

Supersaturation of C_4 : from Zarankiewicz towards Erdős-Simonovits-Sidorenko

Zoltán Lóránt Nagy*

MTA–ELTE Geometric and Algebraic Combinatorics Research Group

ELTE Eötvös Loránd University, Budapest, Hungary

Department of Computer Science

H–1117 Budapest, Pázmány P. sétány 1/C, Hungary

nagyzoli@cs.elte.hu

Abstract

For a positive integer n , a graph F and a bipartite graph $G \subseteq K_{n,n}$ let $F(n+n, G)$ denote the number of copies of F in G , and let $F(n+n, m)$ denote the minimum number of copies of F in all graphs $G \subseteq K_{n,n}$ with m edges. The study of such a function is the subject of theorems of supersaturated graphs and closely related to the Sidorenko-Erdős-Simonovits conjecture as well. In the present paper we investigate the case when $F = K_{2,t}$ and in particular the quadrilateral graph case. For $F = C_4$, we obtain exact results if m and the corresponding Zarankiewicz number differ by at most n , by a finite geometric construction of almost difference sets. $F = K_{2,t}$ if m and the corresponding Zarankiewicz number differ by $c \cdot n\sqrt{n}$ we prove asymptotically sharp results based on a finite field construction. We also study stability questions and point out the connections to covering and packing block designs.

Keywords: extremal graphs, supersaturation, quadrilateral, almost difference sets, designs, finite geometry

1 Introduction

To determine the minimum number of a subgraph F in certain graph families is one of the oldest problems in combinatorics. In fact, it dates back to the results of

*The author is supported by the Hungarian Research Grant (OTKA) No. K. 120154 and by the János Bolyai Research Scholarship of the Hungarian Academy of Sciences

Mantel and Turán [20], who investigated the case of cliques ($F = K_p$) to describe the maximum number of edges of n -vertex graphs for which the number of (not necessarily induced) subgraphs F can still be zero. These results initiated the study of the function

$$\text{ex}(n, F) = \max\{|E(G)| : |V(G)| = n, F \not\subseteq G\},$$

and the so called *extremal graphs* for which the number of edges coincides with the extremal function; furthermore in general, extremal graph theory.

Let us denote by $F(n, G)$ the number of F -subgraphs in the n -vertex graph G , and denote by $F(n, m)$ the minimum number of F -subgraphs in n -vertex graphs having m edges. Clearly, $F(n, m) = 0$ for $m = \text{ex}(n, F)$, while $F(n, \text{ex}(n, F) + 1) > 0$ holds. Following the pioneer unpublished result of Rademacher (see [10]), Erdős and Simonovits [12] started to investigate the order of magnitude of the function $F(n, m) : m > \text{ex}(n, F)$ for arbitrary F and called theorems which asserts that a graph $G = G_n$ contains very many graphs F from a family \mathcal{F} *theorems on supersaturated graphs*. Such theorems are not only interesting in themselves, but also are often useful in establishing other extremal results.

If the number of edges is equal to $m = \text{ex}(n, F) + k$, then we call k the *excess* (over the extremal number).

In the case $F = K_{p+1}$, Erdős [11] proved a stability result when the excess is small.

Theorem 1.1. *For every fixed p there exists a constant $c_p > 0$ such that for any $m = e(T_{n,p}) + k < e(T_{n,p}) + c_p n$,*

$$K_{p+1}(n, m) = K_{p+1}(n, T_{n,p}^{(+k)})$$

holds, where $T_{n,p}^{(+k)}$ is obtained from the Turán graph $T_{n,p}$ by putting k edges in one of its maximal classes so that the new edges form no triangle.

The following stability theorem became a milestone in the study of supersaturated graphs.

Theorem 1.2 (Lovász-Simonovits, [29]). *Let C be an arbitrary constant and let p be a fixed integer. There exist two constants $\delta > 0$ and C' such that in any graph G on n vertices and $m = e(T_{n,p}) + k$ edges ($0 < k < \delta \cdot n^2$), if the number of copies of K_{p+1} in G is at most Ckn^{p-1} then G can be obtained from the Turán graph $T_{n,p}$ by adding or deleting at most $C'k$ edges.*

In general, the cliques can be replaced by any graph F in the problem. It turns out that the analogue of Theorem 1.2 holds for every graph F which has chromatic number $p + 1 > 2$, due to Erdős and Simonovits [12, 20]. The result is based on the application of the celebrated Erdős–Stone–Simonovits theorem.

However if F is bipartite, the situation is significantly different, since the extremal numbers of bipartite graphs have exponents smaller than 2.

Concerning the case when F is bipartite, Erdős and Simonovits also proved

Theorem 1.3 (Erdős-Simonovits, on the number of complete bipartite graphs [12, 20]). *For any $a \leq b$, there exist two constants $c, \gamma > 0$ such that if an n -vertex graph G has $m > cn^{2-\frac{1}{a}}$ edges, then G contains at least*

$$\frac{\gamma \cdot m^{ab}}{n^{2ab-a-b}}$$

copies of $F = K_{a,b}$.

Definition 1.4. *We call $c, \gamma = \gamma(c) > 0$ the density parameters for $K_{a,b}$.*

Corollary 1.5 ([20]). *Let $c > 0$ be an arbitrary constant and G be an n -vertex with m edges. If $m > c \cdot \text{ex}(n, C_4)$, then G contains at least $\gamma m^4/n^4$ copies of C_4 , for some $\gamma > 0$.*

As it was noted in their paper, the random graph with m edges shows that this bound is sharp, which is a natural phenomenon in these kinds of problems. Two fundamental questions arise here. The first is to determine the exact value of γ depending on c . The second one is to find the largest number c' for which the theorem does not hold, and to characterize the number of copies of F appearing in graphs having $\text{ex}(n, F) \leq m \leq (c' + o(1))n^{2-\frac{1}{a}}$ edges.

Notice that the first question has a strong connection to the famous Sidorenko conjecture [32] which was in fact proposed earlier in a slightly different form by Erdős and Simonovits [12]. We consider a discrete form similar to that of Erdős and Simonovits.

