,M)$ denotes the *homomorphism density* of a graph F, which is the expected value of $\prod_{ij \in E(F)} M_{f(i),f(j)}$, where $f : V(F) \to [n]$ is a uniformly chosen random function, see [18, Chapter 5]. In other words,

$$\|G - p\|_{C_{2k}} = n^{-1} \left(\sum_{f:\mathbb{Z}/2k\mathbb{Z} \to V(G)} \; \prod_{i \in \mathbb{Z}/2k\mathbb{Z}} (A_{f(i),f(i+1)} - p) \right)^{1/2k}, \tag{1.3}$$

where the sum is over all (not necessarily injective) maps $f : \mathbb{Z}/2k\mathbb{Z} \to V(G)$ from the residues modulo $2k$ to the vertex set of G.

Proposition 1.2. *Let $k \geq 2$ be a fixed integer. The minimum of $\|G - p\|_{C_{2k}}$ over all n-vertex graphs G (possibly with loops) is*

$$\Theta\left(\min\left\{p(1-p),\, p^{1/2}(1-p)^{1/2} n^{-(k-1)/2k}\right\}\right).$$

Hatami [10] studied which graphs other than even cycles produce a norm when we use the appropriate analogue of (1.3). He showed, among other things, that complete bipartite graphs with both parts of even size do. We also prove a version of Proposition 1.2 for this norm.

The rest of this paper is organized as follows. In Section 2 we prove the lower bound from Theorem 1.1. In Section 3 we prove the upper bound from the same theorem. We consider graphs norms in Section 4, in particular proving Proposition 1.2 there. The final section contains some open questions and concluding remarks. Throughout the paper, we adopt the convention that k is a fixed constant and all asymptotic notation symbols (Ω, O, o and Θ) are with respect to the variable n. To simplify the presentation, we often omit floor and ceiling signs whenever these are not crucial and make no attempts to optimize the absolute constants involved.

2 Lower bound for $u_k(n,p)$ in the range $k \geq 4$

The goal of this section is to prove that $u_k(n,p) = \Omega\big(\max\{D(n,p),\, p(1-p)\}n^{k-2}\big)$. More precisely, we will show that there exists a constant $\varepsilon = \varepsilon(k) > 0$ such that $u_k(G,p) \geq \varepsilon \max\{D(n,p),\, p(1-p)\}n^{k-2}$, for all graphs G on $n \geq k$ vertices and for all $0 < p < 1$. The following lemma shows that it is enough to prove the lower bound for $k = 4$ only.

Lemma 2.1. *For every $k \geq 2$ there is $c_k > 0$ such that $u_{k+1}(G,p) \geq c_k n \cdot u_k(G,p)$ for every graph G of order $n \geq k+1$ and for all $0 < p < 1$.*

Proof. Let $u_F(G,p) := \big|N(F,G) - \mathbf{E}[N(F,\mathbb{G}_{n,p})]\big|$. Take a graph F of order k with $u_F(G,p) = u_k(G,p)$. Let $f(G)$ be the number of pairs (A,x) where A induces F in G and $x \in V(G) \setminus A$. Then $f(G) = (n-k)N(F,G)$ and $\mathbf{E}[f(\mathbb{G}_{n,p})] = (n-k)\mathbf{E}[N(F,\mathbb{G}_{n,p})]$; thus these two parameters differ (in absolute value) by exactly $(n-k)u_k(G,p)$. On the other hand, $f(G)$ can be written as $\sum_{F'} N(F,F')N(F',G)$ where the sum is over $(k+1)$-vertex graphs F'. The expectation of $f(\mathbb{G}_{n,p})$ obeys the same linear identity:

$$\mathbf{E}[f(\mathbb{G}_{n,p})] = \sum_{F'} N(F,F')\,\mathbf{E}[N(F',\mathbb{G}_{n,p})].$$

Thus there is a $(k+1)$-vertex graph F' satisfying $u_{F'}(G,p) \geq c_k n \cdot u_k(G,p)$, where c_k is a constant depending solely on k. □

In the next lemma we prove one of the bounds for $u_4(n,p)$. We remark that it was implicitly proven in [14, Proposition 3.7].

Lemma 2.2. *There exists an absolute constant $\varepsilon > 0$ such that, for every $0 < p < 1$ and for all graphs G on $n \geq 4$ vertices, the inequality $u_4(G,p) \geq \varepsilon p(1-p)n^2$ holds.*

Proof. For every graph G, we can write $e(G)^2$ as

$$e(G)^2 = \sum_F \alpha_F N(F,G), \tag{2.1}$$

where in the summation F ranges over graphs satisfying $2 \leq v(F) \leq 4$, and α_F is a constant depending on F only. For example, if F is an edge then $\alpha_F = 1$ and if $v(F) = 4$, then α_F is the number of ordered pairs of disjoint edges in F.

Let $\varepsilon > 0$ be a sufficiently small constant. Suppose that there is a graph G of order $n \geq 4$ satisfying $u_4(G,p) < \varepsilon p(1-p)n^2$. By applying Lemma 2.1 twice, we conclude that $u_2(G,p) < \frac{\varepsilon p(1-p)}{c_2 c_3}$, where c_i's are given by the lemma. This implies that

$$\left| e(G)^2 - \mathbf{E}\left[e(\mathbb{G}_{n,p})\right]^2 \right| < \left| e(G) - \mathbf{E}\left[e(\mathbb{G}_{n,p})\right] \right| \cdot \left(2p\binom{n}{2} + \frac{\varepsilon p(1-p)}{c_2 c_3} \right) < \frac{3\varepsilon p^2(1-p)}{c_2 c_3}\binom{n}{2}.$$

Since $\mathbf{E}\left[e(\mathbb{G}_{n,p})^2\right] - \mathbf{E}\left[e(\mathbb{G}_{n,p})\right]^2 = \mathbf{Var}[e(\mathbb{G}_{n,p})] = p(1-p)\binom{n}{2}$, we have that

$$\left| e(G)^2 - \mathbf{E}\left[e(\mathbb{G}_{n,p})^2\right] \right| > \frac{p(1-p)}{2}\binom{n}{2}. \tag{2.2}$$

Moreover, the identity (2.1) implies that $\mathbf{E}\left[e(\mathbb{G}_{n,p})^2\right] = \sum_F \alpha_F \mathbf{E}[N(F, \mathbb{G}_{n,p})]$. Thus

$$\sum_{k=2}^{4} \sum_{v(F)=k} \alpha_F\, u_k(G,p) \geq \sum_F \alpha_F \left| N(F,G) - \mathbf{E}[N(F,\mathbb{G}_{n,p})] \right| > \frac{p(1-p)}{2}\binom{n}{2}.$$

The previous inequality together with Lemma 2.1 imply that $u_4(G,p) > \varepsilon p(1-p)n^2$, contradicting our assumption and proving the lemma. □

The previous lemma states that $u_k(n,p) = \Omega(p(1-p)n^{k-2})$, thus to finish the proof of the lower bound, we need to show that $u_k(n,p) = \Omega(D(n,p)n^{k-2})$. The latter bound is a consequence of $u_2(n,p) = D(n,p)$ together with Lemma 2.1, thereby concluding the proof.

