Kneser ranks of random graphs and minimum difference representations

Zoltán Füredi[1][*] and Ida Kantor[2][†]

[1] Alfréd Rényi Institute of Mathematics, Budapest, Hungary
(e-mail: `z-furedi@illinois.edu`)
[2] Charles University, Prague
(e-mail: `idasve@gmail.com`)

Abstract

Every graph $G = (V, E)$ is an induced subgraph of some Kneser graph of rank k, i.e., there is an assignment of (distinct) k-sets $v \mapsto A_v$ to the vertices $v \in V$ such that A_u and A_v are disjoint if and only if $uv \in E$. The smallest such k is called the *Kneser rank* of G and denoted by $f_{\mathrm{Kneser}}(G)$. As an application of a result of Frieze and Reed concerning the clique cover number of random graphs we show that for constant $0 < p < 1$ there exist constants $c_i = c_i(p) > 0$, $i = 1, 2$ such that with high probability

$$c_1 n/(\log n) < f_{\mathrm{Kneser}}(G) < c_2 n/(\log n).$$

We apply this for other graph representations defined by Boros, Gurvich and Meshulam.

A *k-min-difference representation* of a graph G is an assignment of a set A_i to each vertex $i \in V(G)$ such that

$$ij \in E(G) \iff \min\{|A_i \setminus A_j|, |A_j \setminus A_i|\} \geq k.$$

The smallest k such that there exists a k-min-difference representation of G is denoted by $f_{\min}(G)$. Balogh and Prince proved in 2009 that for every k there is a graph G with $f_{\min}(G) \geq k$. We prove that there are constants $c_1'', c_2'' > 0$ such that $c_1'' n/(\log n) < f_{\min}(G) < c_2'' n/(\log n)$ holds for almost all bipartite graphs G on $n + n$ vertices.

1   Kneser representations

A representation of a graph G is an assignment of mathematical objects of a given kind (intervals, disks in the plane, finite sets, vectors, etc.) to the vertices of G in such a way that two vertices are adjacent if and only if the corresponding sets satisfy a certain condition (intervals intersect, vectors have different entries in each coordinate, etc.). Representations of various kinds have been studied extensively, see, e.g., [7], [10], the monograph [15], or from information theory point of view [13]. The representations considered in this paper are assignments $v \mapsto A_v$ to the vertices $v \in V$ of a graph $G = (V, E)$ such that the A_v's are (finite) sets satisfying certain relations.

The *Kneser graph* $\mathrm{Kn}(s, k)$ (for positive integers $s \geq 2k$) is a graph whose vertices are all the k-subsets of the set $[s] := \{1, 2, \dots, s\}$, and whose edges connect two sets if they are disjoint. An assignment $(A_1, \dots, A_n)$ for a graph $G = (V, E)$ (where $V = [n]$) is called a *Kneser representation of rank k* if each A_i has size k, the sets are distinct, and A_u and A_v are disjoint if and only if $uv \in E$.

Every graph on n vertices with minimum degree $\delta < n - 2$ has a Kneser representation of rank $(n - 1 - \delta)$. To see that, define the *co-star* representation $(A_1', \dots, A_n')$ of G. For every $i \in V(G)$, let A_i' be the set of the edges adjacent to i in the complement of G (this is the graph $\overline{G}$ with $V(\overline{G}) = V(G)$ and $E(\overline{G}) = \binom{V(G)}{2} \setminus E(G)$). We have $A_i' \cap A_j' = 1$ if $ij \notin E(G)$, otherwise $A_i' \cap A_j' = 0$, and the maximum size of A_i' is $n - 1 - \delta(G)$. To turn the co-star representation into a Kneser representation add pairwise disjoint sets of labels to the sets $A_1', \dots, A_n'$ to increase their cardinality to exactly $n - 1 - \delta(G)$. The resulting sets $A_1, \dots, A_n$ are all distinct, they have the same intersection properties as $A_1', \dots, A_n'$, and form a Kneser representation of G of rank $n - 1 - \delta(G)$.

Let $\mathcal{G}(n)$ denote the set of $2^{\binom{n}{2}}$ (labelled) graphs on $[n]$ and let $\mathcal{G}(n, k, \mathrm{Kneser})$ denote the family of graphs on $[n]$ having a Kneser representation of rank k. $G \in \mathcal{G}(n, k, \mathrm{Kneser})$ is equivalent to the fact that G is an *induced* subgraph of some Kneser graph $\mathrm{Kn}(s, k)$. We have
$$\mathcal{G}(n, 1, \mathrm{Kneser}) \subseteq \mathcal{G}(n, 2, \mathrm{Kneser}) \subseteq \cdots \subseteq \mathcal{G}(n, n - 1, \mathrm{Kneser}) = \mathcal{G}(n).$$

Let $f_{\mathrm{Kneser}}(G)$ denote the smallest k such that G has a Kneser representation of rank k. We have seen that $f_{\mathrm{Kneser}}(G) \leq n - \delta$. We show that there are better bounds for almost all graphs.

Theorem 1. *There exist constants $c_2 > c_1 > 0$ such that for $G \in \mathcal{G}(n)$ with high probability*
$$c_1 \frac{n}{\log n} < f_{\mathrm{Kneser}}(G) < c_2 \frac{n}{\log n}.$$

We will prove a stronger version as Corollary 12.

2   Minimum difference representations

In *difference representations*, generally speaking, vertices are adjacent if the representing sets are sufficiently different. As an example consider Kneser graphs, where the vertices are adjacent if and only if the representing sets are disjoint. There are other type of representations where one joins sets close to each other, e.g., *t-intersection representations* were investigated by M. Chung and West [6] for dense graphs and Eaton and Rödl [7] for sparse graphs. But these are usually lead to different type of problems, one cannot simply consider the complement of the graph.