Conjecture 1.6 (Erdős-Simonovits, Sidorenko [12]). *Let H be a bipartite graph with vertex sets a_1, \dots, a_r and b_1, \dots, b_s . Let the number of edges of H be m_0 . Let G be a bipartite graph of density α with vertex sets X and Y and let ϕ and ψ be random functions from $\{a_1, \dots, a_r\}$ to X and from $\{b_1, \dots, b_s\}$ to Y . Then the probability that $\phi(a_i)\psi(b_j)$ is an edge of G for every pair i, j such that $a_i b_j$ is an edge of H is at least α^{m_0} .*

Note that in this variant, we not only consider the copies of H but also its images under homomorphisms (edge-preserving mappings) with appropriate weights. This difference between the statement of Conjecture 1.6 and Theorem 1.3 can be essential only if the host graph G is not dense.

The Sidorenko-Erdős-Simonovits conjecture was confirmed for several bipartite graph families, including complete bipartite graphs (see for instance [7, 16, 25, 27]), but is still widely open. One can observe here that in the *dense case*, when $m = \Omega(n^2)$, the

density parameters happen to be the same. Also, this conjecture reveals the reason why Theorem 1.3 was restricted to complete bipartite graphs: the reason was not only our lack of knowledge on the exact value of $\text{ex}(n, F)$ in general for bipartite F , but the structure of graphs with few copies of F in dense graphs is not known as well.

Our aim is to improve the bound in Theorem 1.3 in the *sparse case*, that is, when $|E(G)| = o(n^2)$. In order to do that, from now on we consider only graphs G which are subgraphs of a balanced bipartite graph $K_{n,n}$. We introduce some notations.

Notation 1.7.

- $F(n+n, G)$ denotes the number of F -subgraphs in the bipartite graph G on partition classes of size n .
- $F(n+n, m)$ denotes the minimum number of F -subgraphs a graph G can have where $G \subseteq K_{n,n}$ has m edges. $d(i)$ denotes the degree of a vertex i in G
- $d(X) := d_G(X)$ denotes the co-degree of X , i.e. the number of common neighbors of the vertices of $X \subseteq V(G)$ in G . For the degree of a single vertex y we use the standard notion $d(y)$.

Remark 1.8. The definition of $F(n+n, m)$ implies that

$$z(n, n, a, b) = \max\{m \mid K_{a,b}(n+n, m) = 0\}$$

holds for the well known Zarankiewicz-number, see [20].

In this paper we mainly focus on the questions concerning the density parameters described above for the simplest non-trivial bipartite graph, the quadrilateral.

In the spirit of Theorem 1.3 we establish a general lower bound on $C_4(n+n, m)$ in the next section, which points out the dependence of the density parameters c, γ as follows.

Theorem 1.9. Suppose $m = n(\sqrt{n} + \frac{1}{2}) + \xi(n)$ for $\xi(n) \geq 0$. Then

- (i) $C_4(n+n, m) = \Omega(n)$;
- (ii) $\sqrt{n} \ll \xi(n) \ll n\sqrt{n}$ implies $C_4(n+n, m) \geq (\frac{1}{2} + o(1))\sqrt{n}\xi(n)$;
- (iii) $\xi(n) = Cn\sqrt{n}$ implies $C_4(n+n, m) \geq \left(\frac{C(C+2)(1+C)^2}{4} + o(1)\right)n^2$;
- (iv) $\xi(n) \gg n\sqrt{n}$ implies $C_4(n+n, m) = (1 + o(1))\frac{1}{4}\left(\frac{m}{n}\right)^4$.

Note that sharp lower bounds on $C_4(n+n, m)$ have been proved previously by Allen et. al. [1] and by Füredi and Özkahya [19].

After the proof of Theorem 1.9 we point out the conditions under which the general lower bound is met, namely if there exists a so-called symmetric (almost) design with required parameters. The latter fact in turn implies that for supersaturated extremal problems concerning bipartite graphs, we cannot count on unique stability results similar to Theorem 1.2, as several infinite families of parameters $2 - (v, k, \lambda)$ exist with many non-isomorphic block designs [4].

In Section 3. we investigate the case when the excess is small, i.e. case (ii) in Theorem 1.9. It turns out that in this case some sort of stability does exist, and it relies on some finite geometry argument which leads to the discovery of a new almost difference set family. Then in the next Section we examine the case when the excess is the same order of magnitude as the Zarankiewicz number $z(n, n, 2, 2)$ (case (iii)), and prove the following asymptotic result.

Theorem 1.10. *For any fixed positive integer k , if $m = (\sqrt{k} + o(1))n\sqrt{n}$, then*

$$\frac{C_4(n+n, m)}{n^2} \rightarrow \frac{k(k-1)}{4}$$

while the random balanced bipartite graph contains $\frac{k^2}{4}n^2$ quadrilaterals.

We also extend this theorem for the number of graphs $K_{2,t}$.

Finally in Section 5 we put our results into perspective and point out the connections of this problem with clique-packing or covering problems in graph theory, discussing some open questions.

2 Lower bound on the number of C_4 s in terms of the number of edges

Theorem 2.1 (Theoretical lower bound). *Let G be a subgraph of $K_{n,n}$ with m edges on partition classes X and Y . Then the number of $F = K_{a,b}$ subgraphs in G is at least*

$$\binom{n}{a} \binom{n \binom{\bar{d}}{a} \cdot \binom{n}{a}^{-1}}{b},$$

provided $A \subseteq X$ and $B \subseteq Y$, where $\bar{d} = m/n$ is the average degree in G .

Here we consider the truncated extension of the binomials for which $\binom{n}{k} = \frac{\prod_{i=0}^{k-1} (n-i)}{k!}$ if $n+1 \geq k \geq 1$, otherwise it is zero. Note that this function of n is convex for fixed k if $n > 0$.

Although it is essentially the application of the proof idea of the Kővári-Sós-Turán theorem, we give a proof for the sake of completeness.

Proof. We may fix first a set of cardinality a in X and count the those sets in Y which have exactly b vertices, which are all adjacent to the vertices of the a -set; then we can sum this up to all a -sets. This provides

$$K_{a,b}(n+n, G) = \sum_{A \subseteq X, |A|=a} \binom{d(A)}{b}.$$

By applying Jensen's inequality, we get

$$K_{a,b}(n+n, G) \geq \binom{n}{a} \left(\sum_{A \subseteq X, |A|=a} d(A) \binom{n}{a}^{-1} \right).$$

Note that the sum of the co-degrees counts the way one can choose a vertex from Y and a neighbors of that vertex. Hence we obtain

$$\sum_{A \subseteq X, |A|=a} d(A) = \sum_{y \in Y} \binom{d(y)}{a} \geq n \binom{n^{-1} \sum_y d(y)}{a}$$

after applying Jensen's inequality again. This yields the stated expression. \square

This result enables us to prove Theorem 1.9 to reveal the connection between the density parameters γ and c in Corollary 1.5 for the graph $F = C_4$.