3 Upper bound for $k \geq 3$

In this section, we prove that $u_k(n,p) = O(\max\{D(n,p), p(1-p)\}n^{k-2})$ for fixed $k \geq 3$ and for all p such that $\frac{1}{p(1-p)} = o(n^{1/2})$. Observe that we might assume, without loss of generality, that $p \leq \frac{1}{2}$. Indeed, if $\overline{G}$ denotes the complement of G then $u_k(G,p) = u_k(\overline{G}, 1-p)$, which implies that $u_k(n,p) = u_k(n, 1-p)$. Thus our assumption can be made because the bound $O(\max\{D(n,p), p(1-p)\}n^{k-2})$ is symmetric with respect to p and $1-p$ (recall that $D(n,p) = u_2(n,p) = u_2(n,1-p) = D(n,1-p)$). In addition, note that in the range $p \leq \frac{1}{2}$, it suffices to show that $u_k(n,p) = O(\max\{D(n,p), p\}n^{k-2})$.

To prove the upper bound, we borrow some definitions, results, and proof ideas from [16]. Following their notation, one can count the number of induced subgraphs of G that are isomorphic to H using the following identity

$$N(H,G) = \sum_{H'} \prod_{e \in E(H')} I_G(e) \prod_{e \in E(\overline{H'})} (1 - I_G(e)) \tag{3.1}$$

where we sum over all H' isomorphic to H with $V(H') \subseteq V(G)$, $I_G(e)$ is the indicator that e is an edge in G and $\overline{H'}$ denotes the complement of the graph H'. Observe that H' in the sum is not restricted to subgraphs of G. We define a related sum over the same range of H':

$$S(H,G) = S^{(p)}(H,G) := \sum_{H'} \prod_{e \in E(H')} (I_G(e) - p) \tag{3.2}$$

where p is as before. Replacing $I_G(e)$ by $(I_G(e) - p) + p$ and $1 - I_G(e)$ by $(1-p) - (I_G(e) - p)$ in (3.1), we obtain

$$N(H,G) = \mathbf{E}[N(H,\mathbb{G}_{n,p})] + \sum_{F \in \mathcal{F}_k} a_{F,H}(n,p) S(F,G), \tag{3.3}$$

where $k = v(H)$, $\mathcal{F}_k$ denotes the family of all graphs F without isolated vertices satisfying $2 \leq v(F) \leq k$, and $a_{F,H}(n,p) = O(n^{v(H)-v(F)})$ is a coefficient that does not depend on G. In fact, one can show that $a_{F,H}(n,p) = O(p^{e(H)-\alpha}n^{v(H)-v(F)})$, where α is the maximum number of edges a common subgraph of both H and F can have, but we will not need such an estimate. To prove that there exists a graph G on n vertices such that $u_k(G,p) = O(\max\{D(n,p),p\}n^{k-2})$, it suffices to show that there exists G such that

$$S(F,G) = \begin{cases} O(pn^{v(F)-2}) & \text{for all } F \in \mathcal{F}_k \setminus \{K_2\}, \\ O(D(n,p)) & \text{if } F = K_2. \end{cases} \tag{3.4}$$

A natural candidate for G in (3.4) is the random graph $\mathbb{G}_{n,p}$. The next lemma yields some bounds for $S(F, \mathbb{G}_{n,p})$.

Lemma 3.1. *Let $G \sim \mathbb{G}_{n,p}$. For all $F \in \mathcal{F}_k$, we have*

$$\mathbf{E}[S(F,G)] = 0 \quad \textit{and} \quad \mathbf{E}[S(F,G)^2] \leq p^{e(F)}n^{v(F)}.$$

Proof. By (3.2), we have

$$\mathbf{E}[S(F,G)] = \sum_{F'} \mathbf{E}\left[\prod_{e \in E(F')} (I_G(e) - p)\right],$$

where the sum is over all F' isomorphic to F with $V(F') \subseteq V(G)$. Each expectation on the right-hand side vanishes, by independence and since $\mathbf{E}[I_G(e)] = p$. Thus $\mathbf{E}[S(F,G)] = 0$. We similarly write

$$\mathbf{E}[S(F,G)^2] = \sum_{F',F''} \mathbf{E}\left[\prod_{e \in E(F')} (I_G(e) - p) \prod_{e \in E(F'')} (I_G(e) - p)\right].$$

where the sum is over all pair (F', F'') of graphs isomorphic to F with $V(F') \cup V(F'') \subseteq V(G)$. The expectation term in the above sum vanishes when $F' \neq F''$ and it is bounded by $p^{e(F)}$ when $F' = F''$. Since the number of possible choices for F' is at most $\binom{n}{t} \cdot t! \leq n^t$, where $t = v(F)$, we conclude that $\mathbf{E}[S(F,G)^2] \leq p^{e(F)}n^{v(F)}$. □

Using Chebyschev's inequality (see, e.g., [1, Theorem 4.1.1]), we have that, for all $\lambda > 0$,

$$\mathbf{Pr}\left[\left|S(F, \mathbb{G}_{n,p})\right| \geq \lambda \cdot p^{e(F)/2}n^{v(F)/2}\right] \leq \lambda^{-2}. \tag{3.5}$$

By the union bound combined with (3.5), the random graph $G \sim \mathbb{G}_{n,p}$ satisfies the following property with probability at least 0.96.

Property A. $|S(F,G)| \leq 5|\mathcal{F}_k|^{1/2}p^{e(F)/2}n^{v(F)/2}$ for all graphs $F \in \mathcal{F}_k$.

The inequality $p^{e(F)/2}n^{v(F)/2} \leq pn^{v(F)-2}$ holds whenever $v(F) \geq 4$. This is because every graph on 4 or more vertices in $\mathcal{F}_k$ has at least 2 edges, since no vertex is isolated. In order to find a graph satisfying the conditions expressed in (3.4), we just need to adjust G so that $S(K_2, G) = O(D(n,p))$ and $S(F,G) = O(pn^{v(F)-2})$ when $F \in \mathcal{F}_3 \setminus \{K_2\}$. The adjustment must be performed carefully, to prevent $S(F,G)$ from changing too much for graphs $F \in \mathcal{F}_k$ with $v(F) \geq 4$.