Proof of Theorem 1.9. The result follows from Theorem 2.1. Suppose that $m = n(\sqrt{n} + \frac{1}{2}) + \xi(n)$ for $\xi(n) \geq 0$. Then the lower bound implies that

$$K_{a,b}(n+n, m) \geq \binom{n}{2} \cdot \frac{1}{2} \frac{m(m-n)}{n^2(n-1)} \left(\frac{m(m-n)}{n^2(n-1)} - 1 \right).$$

Substituting $m = n(\sqrt{n} + \frac{1}{2}) + \xi(n)$ we obtain

$$K_{a,b}(n+n, m) \geq \frac{1}{4n^3(n-1)} m(m-n) (m(m-n) - n^2(n-1)) =$$

$$\frac{1}{4n^3(n-1)} (n^3 - n^2/4 + 2n\sqrt{n}\xi(n) + \xi(n)^2) (3n^2/4 + 2n\sqrt{n}\xi(n) + \xi(n)^2)$$

Thus the order of magnitude of $\xi(n)$ determines the leading term in the bound, providing the result. In the case (iv), balanced bipartite random graphs show that the lower bound is sharp, as the given bound equals the expected value of C_4 s in a balanced bipartite random graph of m edges. \square

In brief, we get that if $m/z(n, n, 2, 2) \rightarrow \infty$, then $C_4(n+n, m)$ indeed attained on the random graph apart from a smaller error term, while if $m/z(n, n, 2, 2) \rightarrow (1 + C)$ where $(C > 0)$, then $C_4(n+n, m)$ and the number of C_4 s in the random graph are of same order of magnitude, but the two numbers not equal asymptotically. Finally if $m/z(n, n, 2, 2) \rightarrow 1$, then $C_4(n+n, m)$ is much smaller than the number of C_4 s in the random graph.

2.1 Conditions for sharp results

The theoretical lower bound 2.1 can be a bit sharpened if we take into consideration that both degrees and co-degrees are integer numbers, hence we could apply the following discrete variant of the Jensen inequality.

Lemma 2.2 (Discrete Jensen inequality). *Let $f : \mathbb{Z} \rightarrow \mathbb{R}$ be a convex function. Then for any set of N evaluations,*

$$\sum_{i=1}^N f(x_i) \geq \alpha f\left(\left\lfloor \frac{\sum_i x_i}{N} \right\rfloor\right) + \beta f\left(\left\lceil \frac{\sum_i x_i}{N} \right\rceil\right),$$

where α and β are determined such that $\alpha + \beta = N$ and $\alpha \left\lfloor \frac{\sum_i x_i}{N} \right\rfloor + \beta \left\lceil \frac{\sum_i x_i}{N} \right\rceil = \sum_i x_i$.

If we apply Lemma 2.2 instead of Jensen's inequality, we got a slightly stronger lower bound that is independent of divisibility conditions of the parameters. We apply it for the case $F = K_{2,2}$, and refer to it as *improved theoretical lower bound*.

It is important to note that while the incidence graph of projective planes provides an infinite family of constructions where one can state sharp results for $a = b = 2$, the Zarankiewicz problem is notoriously hard in general thus only bounds are known for $z(n, n, a, b)$ when $b \geq a > 2$. Notably, the bound gained from the discrete Jensen inequality (see the paper of Roman [31]) is not sharp for every value n with $a, b > 2$ fixed. This follows from the improvement by Füredi [18] on the bound of the Zarankiewicz number $z(n, n, s, t)$, who proved that

$$z(n, n, s, t) \leq ((b - a + 1)^{1/a} + o(1))n^{2-1/a},$$

while the well known Kővári-Sós-Turán theorem closely related to Theorem 2.1 only give $z(n, n, s, t) \leq ((b - 1)^{1/a} + o(1))n^{2-1/a}$.

Interestingly, it can be improved in the case $a = b = 2$ as well for certain values of n due to Damásdi, Héger and Szőnyi (see [8]), and the bound on $z(n, n, a, b)$ in the same spirit of Theorem 2.1 turned out to be far from sharp even in this case for particular values of n .

From now on, we focus on the conditions when a graph attains the theoretical lower bound for the case $K_{a,b} = K_{2,2}$ in Theorem 2.1. The following statement is the

straightforward consequence of the two estimations via Jensen's inequality in the proof.

Corollary 2.3. *Suppose that a graph G has at least $n(\sqrt{n} + \frac{1}{2})$ edges. Then G attains the bound of Theorem 2.1 for $K_{a,b} = K_{2,2}$ if and only if*

- G is regular;
- every subset $A \subseteq X$ of size a has the same number of common neighbors, that is, $|d(A)| = |d(A')| \forall A, A' \subseteq X, |A| = |A'| = a$.

Proof. We applied Jensen's inequality for convex functions. If the function is strictly convex on the domain in view, then equality holds exactly when all terms are equal. Since the truncated binomial function $\binom{n}{2}$ is strictly convex above $n = 1$, which means that the Theoretical lower bound is sharp if and only if all degrees in Y are equal and every set $A \subset X$ of size $a = 2$ has equal co-degree too. However, since the expression in Theorem 2.1 is symmetric with respect to swapping a and b if $a = b$, moreover $K_{a,b}(n+n, G) = \sum_{A \subseteq X, |A|=a} \binom{d(A)}{b} = \sum_{B \subseteq Y, |B|=b} \binom{d(B)}{a}$, we obtain equality for the degrees in X and the co-degrees in Y , too. \square

Observe that conditions of the structure of graph attaining the improved theoretical lower bound with Lemma 2.2 can be derived similarly.