Let us investigate what happens to $S(F,G)$ when we add or remove an edge. Note that by "edges", we generally mean edges in the complete graph, i.e., all pairs ij with $i,j \in V(G)$, and not only the pairs that happen to be selected as the edges of G. For each pair ij, with $i,j \in V(G)$, let

$$S_{ij}(F,G) := S(F,G\cup\{ij\}) - S(F,G\setminus\{ij\}), \tag{3.6}$$

where $G\cup\{ij\}$ and $G\setminus\{ij\}$ represent the graphs obtained from G by adding and removing the edge ij, respectivelly. The next lemma gives a bound for the expectation and the variance of $S_{ij}(F,\mathbb{G}_{n,p})$.

Lemma 3.2. *Let $G\sim\mathbb{G}_{n,p}$. For all $F\in\mathcal{F}_k$ with $v(F)\geq 3$ and all pairs $1\leq i<j\leq n$, we have*

$$\mathbf{E}[S_{ij}(F,G)] = 0 \quad \textit{and} \quad \mathbf{E}[S_{ij}(F,G)^2] \leq k^2 p^{e(F)-1} n^{v(F)-2}.$$

Proof. By expanding the definition in (3.6), we have

$$S_{ij}(F,G) = \sum_{F'} \prod_{e\in E(F')\setminus\{ij\}} (I_G(e)-p),$$

where the sum is over all F' isomorphic to F with $V(F')\subseteq V(G)$ and $\{i,j\}\in E(F')$. By independence and the linearity of expectation, we have $\mathbf{E}[S_{ij}(F,G)]=0$. For the last part of the lemma, we write

$$\mathbf{E}[S_{ij}(F,G)^2] = \sum_{F',F''} \mathbf{E}\left[\prod_{e\in E(F')\setminus\{ij\}} (I_G(e)-p) \prod_{e\in E(F'')\setminus\{ij\}} (I_G(e)-p)\right].$$

where the sum is over all pair (F',F'') of graphs isomorphic to F with $V(F')\cup V(F'')\subseteq V(G)$ and $\{i,j\}\in E(F')\cap E(F'')$. The expectation term in the above sum vanishes when $F'\neq F''$ and it is bounded by $p^{e(F)-1}$ when $F'=F''$. Since the number of possible choices for F' is at most $k^2 n^{v(F)-2}$, we conclude that $\mathbf{E}[S_{ij}(F,G)^2]\leq k^2p^{e(F)-1}n^{v(F)-2}$. □

The family $\mathcal{F}_3$ consists of three graph: cliques K_2 and K_3 as well as the 2-path P_2, the unique graph on three vertices having exactly two edges. So, we just need to adjust $S(K_2,G)$, $S(K_3,G)$ and $S(P_2,G)$. Let $\varepsilon=\varepsilon(k)>0$ be sufficiently small.

Take a pair ij of vertices. Let Z_s^*, for $0\leq s\leq 2$, denote the number of vertices $w\in V(G)\setminus\{i,j\}$ such that exactly s of the pairs iw and jw belong in G. The triple (Z_0^*,Z_1^*,Z_2^*) has a multinomial distribution for $G\sim\mathbb{G}_{n,p}$. We shift to zero mean by setting

$$\begin{aligned} Z_0 &:= Z_0^* - (n-2)\cdot(1-p)^2, \\ Z_2 &:= Z_2^* - (n-2)\cdot p^2. \end{aligned}$$

Note that Z_0 and Z_2 determine $S_{ij}(F,G)$ for every 3-vertex graph F, which is the amount by how much $S(F,G)$ changes when we "flip" ij. Namely, using $Z_1^*=n-2-Z_0^*-Z_2^*$, we obtain

$$\begin{aligned} S_{ij}(P_2,G) &= 2(1-p)Z_2^* + (1-2p)Z_1^* - 2pZ_0^* = Z_2 - Z_0 =: Y_1 \\ S_{ij}(K_3,G) &= (1-p)^2Z_2^* - p(1-p)Z_1^* + p^2Z_0^* = (1-p)Z_2 + pZ_0 =: Y_2. \end{aligned} \tag{3.7}$$

In the next statement we show that for any fixed open rectangle $R\subseteq\mathbb{R}^2$, there exists $\varepsilon=\varepsilon(R)$ such that $\left(\frac{Y_1}{\sqrt{2p(1-p)n}},\frac{Y_2}{p(1-p)\sqrt{n}}\right)\in R$ with probability at least ε, provided that p is not too small.

Lemma 3.3. *For fixed reals $\alpha_1 < \alpha_2$ and $\beta_1 < \beta_2$ there exists $\varepsilon = \varepsilon(\alpha_1, \alpha_2, \beta_1, \beta_2) > 0$, such that with probability at least ε, we have*

$$\alpha_1 \leq \frac{Y_1}{\sqrt{p(1-p)n}} \leq \alpha_2 \quad \text{and} \quad \beta_1 \leq \frac{Y_2}{p(1-p)\sqrt{n}} \leq \beta_2, \tag{3.8}$$

provided that $\frac{1}{p(1-p)} = o(n^{1/2})$.

Proof. Without loss of generality, we again assume that $p \leq \frac{1}{2}$. Our first observation is that, in order to prove the lemma, it suffices to show that for any fixed reals $\alpha_1' < \alpha_2'$ and $\beta_1' < \beta_2'$, there exists $\varepsilon = \varepsilon(\alpha_1', \alpha_2', \beta_1', \beta_2') > 0$ such that with probability at least ε, we have

$$\alpha_1' \sqrt{p(1-p)n} \leq Z_0 \leq \alpha_2' \sqrt{p(1-p)n} \quad \text{and} \quad \beta_1' p\sqrt{n} \leq Z_2 \leq \beta_2' p\sqrt{n}. \tag{3.9}$$

This is because of the following reason. Let $T_p : \mathbb{R}^2 \to \mathbb{R}^2$ be a family of linear transformations indexed by $p \in \left[0, \frac{1}{2}\right]$, given by $T_p(\alpha, \beta) := \left(\beta\sqrt{\frac{p}{1-p}} - \alpha, \beta + \alpha\sqrt{\frac{p}{1-p}}\right)$. If $\alpha = \frac{Z_0}{\sqrt{p(1-p)n}}$ and $\beta = \frac{Z_2}{p\sqrt{n}}$ then

$$\left(\frac{Y_1}{\sqrt{p(1-p)n}}, \frac{Y_2}{p(1-p)\sqrt{n}}\right) = T_p(\alpha, \beta).$$

Since T_p is an equicontinuous family of surjective linear transformations, there exist $\alpha_1' < \alpha_2'$ and $\beta_1' < \beta_2'$ such that if $(\alpha, \beta) \in [\alpha_1', \alpha_2'] \times [\beta_1', \beta_2']$ then $T_p(\alpha, \beta) \in [\alpha_1, \alpha_2] \times [\beta_1, \beta_2]$. Hence for (3.8) to hold with probability bounded away from zero, it suffices that (3.9) holds with probability at least ε.

We turn to show that (3.9) is satisfied with probability bounded away from zero. We have $Z_2 = Z_2^* - (n-2)p^2$ and $Z_0 = Z_0^* - (n-2)(1-p)^2$, so the first condition of (3.9) is equivalent to

$$(n-2)(1-p)^2 + \alpha_1' \sqrt{p(1-p)n} \leq Z_0^* \leq (n-2)(1-p)^2 + \alpha_2' \sqrt{p(1-p)n}, \tag{3.10}$$

while the second is equivalent to

$$(n-2)p^2 + \beta_1' p\sqrt{n} \leq Z_2^* \leq (n-2)p^2 + \beta_2' p\sqrt{n}. \tag{3.11}$$

We begin by sampling Z_2^*. We know that Z_2^* is distributed according to the binomial distribution $Z_2^* \sim \mathrm{Bin}(n-2, p^2)$, and thus we have $\mathbf{Var}[Z_2^*] = \Theta(p^2 n)$. By de Moivre-Laplace theorem (see, e.g., Theorem VII.3.2 in [9]), we have that (3.11) is satisfied with probability at least $\varepsilon' = \varepsilon'(\beta_1', \beta_2')$. We now sample Z_0^* conditioned on the event E_β, which states that $Z_2^* = (n-2)p^2 + \beta p\sqrt{n}$ for some $\beta_1' \leq \beta \leq \beta_2'$. Conditioned on E_β, we have $Z_0^* \sim \mathrm{Bin}\left((n-2)(1-p^2) - \beta p\sqrt{n}, \frac{(1-p)^2}{1-p^2}\right)$, thus

$$\mathbf{E}[Z_0^* | E_\beta] = (n-2)(1-p)^2 - \beta p \frac{(1-p)^2}{1-p^2}\sqrt{n} \quad \text{and} \quad \mathbf{Var}[Z_0^* | E_\beta] = \Theta(p(1-p)n).$$

Since the term $\beta p \frac{(1-p)^2}{1-p^2}\sqrt{n}$ is negligible when compared to $\sqrt{p(1-p)n}$, by de Moivre-Laplace theorem, we have that (3.10) is satisfied with probability at least $\varepsilon'' = \varepsilon''(\alpha_1', \alpha_2')$. Therefore (3.9) is satisfied with probability at least $\varepsilon = \varepsilon'\varepsilon''$, which concludes the proof. □

Let E_1 consist of those pairs e for which

$$e \in E(G), \quad \sqrt{pn} < Y_1(e) \quad \text{and} \quad p\sqrt{n} < Y_2(e). \tag{3.12}$$

Let $I_1(e)$ be the indicator random variable for E_1. For the random graph $G \sim \mathbb{G}_{n,p}$, the first condition $e \in E(G)$ for e to be in E_1 is independent of the other two conditions. Thus we can assume that $\mathbf{E}[I_1(e)] \geq \varepsilon p$. We have $|E_1| = \sum_e I_1(e)$, hence $\mathbf{E}[\,|E_1|\,] \geq \varepsilon p\binom{n}{2}$. We re-write the variance of $|E_1|$ as the sum of pairwise covariances of its components:

$$\mathbf{Var}[\,|E_1|\,] = \sum_{e \cap e' = \emptyset} \mathbf{Cov}[I_1(e), I_1(e')] + \sum_{e \cap e' \neq \emptyset} \mathbf{Cov}[I_1(e), I_1(e')]. \tag{3.13}$$

When e and e' have no common vertices, then $I_1(e)$ can only influence $I_1(e')$ through the four common edges with both endpoints in $e \cup e'$. The probability that Y_1 or Y_2 is within 8 from the thresholds in (3.12) is $o(1)$ by de Moivre-Laplace theorem. It follows that the covariance $\mathbf{Cov}[I_1(e), I_1(e')] = o(p^2)$. Thus the first sum in (3.13) has $O(n^4)$ terms, each $o(p^2)$. Since the second sum has $O(n^3)$ terms, each at most p^2, the variance of $|E_1|$ is $o(n^4p^2)$. By Chebyschev's inequality,

$$\mathbf{Pr}[\,|E_1| < \varepsilon pn^2/4\,] = o(1).$$

Next, we put e in one of sets as follows:

$$\begin{aligned}
E_2 &:= \{e : e \in E(G),\ \sqrt{pn} < Y_1(e) \text{ and } Y_2(e) < -p\sqrt{n}\}\\
E_3 &:= \{e : e \in E(G),\ Y_1(e) < -\sqrt{pn} \text{ and } p\sqrt{n} < Y_2(e)\}\\
E_4 &:= \{e : e \in E(G),\ Y_1(e) < -\sqrt{pn} \text{ and } Y_2(e) < -p\sqrt{n}\}\\
E_5 &:= \{e : e \notin E(G),\ |Y_1(e)| < 0.1\sqrt{pn} \text{ and } |Y_2(e)| < 0.1p\sqrt{n}\}.
\end{aligned}$$

The argument above implies that asymptotically almost surely $|E_i| \geq \varepsilon pn^2/4$ for all $i = 1, \ldots, 4$. Similarly, one can show that $|E_5| \geq \varepsilon n^2/4$ asymptotically almost surely (E_5 might be much "denser" than the other sets because we dropped the requirement $e \in E(G)$). Finally, using the standard Chernoff estimates one can show that asymptotically almost surely $\Delta(G) \leq 2np$ for $G \sim \mathbb{G}_{n,p}$. In particular, the following property is satisfied with probability at least 0.99 when n is large.

Property B. $|E_i| \geq \varepsilon pn^2/4$ for $i = 1, \ldots, 4$. Moreover $|E_5| \geq \varepsilon n^2/4$, and $\Delta(G) \leq 2\varepsilon n$.