Corollary 2.4. *Suppose that a graph G has at least $n(\sqrt{n} + \frac{1}{2})$ edges. Then G attains the improved theoretical lower bound for $K_{2,2}$ if and only if*

- G is almost regular, that is, $|d(v) - d(v')| \leq 1 \quad (\forall v, v' \in V(G))$;
- every subset $A \subseteq X$ of size a has almost the same number of common neighbors, that is, $||d(A)| - |d(A')|| \leq 1 \quad \forall A, A' \subseteq X, |A| = |A'| = a$.

2.2 Case of equality with the theoretical lower bound

In this preliminary subsection we make the connection between balanced bipartite graphs attaining the theoretical lower bound and symmetric block designs or symmetric almost designs, furthermore we recall the definitions of these structures, together with some useful observations which will be essential in Section 3.

Definition 2.5 (Design). $\mathcal{D}(\mathcal{P}, \mathcal{B})$ is a symmetric $2 - (v, k, \lambda)$ block design if \mathcal{P} is a set of points, \mathcal{B} is a set of k -uniform blocks, with

- $|\mathcal{P}| = v = |\mathcal{B}|$, and
- for every pair of points $\{x, y\} \subset \mathcal{P}$ the number of incident blocks to this pair $|B_i \in \mathcal{B} : x, y \in B_i| = \lambda$.

Note that the equality $v\binom{k}{2} = \binom{v}{2}\lambda$ connects the parameters of symmetric designs. Corollary 2.3 directly implies the following.

Corollary 2.6 (Every design provides equality). *In the incidence graphs of a symmetric $2 - (v, k, \lambda)$ block designs we have*

- $|d(i) - d(i')| = 0 \quad (\forall i, i' \in V(G));$
- $|d(\{j, k\}) - d(\{j', k'\})| = 0 \quad (\forall \{j, k\}, \{j', k'\} \in \binom{V(G)}{2}),$ thus these graphs provide equality in Theorem 2.1

A bit more general concept is the so-called almost design or *adesign*.

Definition 2.7 (Adesign). $\mathcal{D}(\mathcal{P}, \mathcal{B})$ is a symmetric $2 - (v, k, \lambda)$ block adesign (almost design) if \mathcal{P} denotes a set of points, \mathcal{B} is a set of k -uniform blocks, with

- $|\mathcal{P}| = v = |\mathcal{B}|,$ and
- for every pair of points $\{x, y\} \subset \mathcal{P}$ the number of incident blocks to this pair $|B_i \in \mathcal{B} : x, y \in B_i|$ equals λ or $\lambda + 1$.

Similarly to the designs, it is easy to see that the more general Corollary 2.4 implies

Corollary 2.8 (Every Adesign provides equality). *In the incidence graphs of symmetric $2 - (v, k, \lambda)$ adesigns we have*

- $|d(i) - d(i')| = 0 \quad (\forall i, i' \in V(G));$
- $|d(\{j, k\}) - d(\{j', k'\})| \leq 1 \quad (\forall \{j, k\}, \{j', k'\} \in \binom{V(G)}{2}),$ thus these graphs provide equality in the improved theoretical lower bound.

Various infinite families of designs and adesigns exist, see [4] for recent surveys. For our purposes, we describe a special family of designs, which are connected to difference sets.

Definition 2.9. A (v, k, λ) difference set is a subset D of size k of (an additive) group G of order v such that every non-zero element of G can be expressed as a difference $d - d'$ with a pair of elements of D in exactly λ ways. A difference set D is said to be cyclic or Abelian if the group G has the corresponding property.

A difference set with $\lambda = 1$ is called *planar*, and it is well known due to Singer that there always exist planar difference sets for $G = \mathbb{Z}_n$ with $n = q^2 + q + 1$ where q is some power of a prime, as cyclic projective planes provide such examples. For further details, we refer to [4, 9].

Definition 2.10. A (v, k, λ) almost difference set is a subset D of size k of (an additive) group G of order v such that every non-zero element of G can be expressed as $d - d'$ with a pair of elements of D in exactly λ or $\lambda + 1$ ways, with both cases appearing.

Remark 2.11. *If D is an (almost) difference set, and $g \in G$, then $-D = \{-d : d \in D\}$ is also an (almost) difference set, furthermore $g + D = \{g + d : d \in D\}$ is also an (almost) difference set, and is called a translate of D . The set of translates of D yields a $2 - (v, k, \lambda)$ design (or adesign).*

A key concept concerning difference sets is the multiplier.

Definition 2.12. *A multiplier of a difference set D in group G is a group automorphism ϕ of G such that $D^\phi = g + D$ for some $g \in G$. If G is Abelian and ϕ is the automorphism that maps $h \rightarrow t \cdot h$, then t is called a (numerical) multiplier.*

3 $F = C_4$, when the excess is small - connection to finite geometries

In this section we prove that slightly above the Zarankiewicz number (in cases (i) and (ii) of Theorem 1.9) it is possible to attain the improved lower bound by adding edges to the incidence graph $G(\Pi_q)$ of a projective plane Π_q . On the other hand, this extension method only works in a rather small domain. Finally we describe the incidence structure of points and lines corresponding to the added edges. For an introduction on finite geometries, we refer to [3].

Theorem 3.1. *If $n = q^2 + q + 1$ for a prime power $q = p^h$ and $m \leq z(n, n, 2, 2) + n$, then $C_4(n + n, m)$ meets the improved theoretical lower bound.*

Proof. It is clearly enough to show a construction for $m = z(n, n, 2, 2) + n$ which contains the incidence graph of a projective plane Π_q of order q . To this end, we will add a matching to $G(\Pi_q)$ between non-incident points and lines in the following way. Consider a planar difference set D corresponding to a cyclic projective plane of order q . We would like to add another element g of \mathbb{Z}_n to D such that $D \cup g$ is an almost difference set, i.e. every non-zero element appears once or twice as a difference. Observe that g cannot be added if and only if g is already an element of D or $d - g = g - d' \Leftrightarrow d + d' = 2g$ for a pair of different elements in D . Hence $n - |D| - \binom{|D|}{2} = \binom{q}{2}$ elements can complete D to an almost difference set D^* of size $q + 2$. We call such elements *completion elements* (w.r.t. D).