Let E^* denote the set of pairs ij, where $i, j \in V(G)$, such that

$$|S_{ij}(F, G)| > 4k \cdot \varepsilon^{-1/2} |\mathcal{F}_k|^{1/2} p^{e(F)/2-1} n^{v(F)/2-1} \tag{3.14}$$

for at least one $F \in \mathcal{F}_k$. Chebyschev's inequality together with Lemma 3.2 implies that $\mathbf{Pr}[ij \in E^*] \leq \varepsilon p/16$. Hence $\mathbf{E}[\,|E^*|\,] \leq \varepsilon pn^2/32$. By Markov's inequality, $\mathbf{Pr}[\,|E^*| > \varepsilon pn^2/8\,] < \frac{1}{4}$. Thus $G \sim \mathbb{G}_{n,p}$ satisfies the following property with probability at least 0.75.

Property C. E^* has size at most $\varepsilon pn^2/8$.

Finally, we state and prove the following result that asserts the existence of large matchings in relatively dense graphs.

Proposition 3.4. *Let H be a graph and let $\Delta := \Delta(H)$. There exists a matching in H of size at least $\frac{e(H)}{2\Delta}$. In particular, if $m < \Delta$ then H contains a subgraph H' with maximal degree $\Delta(H') \leq m$ and $e(H') \geq \frac{m}{4\Delta} e(H)$.*

Proof. Let M be a maximal matching in H, and assume M has $k < \frac{e(H)}{2\Delta}$ pairs. All the edges of H have at least one endpoint in $V(M)$, hence

$$e(H) \leq |V(M)| \cdot \Delta = 2k \cdot \Delta < e(H),$$

a contradiction. We remark that the bound $\frac{e(H)}{2\Delta}$ is not tight but it suffices for our purposes.

To construct H', we start with the empty graph. At each step of the construction, we apply the first assertion of the proposition to the graph $H \setminus H'$, in order to obtain a matching M having exactly $\left\lceil \frac{e(H)}{4\Delta} \right\rceil$ edges. We then add all the edges from M to H'. We repeat this step exactly m times. Since we always have $e(H') \leq m \cdot \left\lceil \frac{e(H)}{4\Delta} \right\rceil < \frac{e(H)}{2}$, and thus $e(H \setminus H') > \frac{e(H)}{2}$, it is always possible to find such M, in all the steps of the process. □

An important corollary of Proposition 3.4 is as follows.

Corollary 3.5. *Let $C > 0$ be fixed. If Properties **B** and **C** simultaneously hold for a graph G and n is sufficiently large, there exists a graph H' having at least Cn edges from each $E_i \setminus E^*$, $i = 1, \ldots, 5$, such that $\Delta(H') \leq 320C/\varepsilon$.*

Proof. Because of Property **C**, we have $|E^*| \leq \varepsilon pn^2/8$, which, together with Property **B**, implies that $|E_i \setminus E^*| \geq \varepsilon pn^2/8$ for $i = 1, \ldots, 4$, and $|E_5| \geq \varepsilon n^2/8$. Let H_i be the graph on $V(G)$ having edge set $E_i \setminus E^*$. We have $\Delta(H_i) \leq \Delta(G) \leq 2np$ for $i = 1, \ldots, 4$ and $\Delta(H_5) \leq n$. Hence $\frac{e(H_i)}{\Delta(H_i)} \geq \frac{\varepsilon n}{16}$ for all $i = 1, \ldots, 5$. By Proposition 3.4 applied with $m = 64C/\varepsilon < \min\{\Delta(H_i) : i = 1, \ldots, 5\}$, each H_i contains a subgraph H_i' having at least Cn edges such that $\Delta(H_i') \leq m$. Let $H' = \bigcup_{i=1}^{5} H_i'$. Clearly $\Delta(H') \leq 5m = 320C/\varepsilon$ and H' contains at least Cn each from each $E_i \setminus E^*$, thereby proving the corollary. □

Proof of the upper bound in Theorem 1.1. Given $p \in (0, 1)$ and $k \geq 4$, choose small $\varepsilon > 0$ and then sufficiently large C. Let $n \to \infty$. By the union bound, $G \sim \mathbb{G}_{n,p}$ satisfies Properties **A**, **B** and **C** with probability at least 0.7. Hence there exists a graph G on n vertices satisfying the three properties simultaneously. Fix such G.

From Corollary 3.5, there exits a graph H' having at least Cn edges from each $E_i \setminus E^*$, such that $\Delta := \Delta(H') \leq 320C/\varepsilon$. Let $E' = E(H')$.

In what follows, we change $E(G)$ on pairs, all of which will belong to E'. Note that at any intermediate step, the effect of (for instance) removing an edge $ij \in E' \cap E_1$ from $E(G)$ on $S(P_2, G)$ and $S(K_3, G)$ is not quite given by the initial values of $Y_1(ij)$ and $Y_2(ij)$, since certain edges iw, jw might have been changed. But E' was defined in such a way that there are most $2\Delta = o(\sqrt{pn})$ changed edges which affect either Y_1 or Y_2. So, the removal of $ij \in E_1 \setminus E^*$ from $E(G)$ at any intermediate stage, still decreases $S(P_2, G)$ by an amount between $\sqrt{pn} - 2\Delta$ and $4k\varepsilon^{-1/2}|\mathcal{F}_k|^{1/2}\sqrt{pn} + 2\Delta < \varepsilon^{-1}\sqrt{pn}$. Similarly, because $\Delta = o(p\sqrt{n})$, the same operation decreases $S(K_3, G)$ by an amount between $p\sqrt{n} - 2\Delta$ and $4k\varepsilon^{-1/2}|\mathcal{F}_k|^{1/2}p\sqrt{n} + 2\Delta < \varepsilon^{-1}p\sqrt{n}$.

By Property **A**, we know that $|S(K_2, G)| \leq 5|\mathcal{F}_k|^{1/2}p^{1/2}n$. If $S(K_2, G) \geq 1$, we can pick an $e \in E' \setminus E_5$ and remove it from G. This has the effect of reducing $S(K_2, G)$ by 1. If $S(K_2, G) \leq -1$, then we can pick an $e \in E' \cap E_5$ and add it to G. This new edge increases the value of $S(K_2, G)$ by 1. Iterate this process at most $5|\mathcal{F}_k|^{1/2}p^{1/2}n$ times to obtain a graph G such that $|S(K_2, G)| = D(n, p)$, always using a different edge e. This is possible because there are at least Cn edges from $E' \cap E_i$, for each i.