The incidence graph of the adesign formed by the translates of D^* thus provides such an example in view of Corollary 2.8. \square

Remark 3.2. *The constructed $D \cup g$ provides a new infinite family of almost difference sets with parameters $(q^2 + q + 1, q + 2, 1)$.*

Remark 3.3. *Note that during the completion, we used the fact that $2 \nmid n$. Also, a slightly more involved calculation shows that any difference set with $\lambda = 2$ can be completed by an element to an almost different set.*

In the geometric point of view, as the translates of the planar difference set D are corresponding to the lines of the plane, one might ask about the structure of the points that can be added to the lines to obtain the desired almost difference set with its translates. This geometric structure of the completion elements is described in the following theorems.

Theorem 3.4. *If the order of the projective plane is even, then the completion elements of the difference set D are the points not incident to a dual hyperoval.*

Theorem 3.5. *If the order of the projective plane is odd, then the completion elements of the difference set D are the points not incident to a set of $q+1$ ovals which pairwise intersect each other in exactly one point.*

Recall that an oval, resp. hyperoval is a set of $q+1$, resp. $q+2$ points in the projective plane of order q no three of which are collinear. This also implies that any line intersects the hyperoval either in 0 or 2 points, hence hyperovals are maximal (k, n) arcs as well. Hyperovals only exist if the order is even, and in that case every oval ($q+1$ -arc) extends uniquely to a hyperoval. (For further details, see [3].)

A key ingredient in the proofs is the following algebraic result due to M. Hall [21].

Result 3.6 (Hall [21], see also [24]). *Let D be an Abelian difference set for a projective plane of order q . Then 2 and $1/2$ are (numerical) multipliers of D if and only if q is even. Moreover, $\frac{1}{2}D$ is the set of absolute points of a certain polarity, and it forms an oval if q is odd or a line if q is even.*

Proof of Theorem 3.4 and 3.5. In the proof of Theorem 3.1, we saw that $g \in \mathbb{Z}_n$ is not a completion element if and only if $2g = d + d'$ for a pair of (not necessarily distinct) elements d, d' of D , hence these elements g have the form $\frac{d+d'}{2}$. This implies that the set of such elements are determined by the union of the some translates of $\frac{D}{2}$, namely

$$\bigcup_{d \in D} \left\{ \frac{d}{2} + \frac{D}{2} \right\}.$$

Note that these translates intersect each other in exactly one element and no three of them share a common element. It follows from Hall's result that if q is even, then these are translates of D since $\frac{1}{2}$ is a multiplier, hence they are lines as well, which form a dual hyperoval together with D .

If q is odd, then these are translates of an oval according to Hall's result, again with no three of them share a common element and no two of them have more than one common element, completing the proof. (We note that such a structure is called a projective bundle and has many interesting properties, see [24] and references therein.) \square

Via Theorem 3.1 one can add n edges to the incidence graph of the projective plane in many ways to get a graph which attains the improved theoretical lower bound corresponding to the number of C_4 s. However, if the excess over $z(n, n, 2, 2)$ is more than n , this is not the case.

Proposition 3.7. *Suppose that $n = q^2 + q + 1$, then $C_4(n + n, G)$ cannot meet the improved theoretical lower bound on graphs G with m edges containing $G(\Pi_q)$ as a subgraph if $z(n, n, 2, 2) + n < m \leq \sqrt{2} \cdot z(n, n, 2, 2)$.*

Proof. Suppose to the contrary that there exists such a construction with excess more than n . The edges not contained in $G(\Pi_q)$ can be considered as newly introduced incidences between lines and points. Since the excess is more than n , there exists a point P which is now joined with two lines e and f for which $P \not\sim e$ and $P \not\sim f$ in Π_q . Lines e and f have a unique intersection $Q = e \cap f$ in Π_q , while P and Q determine a unique line g in Π_q , different from e and f . Consequently, $N(P, Q) \supseteq \{e, f, g\}$ in the construction, which can only happen in an extremal construction if every point pair shares at least two common lines, thus the statement follows. \square

We remark in the end of this section that a recent result of Ferber, Hod, Krivelevich and Sudakov [15] proved the existence of certain $2 - (n, k, 1)$ almost designs. However, their result cannot be applied to our case since they assume that the number of blocks is much larger than the number of points, but it could be applied for $F(n, n', m)$ to determine the minimum number of C_4 s in m -edge subgraphs of very much unbalanced bipartite graphs $K_{n, n'}$ ($n \ll n'$).

4 When the excess is of order $n\sqrt{n}$ - generalizing the ideas of Mörs

In his paper [30], Mörs proved an asymptotic result for the Zarankiewicz number $z(n, n, 2, k + 1)$. In this subsection we show a family of graphs G for which $K_{2,2}(n + n, G) = K_{2,2}(n + n, m)$ for $m = O(n\sqrt{n})$. The idea behind the construction is the same, namely to exploit the additive and multiplicative structures of a finite field \mathbb{F}_q , but the construction below is defined in a much simpler manner and also is more general. This idea appeared also in the paper of Füredi [17] concerning $ex(n, K_{2,t})$.

Construction 4.1. *Let G be a bipartite graph on $V_1 \cup V_2$ with edge set $E := E(G)$ as follows.*

$V_1 = V_2 = A \times B$ where $A \sim \mathbb{F}_q$ and $B = \{1, 2, \dots, \frac{q-1}{k}\}$, furthermore

$$\{(a, b), (\alpha, \beta)\} \in E \Leftrightarrow \exists j \in \mathbb{Z}_k : g^\beta a + g^b \alpha = g^{j \cdot \frac{q-1}{k}} \text{ in } \mathbb{F}_q \quad (4.1)$$

for a fixed primitive root $g \in \mathbb{F}_q^\times$. Here $(a, b) \in V_1$ and $(\alpha, \beta) \in V_2$.

We refer to this graph G , depending on parameters q and k as $G^{(q,k)}$

Proposition 4.2. *The graph in Construction 4.1 is regular and $E(G^{(q,k)}) = \frac{q(q-1)^2}{k}$.*

Proof. Our aim is to prove that for every vertex $v = (a, b) \in V_1$, $d(v) = q - 1$. Pick an arbitrary vertex (a, b) . The defining equation 4.1 yields that for every choice $\beta \in B, j \in \mathbb{Z}_k$, exactly one $\alpha \in A$ fulfills the condition. The proposition follows if we observe that different pairs $(\beta, j), (\beta, j')$ cannot determine the same $\alpha \in A$. \square

Proposition 4.3. *The co-degree for any pair of vertices is either 0 or k in $G^{(q,k)}$, and $K_{2,2}(\frac{q(q-1)}{k}, G^{(q,k)}) = \frac{q(q-1)^3}{4} (1 - \frac{1}{k})$.*

Proof. We determine the co-degree of each pair $v = (a, b), v' = (a', b')$ from V_1 .

Case 1. $a = a' = 0, b \neq b'$. In that case, we look for the possible solutions of the system of equations

$$\begin{cases} \exists j \in \mathbb{Z}_k : g^b \alpha = g^{j \cdot \frac{q-1}{k}} \\ \exists j' \in \mathbb{Z}_k : g^{b'} \alpha = g^{j' \cdot \frac{q-1}{k}} \end{cases}$$

A solution (α, β) is called *admissible*, if $\alpha \in A$ and $\beta \in B$ are both satisfied. Clearly $\alpha \neq 0$ must hold, but that yields $g^{b-b'} = g^{(j-j') \cdot \frac{q-1}{k}}$, which is impossible, so the co-degree is 0 in this case.

Case 2. $a \neq 0$. We can rewrite the system of equations as

$$\begin{cases} \exists j \in \mathbb{Z}_k : g^\beta a' + \frac{a'}{a} g^b \alpha = \frac{a'}{a} g^{j \cdot \frac{q-1}{k}} \\ \exists j' \in \mathbb{Z}_k : g^\beta a' + g^{b'} \alpha = g^{j' \cdot \frac{q-1}{k}}. \end{cases}$$

Subtracting the two equations we get

$$\left(\frac{a'}{a} g^b - g^{b'} \right) \alpha = \frac{a'}{a} g^{j \cdot \frac{q-1}{k}} - g^{j' \cdot \frac{q-1}{k}}.$$

Case 2A. $\frac{a'}{a} g^b - g^{b'} = 0$. This implies $\frac{a'}{a} = g^{j^* \cdot \frac{q-1}{k}}$ for some $j^* \in \mathbb{Z}_k$. Comparing this remark with the subcase condition, we get a contradiction, as $\frac{a'}{a} = g^{b'-b}$, however to obtain an admissible pair b, b' we must have $\frac{a'}{a} = 1$, but then $b' = b$ would also follow.

Case 2B. $\frac{a'}{a} g^b - g^{b'} = \gamma \neq 0$. This allows us to express α as

$$\alpha = \frac{1}{\gamma} \left(\frac{a'}{a} g^{j \cdot \frac{q-1}{k}} - g^{j' \cdot \frac{q-1}{k}} \right), \quad (4.2)$$

and then we obtain g^β as

$$g^\beta = (g^{j' \cdot \frac{q-1}{k}} - g^{b'} \alpha) / a'. \quad (4.3)$$

Clearly, α and thus β depend on the choice of j and j' , so we have to guarantee that these solutions are admissible. We claim that for any fixed j, j' , exactly one choice from $\{j + i, j' + i\} : i \in \mathbb{Z}_k$ provides an admissible pair. Indeed, if α and g^β are determined by j and j' , then $j + i$ and $j' + i$ give $g^{i \cdot \frac{q-1}{k}} \alpha$ and $g^{i \cdot \frac{q-1}{k}} g^\beta$, hence this set gives exactly one admissible candidate for β .

Observe on the other hand, that these admissible pairs provided by some j and j' are different. Indeed, a fixed α and g^β determine exactly j' (in the equation 4.3) and then together with j' , j is also determined exactly (in the equation 4.2). Thus all in all, we get that there exist exactly k common neighbors for $v = (a, b), v' = (a', b')$ in Case 2B.

This gives

$$\frac{1}{2} |V_1| \cdot \left(|V_1| - \frac{q-1}{k} \right) \binom{k}{2} = \frac{q(q-1)^3}{4} \left(1 - \frac{1}{k} \right) \quad (4.4)$$

$K_{2,2}$ subgraphs, taking into consideration that for every (a, b) , exactly $\frac{q-1}{k}$ different other vertex from V_1 has the property that their pairwise co-degree with (a, b) is zero according to Case 1 or Case 2A. \square

Remark 4.4. *It is easy to see that if the edge set of $G^{(q,k)}$ was defined by*

$$\{(a, b), (\alpha, \beta)\} \in E \Leftrightarrow \exists j \in \mathbb{Z}_k : g^\beta a + g^b \alpha = g^\delta g^{j \cdot \frac{q-1}{k}} \quad \text{in } \mathbb{F}_q \quad (4.5)$$

for some $\delta \in \mathbb{Z}$, then the proof of Proposition 4.3 would work out exactly the same way.

Comparing Proposition 4.2, 4.3 to Theorem 1.9, the proof of Theorem 1.10 easily follows.

Proof of Theorem 1.10. Choose a prime p for which $p \equiv 1 \pmod{k}$. Then Construction 4.1 on $n + n = 2 \frac{p(p-1)}{k}$ vertices contains $\frac{p(p-1)^2}{k} = \sqrt{k} \sqrt{\frac{p-1}{p}} n \sqrt{n}$ edges and $\frac{p(p-1)^3}{4} \left(1 - \frac{1}{k} \right) = \frac{\sqrt{k}^2 (\sqrt{k}+1)(\sqrt{k}-1)}{4} \frac{p-1}{p} n^2$. This shows that the construction provides asymptotically sharp result for every fix $k \in \mathbb{Z}$ if we take into consideration the density theorem of 1 (mod k) primes due to Huxley [22]. \square

In Proposition 4.3 we prove that there exist exactly k common neighbors for a pair $v, v' \in V_1$ which satisfies the condition of Case 2B, otherwise their co-degree is zero. This enables us to generalize the statement of Theorem 1.10 to provide asymptotic results on the number of $K_{2,t}$ graphs in a similar manner.

Theorem 4.5. *For any fixed positive integers $t > 2$ and k , if $m = (\sqrt{k} + o(1))n\sqrt{n}$, then*

$$\frac{K_{2,t}(n+n, m)}{n^2} \rightarrow \binom{k}{t}.$$

Proof. From the theoretical lower bound Theorem 2.1 one can derive that either the trivial lower bound $K_{2,t}(n+n, m) \geq 0$ in the case $k < t$ or is at least $2\binom{k}{t}\binom{n}{2}$, asymptotically. Note here that Theorem 2.1 counted the copies of the subgraphs with the prescription that $K_{2,t}$ is an ordered injection to $K_{n,n}$, which makes no difference in the case when $t = 2$, in contrary to the general case. This explains the factor 2 in the formula.