Our next goal is to make both $|S(K_3, G)|$ and $|S(P_2, G)|$ small without changing $S(K_2, G)$. Since all the operations were performed on the graph H', whose edges are disjoint from E^* and having maximum degree at most Δ, we know that $|S(K_3, G)| \leq pS_0$ and $|S(P_2, G)| \leq p^{1/2}S_0$, where

$$S_0 = 5|\mathcal{F}_k|^{1/2}p^{1/2}n^{3/2} + 5|\mathcal{F}_k|^{1/2}p^{1/2}n \cdot \varepsilon^{-1}\sqrt{n}.$$

We repeat the following step $(C - 5|\mathcal{F}_k|^{1/2})p^{1/2}n$ times. Consider the current graph G. There are four cases depending on whether each of $S(K_3, G)$ and $S(P_2, G)$ is positive or not. First suppose that they are both positive. Pick an edge $e \in E' \cap E_1$ and an $e' \in E' \cap E_5$

and replace e with e' in G. This operation preserves the value of $S(K_2, G)$, and has the effect of reducing both $S(K_3, G)$ and $S(P_2, G)$. It reduces $S(K_3, G)$ by between $0.8p\sqrt{n}$ and $2\varepsilon^{-1}p\sqrt{n} < pn$. Thus if (initially) $S(K_3, G) \geq pn$, then this value is lowered by at least $0.8p\sqrt{n}$. Regarding $S(P_2, G)$, the operation reduces it by between $0.8\sqrt{pn}$ and $2\varepsilon^{-1}\sqrt{pn} < pn$. Likewise, if $S(K_3, G) < 0$ and $S(P_2, G) > 0$, we replace an $e \in E' \cap E_2$ by an $e' \in E' \cap E_5$, and similarly in the other two cases. We iterate this process, always using edges e and e' that have not been used before. This is possible since E' contains at least Cn edges from each E_i. Also, once one of $|S(K_3, G)|$ or $|S(P_2, G)|$ becomes less than pn, it stays so for the rest of process. Since $(C - 5|\mathcal{F}_k|^{1/2})p^{1/2}n \cdot 0.8\sqrt{n} > S_0$, we have that $\max\{|S(K_3, G)|, |S(P_2, G)|\} < pn$ at the end.

The iterative process might change the value of $S(F, G)$ for $F \in \mathcal{F}_k$ with $v(F) \geq 4$. However, since all flips were performed on a bounded degree graph H' disjoint from E^*, we have that

$$\begin{aligned} |S(F,G)| &\leq 5|\mathcal{F}_k|^{1/2}p^{e(F)/2}n^{v(F)/2} + Cp^{1/2}n \cdot 4k\varepsilon^{-1/2}|\mathcal{F}_k|^{1/2}p^{e(F)/2-1}n^{v(F)/2-1} \\ &+ 4(C+\Delta)^2(v(F))^4pn^{v(F)-2} = O(pn^{v(F)-2}), \end{aligned}$$

provided that $e(F) \geq 3$, where the last summand is a generous upper bound on the number of copies of F that use more than one changed edge. For the case $e(F) \leq 2$, we have only one graph, namely $F = 2K_2$ (two parallel edges). But

$$S(K_2, G)^2 = p(1-p)\binom{n}{2} + (1-2p)S(K_2, G) + 2S(P_2, G) + 2S(2K_2, G),$$

therefore $|S(2K_2, G)| = O(pn^2)$, and hence $S(F, G) = O(pn^{v(F)-2})$ for all $F \in \mathcal{F}_k \setminus \{K_2\}$, finishing the proof of the upper bound. $\square$

4 Shatten norms and other related norms

We begin with the proof of Proposition 1.2

Proof of Proposition 1.2. Let $s = 2k$ and let G be a graph (possibly with loops) of order $n \to \infty$. Recall that the total number of edges (including loops) that G might have is at most $\binom{n+1}{2}$. Let $\overline{G}$ be the complement of G. In this definition of the complement, loopless vertices are mapped to loops and vice versa. Without loss of generality we may assume that $p \leq \frac{1}{2}$. This is because $\|G - p\|_{C_s}^s = \|\overline{G} - (1-p)\|_{C_s}^s$ and the expression in the statement we are trying to prove is symmetric with respect to p and $1 - p$.

Let $M = A - pJ$ be the shifted adjacency matrix of G, that is,

$$M_{ij} = \begin{cases} 1-p, & \text{if } ij \in E(G), \\ -p, & \text{otherwise,} \end{cases} \qquad 1 \leq i, j \leq n. \tag{4.1}$$

It is a symmetric real matrix so it has real eigenvalues $\lambda_1 \geq \cdots \geq \lambda_n$. For an even integer $s \geq 4$, we have

$$\sum_{i=1}^{n} \lambda_i^s = \mathrm{tr}(M^s) = n^s \|G - p\|_{C_s}^s.$$

Also, $\sum_{i=1}^n \lambda_i^2 = \sum_{i,j=1}^n M_{ij}^2 = (1-p)^2 e(G) + p^2 e(\overline{G})$.

From now on we split the analysis of the lower bound for $\|G-p\|_{C_s}^s$ into two cases. In the first case, we assume that $e(G) \geq p\binom{n+1}{2}/2$. This (together with $p \leq \frac{1}{2}$) implies that

$$\sum_{i=1}^{n} \lambda_i^2 \geq p(1-p)\binom{n+1}{2}/4.$$

By the inequality between the arithmetic and k-th power means for $k \geq 2$ applied to non-negative numbers $\lambda_1^2, \ldots, \lambda_n^2$ (or just by the convexity of $x \mapsto x^k$ for $x \geq 0$), we conclude that

$$\left(\frac{\lambda_1^{2k} + \cdots + \lambda_n^{2k}}{n}\right)^{1/k} \geq \frac{\lambda_1^2 + \cdots + \lambda_n^2}{n} \geq p(1-p)n/8.$$

Thus $n^{2k}\|p - G\|_{C_{2k}}^{2k} = \sum_{i=1}^n \lambda_i^{2k} = \Omega(p^k(1-p)^k n^{k+1})$, giving the required lower bound in the first case.

In the second case, we assume that $e(G) < p\binom{n+1}{2}/2$. Since λ_n is the smallest eigenvalue of M, we have $\lambda_n = \min\{\langle Mv, v\rangle : \|v\|_2 = 1\}$. So if we choose $v = \left(\frac{1}{\sqrt{n}}, \ldots, \frac{1}{\sqrt{n}}\right) \in \mathbb{R}^n$, we obtain

$$\lambda_n \leq \langle Mv, v\rangle = \frac{(1-p)e(G) - pe(\overline{G})}{n} \leq -p(1-p)n/4.$$

This implies that $\sum_{i=1}^n \lambda_i^{2k} \geq \lambda_n^{2k} = \Omega(p^{2k}(1-p)^{2k}n^{2k})$, thereby proving the lower bound in the second case.

On the other hand, for the upper bound we have two constructions. Again we assume that $p \leq \frac{1}{2}$. The first construction is very simple: the empty graph. If G is empty, a straightforward computation shows that

$$\|G - p\|_{C_{2k}} = p \leq 2p(1-p),$$

and this proves the upper bound whenever $p \leq n^{-(k-1)/k}$. For the second construction, we consider $G \sim \mathbb{G}_{n,p}^{\text{loop}}$ to be a random graph with loops, where every possible pair or loop belongs to G independently with probability p. Here we assume that $p > n^{-(k-1)/k}$. Let $X = n^{2k}\|G - p\|_{C_{2k}}^{2k}$. Write $X = \sum_{f:\mathbb{Z}/2k\mathbb{Z} \to V(G)} X_f$, where $X_f = \prod_{i \in \mathbb{Z}/2k\mathbb{Z}} M_{f(i),f(i+1)}$ and $M = A - pJ$ is as before. Then the expectation of X_f is 0 unless for every i there is $j \neq i$ with $\{f(j), f(j+1)\} = \{f(i), f(i+1)\}$, that is, every edge of C_{2k} is glued with some other edge. If f is a map with $\mathbf{E}[X_f] \neq 0$ then the image under f of the edge set of C_{2k} is a connected multi-graph where every edge (or loop) appears with even multiplicity, so it contains at most $k+1$ vertices. Since the number of maps f for which the aforestated image of C_{2k} contains at most e distinct edges (ignoring multiplicity) is $O(n^{e+1})$, we have

$$\mathbf{E}[X] = O\left(\sum_{e=1}^{k} n^{e+1}p^e(1-p)^e\right) = O(n^{k+1}p^k(1-p)^k),$$

since $p > n^{-1}$. Now take an outcome G such that the value of X is at most its expected value. This finishes the proof of the proposition. □

A related result of Hatami [10] shows that a complete bipartite graph $F = K_{2k,2m}$ with both part sizes, $2k$ and $2m$, being even also gives a norm by a version of (1.3). This norm, for $G - p$, is $\|G - p\|_F := t(F, M)^{1/(2k+2m)} = n^{-1}X^{1/(2k+2m)}$, where M is as in (4.1),

$$X = \sum_{f:A \cup B \to V(G)} \prod_{a \in A} \prod_{b \in B} M_{f(a),f(b)},$$

and A, B are fixed disjoint sets of sizes $2k$ and $2m$ respectively.

Proposition 4.1. *Let $F = K_{2k,2m}$ with $1 \leq k \leq m$. The minimum of $\|G - p\|_F$ over n-vertex graphs G (possibly with loops) is*

$$\Theta\left(\min\left\{p^{4km}(1-p)^{4km},\, p^{2km}(1-p)^{2km}n^{-k}\right\}^{1/(2m+2k)}\right).$$

Proof. For the same reasons stated in the beginning of the proof of Proposition 1.2 we may assume, without loss of generality, that $p \le \frac{1}{2}$. We begin with the lower bound. We rewrite X by grouping all maps $f : A \cup B \to V(G)$ by the restriction of f to A. For every fixed $h : A \to V(G)$, we have

$$\sum_{g:B\to V(G)} \prod_{a\in A} \prod_{b\in B} M_{h(a),g(b)} = \left(\sum_{u\in V(G)} \prod_{a\in A} M_{h(a),u} \right)^{2m} \ge 0.$$

As in the proof of Proposition 1.2, we divide the analysis into two cases. In the first case, we assume that $e(G) \ge p\binom{n+1}{2}/2$. Let $\mathcal{H}$ be the set of all $h : A \to V(G)$ such that $h(2i-1) = h(2i)$ for all $i \in [k]$, where we assumed $A := [2k]$. Note that $|\mathcal{H}| = n^k$. If $h \in \mathcal{H}$ we have

$$\sum_{u\in V(G)} \prod_{a\in A} M_{h(a),u} = \sum_{u\in V(G)} \prod_{i\in[k]} M^2_{h(2i),u}.$$

Thus by the convexity of $x \mapsto x^{2m}$ for $x \in \mathbb{R}$ and the convexity of $x \mapsto x^k$ for $x \ge 0$, we have

$$\begin{aligned}
X &= \sum_{h:A\to V(G)} \left(\sum_{u\in V(G)} \prod_{a\in A} M_{h(a),u} \right)^{2m} \ge \sum_{h\in\mathcal{H}} \left(\sum_{u\in V(G)} \prod_{i\in[k]} M^2_{h(2i),u} \right)^{2m} \\
&\ge n^k \left(\frac{1}{n^k} \sum_{h\in\mathcal{H}} \sum_{u\in V(G)} \prod_{i\in[k]} M^2_{h(2i),u} \right)^{2m} = n^k \left(\frac{1}{n^k} \sum_{u\in V(G)} \left[\sum_{v\in V(G)} M^2_{v,u} \right]^k \right)^{2m} \\
&\ge n^k \left(\frac{1}{n^{k-1}} \left[\frac{1}{n} \sum_{u\in V(G)} \sum_{v\in V(G)} M^2_{v,u} \right]^k \right)^{2m} = n^k \left(\frac{1}{n^{k-1}} \left[\frac{(1-p)^2 e(G) + p^2 e(\overline{G})}{n} \right]^k \right)^{2m} \\
&= \Omega\left(p^{2km} (1-p)^{2km} n^{k+2m} \right),
\end{aligned}$$

which proves the lower bound in the first case.

In the second case, we assume that $e(G) < p\binom{n+1}{2}/2$. By the convexity of $x \mapsto x^{2m}$ and $x \mapsto x^{2k}$ for all $x \in \mathbb{R}$, we have

$$\begin{aligned}
X &= \sum_{h:A\to V(G)} \left(\sum_{u\in V(G)} \prod_{a\in A} M_{h(a),u} \right)^{2m} \ge n^{2k} \left(\frac{1}{n^{2k}} \sum_{h:A\to V(G)} \sum_{u\in V(G)} \prod_{i\in[2k]} M_{h(i),u} \right)^{2m} \\
&= n^{2k} \left(\frac{1}{n^{2k}} \sum_{u\in V(G)} \left[\sum_{v\in V(G)} M_{v,u} \right]^{2k} \right)^{2m} \ge n^{2k} \left(\frac{1}{n^{2k-1}} \left[\frac{1}{n} \sum_{u\in V(G)} \sum_{v\in V(G)} M_{v,u} \right]^{2k} \right)^{2m} \\
&= n^{2k} \left(\frac{1}{n^{2k-1}} \left[\frac{(1-p)e(G) - pe(\overline{G})}{n} \right]^{2k} \right)^{2m} = \Omega\left(p^{4km} (1-p)^{4km} n^{2k+2m} \right),
\end{aligned}$$

which proves the lower bound in the second case.