If $k < t$, the graph $G^{(q,k)}$ clearly contains no copy of $K_{2,t}$. On the other hand, Proposition 4.3 about the codegrees of $G^{(q,k)}$ implies that we can derive

$$K_{2,t}(n+n, G^{(q,k)}) = \frac{q(q-1)^3}{k^2} \binom{k}{t}$$

just as we got equation 4.4. Then the result follows. \square

5 Concluding remarks and related problems

Theorem 1.9 together with Theorem 4.5 suggest that in fact the following stronger version of the conjecture of Erdős-Simonovits and Sidorenko might also hold.

Conjecture 5.1. *For a bipartite graph F , if $m/ex_{bi}(n+n, F) \rightarrow \infty$ holds for the number m of edges in a balanced bipartite graphs, then the random graph with m edges has in expectation the asymptotically smallest number of copies of F .*

Observe that this is even stronger than the following

Conjecture 5.2. *(Erdős-Simonovits)[13, 33] For every bipartite graph F , there exist constants $\beta \in (0, 1), c > 0$ and n_0 such that graph G on $n \geq n_0$ vertices with $m \geq n^{2-\beta}$ edges contains at least $cn^{|V(F)|}p^{|E(F)|}$ copies of F where $p = \frac{m}{\binom{n}{2}}$ is the edge density of G .*

Note that the expected number of copies of H in a random graph $G(n, p)$ is roughly $\frac{1}{|Aut(H)|}n^{|V(H)|}p^{|E(H)|}$. Therefore the conjecture above in some sense asserts that random graphs contain (or, is close to containing) the minimum number of copies of F .

Two obvious ways to extend our results is to consider a sample graph F from larger set of bipartite graphs, and to consider the case when the host graph is K_n and

not $K_{n,n}$. The latter case, which is strongly connected to the investigated bipartite case is considered to be slightly more involved in general. Nevertheless, in the case of $F = K_{2,t}$, the construction of Füredi [17] provides a matching bound to the corresponding lower bound, in the case when the number of edges is of order $n\sqrt{n}$.

An interesting consequence of Theorem 2.1 is that while many extremal graph problems have stability versions, it is not quite the case for $F(n, m)$ in general. Indeed, several infinite families of block design parameters are known, for which many non-isomorphic block designs exist. This motivates the following problem.

Problem 5.3. *Bound the maximal edit distance between incidence graphs of non-isomorphic block designs with same parameters $2 - (v, k, \lambda)$, in terms of their parameters.*

The celebrated result of Wilson (see in [4]) states that if n is large enough, and some obvious divisibility conditions hold, then there exist block designs of parameters $2 - (v, k, \lambda)$.

However, n should be much larger than k^2 in the theorem, so it cannot be applied in general to obtain incidence graphs of block designs (approximately) attaining the supersaturation number $C_4(n + n, m)$ with some prescribed m . To overcome this, one may aim for almost uniform hypergraphs which cover every pair of vertices almost equal number of times.

More exactly, it would be interesting to see how close can we get to a design structure with hypergraphs having hyperedge size mostly k , while the covering number of every pair is mostly $\lfloor n\binom{k}{2}^{-1} \rfloor$ or $\lceil n\binom{k}{2}^{-1} \rceil$ in the most interesting case of Theorem 1.9 when $k = \Theta(\sqrt{n})$. Packing and covering designs represent known particular cases. A good construction would follow from an optimal $2 - (v, k, \lambda)$ *packing* or a $2 - (v, k, \lambda)$ *covering* in which every edge appears in at most or at least one block, respectively. (For details, we refer to [4, 6]). But so far, the corresponding well-known Schönheim and Johnson bound is not known to be asymptotically tight in the region in view.

Similarly to the case of almost designs, it would also be interesting to investigate the existence of almost difference sets, or structures close to an almost difference set, as well. Consider the following construction.

Construction 5.4. *Let \mathcal{G} be a finite Abelian group of order n , written additively. Take a subset $A \subset \mathcal{G}$ of size k . Consider the balanced bipartite graph $G(\mathcal{G}, A)$ on vertex classes $\{A + g | g \in \mathcal{G}\}$ and $\{g | g \in \mathcal{G}\}$, where edges correspond to inclusion.*

Then for fixed n and k , the closest A is to be a difference set, the fewest copy of C_4 is determined by Construction 5.4. More precisely, let $h_t(\mathcal{G}, A)$ be defined as

$$h_t(\mathcal{G}, A) = \sum_{0 \neq g \in \mathcal{G}} (|(a, a') : a - a' = g, a, a' \in A|)^t.$$

Then one can derive the following statement.

Proposition 5.5. *If n is odd, then $C_4(n + n, G(\mathcal{G}, A)) = \frac{n}{4} (h_2(\mathcal{G}, A) - h_1(\mathcal{G}, A))$*

The condition on the parity of n follows from the fact that we must distinguish between the cases when g and $-g$ can actually coincide or not for some elements of \mathcal{G} , but similar formula can be derived in the even order case too.

Observe that $h_1(\mathcal{G}, A)$ is fixed if we prescribe n and k , thus actually minimizing $C_4(n + n, G(\mathcal{G}, A))$ is equivalent to minimize

$$\Psi_2(\mathcal{G}, A) := \sum_{0 \neq g \in \mathcal{G}} \left(|(a, a') : a - a' = g, a, a' \in A| - \frac{k(k-1)}{n-1} \right)^2$$

Note that $\frac{k(k-1)}{n-1}$ is simply the average occurrence of a difference, thus $\Psi_2(\mathcal{G}, A) = 0$ if and only if A is a difference set.

This proposition motivates the following problem.

Problem 5.6. *Prove a general upper bound on $\min_{A \subset \mathcal{G}, |A|=k} \Psi_2(\mathcal{G}, A)$.*

Putting into perspective, the supersaturation problem we studied can be also seen as problem where we search for the extreme number of graphs C_4 or $K_{2,t}$ in graphs without odd cycles and given edge density, which fits in as a particular case of a recently widely investigated general framework.