We turn to the upper bound. We need two construction. The first one is again the empty graph. If G is empty then

$$\|G - p\|_F = p^{2km/(k+m)},$$

and this proves the upper bound whenever $p \le n^{-1/2m}$. The second construction is the random graph $G \sim \mathbb{G}^{\text{loop}}_{n,p}$. Write X as the sum of X_f over $f : A \cup B \to V(G)$. Each f with

$\mathbf{E}[X_f] \neq 0$ maps $E(K_{2k,2m})$ into a connected multi-graph where every edge appears with even multiplicity. Consider the equivalence relation on $A \cup B$ given by such f, where two vertices in $A \cup B$ are equivalent if their images under f coincide. If non-trivial classes miss some $a \in A$ and some $b \in B$, then $f(a)f(b)$ is a singly-covered edge, a contradiction. Thus, non-trivial classes have to cover at least one of A or B entirely, so the number of identifications is at least $\min\{|A|,|B|\}/2 = k$. It follows that the image of F under f has at most $k+2m$ vertices. In fact, if the image of F under f contains exactly $2k+2m-t$ vertices (where $t \geq k$), the number of distinct edges in the image of F by f is at least $4km-2mt$. This is because every "identification" of vertices under the same equivalence class of f can "destroy" at most $2m$ edges. Therefore

$$\mathbf{E}[X] = O\left(\sum_{t=k}^{2k+2m-1} n^{2k+2m-t} p^{4km-2mt}(1-p)^{4km-2mt}\right) = O(n^{k+2m}p^{2km}(1-p)^{2km}),$$

since $p > n^{-1/(2m)}$. Now take an outcome G such that the value of X is at most its expected value. This finishes the proof of the proposition. □

5 Concluding remarks and open questions

Observe that the result of Chung, Graham, Wilson [5] implies that there cannot be a graph G with $t(K_2, A) = p$ and $t(C_4, A) = p^4$ where $0 < p < 1$ and A is the adjacency matrix of G. (Indeed, otherwise the uniform blow-ups of G would form a quasirandom sequence, which is a contradiction.) This argument does not work with the subgraph count function $N(F, G)$. We do not know if the fact that $u_k(n,p)$ can be zero infinitely often for $k = 3$ (when p is rational) but not for $k = 4$ can directly be related to the fact that quasirandomness is forced by 4-vertex densities.

Let $\mathbb{G}_{n,m}$ be the random graph on $[n]$ with m edges, where all $\binom{\binom{n}{2}}{m}$ outcomes are equally likely. Janson [12] completely classified the cases when the random variable $N(F, \mathbb{G}_{n,m})$ satisfies the Central Limit Theorem where $n \to \infty$ and $m = \lfloor p\binom{n}{2} \rfloor$. He showed that the exceptional F are precisely those graphs for which $S^{(p)}(H, F) = 0$ for every H from the following set: connected graphs with 5 vertices and graphs without isolated vertices with 3 or 4 vertices. It is an open question if at least one such pair (F, p) with $p \neq 0, 1$ exists, see, e.g., [12, Page 65] and [13, Page 350]. Note that nothing is stipulated about $S^{(p)}(K_2, F)$. In fact, it has to be non-zero e.g. by Theorem 1.1; moreover, [12, Theorem 4] shows that, for given $v(F)$ and p, the number of edges in such hypothetical F is uniquely determined. This indicates that the general question of understanding possible joint behaviour of the S-statistics may be difficult.

It would be interesting to extend Theorem 1.1 to a wider range of p, or to other structures such as, for example, r-uniform hypergraphs.

References

[1] N. Alon and J. Spencer, *The probabilistic method*, 3d ed., Wiley Interscience, 2008.

[2] A. D. Barbour, M. Karoński, and A. Ruciński, *A central limit theorem for decomposable random variables with applications to random graphs*, J. Combin. Theory (B) **47** (1989), 125–145.

[3] B. Bollobás and A. D. Scott, *Discrepancy in graphs and hypergraphs*, in More sets, graphs and numbers, Ervin Gyori, Gyula O.H. Katona and Laszlo Lovász, eds, pp. 33–56, Bolyai Soc. Math. Stud. **15**, Springer, Berlin, 2006.

[4] ———, *Intersections of hypergraphs*, Journal of Combinatorial Theory, Series B **110** (2015), 180–208

[5] F. R. K. Chung, R. L. Graham, and R. M. Wilson, *Quasi-random graphs*, Combinatorica **9** (1989), 345–362.

[6] P. Erdős, *On combinatorial questions connected with a theorem of Ramsey and van der Waerden*, Mat. Lapok **14** (1963), 29–37.

[7] P. Erdős, M. Goldberg, J. Pach, and J. Spencer, *Cutting a graph into two dissimilar halves*, J. Graph Theory **12** (1988), 121–131.

[8] P. Erdős and J. Spencer, *Imbalances in k-colorations*, Networks **1** (1971/72), 379–385.

[9] W. Feller, *An introduction to probability theory and its applications*, 3d ed., Wiley-Intersci. Publ., 1968.

[10] H. Hatami, *Graph norms and Sidorenko's conjecture*, Israel J. Math. **175** (2010), 125–150.

[11] S. Janson, *A functional limit theorem for random graphs with applications to subgraph count statistics*, Random Struct. Algorithms **1** (1990), 15–37.

[12] ———, *Orthogonal decompositions and functional limit theorems for random graph statistics*, Mem. Amer. Math. Soc. **111** (1994), no. 534, vi+78.

[13] ———, *A graph Fourier transform and proportional graphs*, Random Struct. Algorithms **6** (1995), 341–351.

[14] S. Janson and J. Kratochvíl, *Proportional graphs*, Random Struct. Algorithms **2** (1991), 209–224.

[15] S. Janson and K. Nowicki, *The asymptotic distributions of generalized U-statistics with applications to random graphs*, Probab. Theory Related Fields **90** (1991), 341–375.

[16] S. Janson and J. Spencer, *Probabilistic construction of proportional graphs*, Random Struct. Algorithms **3** (1992), 127–137.

[17] J. Kärrman, *Existence of proportional graphs*, J. Graph Theory **17** (1993), 207–220.

[18] L. Lovász, *Large networks and graph limits*, Colloquium Publications, Amer. Math. Soc, 2012.