Problem 5.7. *Determine the largest number of subgraphs F in graphs of order n not containing a forbidden family \mathcal{F} of graphs.*

The most notable recent results in the direction concern the case $F = K_3$ and $\mathcal{F} = \{C_5\}$ [2, 5], and $F = K_m$ and $\mathcal{F} = \{K_{s,t}\}$ [2]. It has also extensions in extremal hypergraph theory as well [23, 28].

Acknowledgement

The author would like to thank Bence Csajbók and Francesco Pavese for their remarks concerning Theorem 3.4 and the anonymous referees whose suggestions improved the presentation of the paper.

References

- [1] Allen, P., Keevash, P., Sudakov, B., Verstraëte, J. (2014). Turán numbers of bipartite graphs plus an odd cycle. *Journal of Combinatorial Theory, Series B*, 106, 134–162.
- [2] Alon, N., Shikhelman, C. (2016). Many T copies in H-free graphs. *Journal of Combinatorial Theory, Series B*, 121, 146–172.

- [3] Ball, S. (2015). *Finite geometry and combinatorial applications* (Vol. 82). Cambridge University Press.
- [4] Beth, T., Jungnickel, D., Lenz, H. (1999). *Design theory* (Vol. 69). Cambridge University Press.
- [5] Bollobás, B., Győri, E. (2008). Pentagons vs. triangles. *Discrete Mathematics*, 308(19), 4332–4336.
- [6] Chee, Y. M., Colbourn, C. J., Ling, A. C., Wilson, R. M. (2013). Covering and packing for pairs. *Journal of Combinatorial Theory A*, 120(7), 1440–1449.
- [7] Conlon, D., Fox, J., Sudakov, B. (2010). An approximate version of Sidorenko’s conjecture. *Geometric and Functional Analysis*, 20(6), 1354–1366.
- [8] Damásdi, G., Héger, T., Szőnyi, T. (2013). The Zarankiewicz problem, cages, and geometries. *Annales Universitatis Scientiarum Budapestinensis de Rolando Eötvös Nominatae Sectio Mathematica*, 56(1), 3–37.
- [9] Ding, C., Pott, A., Wang, Q. (2014). Constructions of almost difference sets from finite fields. *Designs, codes and cryptography*, 72(3), 581–592.
- [10] Erdős, P. (1962). On a theorem of Rademacher-Turán. *Illinois J. Math*, 6, 1–3.
- [11] Erdős, P. (1962). On the number of complete subgraphs contained in certain graphs, *Magyar Tud. Akad. Mat. Kutató Int. Közl* 7(3), 459–464.
- [12] Erdős, P., Simonovits, M. (1983). Supersaturated graphs and hypergraphs. *Combinatorica*, 3(2), 181–192.
- [13] Erdős, P., Simonovits, M. (1984). Cube-supersaturated graphs and related problems. *Progress in graph theory* (Waterloo, Ont., 1982), 203–218.
- [14] Erdős, P., Simonovits, M. (1966). A limit theorem in graph theory. *Studia Sci. Math. Hungar.*, 1, 51–57.
- [15] Ferber, A., Hod, R., Krivelevich, M., Sudakov, B. (2014). A construction of almost Steiner systems. *Journal of Combinatorial Designs*, 22(11), 488–494.
- [16] Fox, J., Wei, F. (2017). On the Local Approach to Sidorenko’s Conjecture. *Electronic Notes in Discrete Mathematics*, 61, 459–465.
- [17] Füredi, Z. (1996). New asymptotics for bipartite Turán numbers. *Journal of Combinatorial Theory A*, 75(1), 141–144.
- [18] Füredi, Z. (1996). An upper bound on Zarankiewicz’s problem. *Combinatorics, Probability and Computing*, 5(1), 29–33.
- [19] Füredi, Z., Özkahya, L. (2009). On 14-cycle-free subgraphs of the hypercube. *Combinatorics, Probability and Computing*, 18(5), 725–729.
- [20] Füredi, Z., Simonovits, M. (2013). The history of degenerate (bipartite) extremal graph problems. In *Erdős Centennial* (pp. 169–264). Springer Berlin Heidelberg.
- [21] Hall, M. Jr. (1947). Cyclic projective planes. *Duke Mathematical Journal*, 14, 1079–1090.
- [22] Huxley, M. N. (1971). On the difference between consecutive primes. *Inventiones mathematicae*, 15(2), 164–170.
- [23] Jiang, T., Ma, J. (2016). Cycles of given lengths in hypergraphs. ArXiv preprint arXiv:1609.08212.
- [24] Jungnickel D., Vedder, K. (1984). On the geometry of planar difference sets. *European Journal of Combinatorics*, 5, 143–148.
- [25] Kim, J. H., Lee, C., Lee, J. (2016). Two approaches to Sidorenko’s conjecture. *Transactions of the American Mathematical Society*, 368(7), 5057–5074.

- [26] Kiss, Gy., Malnič, A.; Marušič, D. (1999). A New Approach to Arcs. *Acta Mathematica Hungarica*, 84(3), 181–188.
- [27] Li, J. X., Szegedy, B. On the logarithmic calculus and Sidorenko’s conjecture. To appear in *Combinatorica*.
- [28] Loh, P. S., Tait, M., Timmons, C., Zhou, R. M. (2017). Induced Turán numbers. *Combinatorics, Probability and Computing*, 1–15.
- [29] Lovász, L., Simonovits, M. (1983). On the number of complete subgraphs of a graph II. *Studies in pure mathematics*, 459–495.
- [30] Mörs, M. (1981). A new result on the problem of Zarankiewicz. *Journal of Combinatorial Theory A*, 31(2), 126–130.
- [31] Roman, S., A problem of Zarankiewicz *Journal of Combinatorial Theory A*, 1975 18(2), 187–198.
- [32] Sidorenko, A.F. (1991). Inequalities for functionals generated by bipartite graphs, *Diskret. Mat.* 3, 50–65 (in Russian), *Discrete Mathematics and its Applications*, 2, 489–504. (in English)
- [33] Simonovits, M. (1984). Extremal graph problems, degenerate extremal problems, and supersaturated graphs, *Progress in graph theory* (Waterloo, Ont., 1982), Academic Press, Toronto, ON, pp. 419–437.