This paper is mostly focused on k-min-difference representations (and its relatives), defined by Boros, Gurvich and Meshulam in [5] as follows.

Definition 2. *Let G be a graph on the vertices $[n] = \{1, \ldots, n\}$. A k-min-difference representation $(A_1, \ldots, A_n)$ of G is an assignment of a set A_i to each vertex $i \in V(G)$ so that*

$$ij \in E(G) \iff \min\{|A_i \setminus A_j|, |A_j \setminus A_i|\} \geq k.$$

Let $\mathcal{G}(n, k, \min)$ be the set of graphs with $V(G) = [n]$ that have a k-min-difference representation. The smallest k such that $G \in \mathcal{G}(n, k, \min)$ is denoted by $f_{\min}(G)$.

The co-star representation (which was investigated by Erdős, Goodman, and Pósa [8] in their classical work on clique decompositions) shows that $f_{\min}(G)$ exists and it is at most $n - 1 - \delta(G)$.

Boros, Collado, Gurvits, and Kelmans [4] showed that many n-vertex graphs, including all trees, cycles, and line graphs, the complements of the above, and P_4-free graphs, belong to $\mathcal{G}(n, 2, \min)$. They did not find any graph with $f_{\min}(G) \geq 3$. Boros, Gurvitch and Meshulam [5] asked whether the value of $f_{\min}$ over all graphs is bounded by a constant. This question was answered in the negative by Balogh and Prince [3], who proved that for every k there is an n_0 such that whenever $n > n_0$, then for a graph G on n vertices we have $f_{\min}(G) \geq k$ with high probability. Their proof used a highly non-trivial Ramsey-type result due to Balogh and Bollobás [2], so their bound on n_0 is a tower function of k.

Our main result is a significant improvement of the Balogh-Prince result. Let $\mathcal{G}(n, n)$ denote the family of 2^{n^2} bipartite graphs G with partite sets V_1 and V_2, $|V_1| = |V_2| = n$.

Theorem 3. *There is a constant $c > 0$ such that for almost all bipartite graphs $G \in \mathcal{G}(n, n)$ one has $f_{\min}(G) \geq cn/(\log n)$.*

Let H be a graph on $\log n$ vertices with $f_{\min}(H) \geq c \log n/(\log \log n)$. One of the basic facts about random graphs is that almost all graphs on n vertices contain H as an induced subgraph. The following theorem is an easy consequence of this fact together with Theorem 3.

Corollary 4. *There is a constant $c > 0$ such that almost all graphs G on n vertices satisfy*

$$f_{\min}(G) \geq \frac{c \log n}{\log \log n}.$$

3 On the number of graphs with k−min-dif representations

3.1 The structure of min-dif representations of bipartite graphs

Analogously to previous notation, $\mathcal{G}(n, k, \min)$ (and $\mathcal{G}(n, n, k, \min)$) denotes the family of (bipartite) graphs G with n labelled vertices V (partite sets V_1 and V_2, $|V_1| = |V_2| = n$, respectively) with $f_{\min}(G) \leq k$. Our aim in this Section is to show that there exists a constant $c > 0$ such that $|\mathcal{G}(n, n, k, \min)| = o(2^{n^2})$ if $k < cn/(\log n)$. This implies that for almost all bipartite graphs on $n + n$ vertices $f_{\min}(G) \geq cn/(\log n)$.

A k-min-difference representation $(A_i : i \in V)$ of G is *reduced* if deleting any element x from all sets that contain it yields a representation of a graph different from G. Note that

$$|A_i \setminus A_j| - 1 \leq |(A_i \setminus x) \setminus (A_j \setminus x)| \leq |A_i \setminus A_j|$$

so the graph G' corresponding to the k-representation $(A_i \setminus x : i \in V)$ has no more edges than G, $E(G') \subseteq E(G)$. There is a natural partition of the elements of $\bigcup A_i$: for every $\emptyset \neq I \subseteq [n]$, we have the subset $(\bigcap_{i \in I} A_i) \cap (\bigcap_{j \notin I} \overline{A_j})$ where $\overline{A_j}$ is the complement of the set A_j. We call these subsets *atoms*. If a k-min-difference representation is reduced, then no atom has more than k elements. It follows that the ground set $\bigcup A_i$ of a reduced representation of an n-vertex graph has no more than $k2^n$ elements. Lemma 5 improves on this observation.

Lemma 5. *Let G be a graph with n vertices and $(A_1, \ldots, A_n)$ a reduced k-min-difference representation of G. Then*

$$\left| \bigcup A_i \right| \leq 2e(G)k \leq kn^2.$$

Proof. Define the sets $A_{i,j} := A_i \setminus A_j$ in the cases $ij \in E(G)$, and $|A_i \setminus A_j| = k$. Let $S := \bigcup A_{i,j}$. The number of elements in S is bounded above by the quantity $|E(G)| \cdot 2k$. We claim that $S = \bigcup A_i$. Otherwise, if there is an element $x \in (\bigcup A_i) \setminus S$, then the representation can be reduced, $(A_i \setminus x : i \in V)$ defines the same graph as $(A_i : i \in V)$. □

The upper bound in Lemma 5 can be significantly improved for bipartite graphs.

Lemma 6. *Let $G \in \mathcal{G}(n, n)$ be a bipartite graph with $n + n$ labeled vertices, $G \in \mathcal{G}(n, n, k, \min)$. Let $(A_1, \ldots, A_n)$ and $(B_1, \ldots, B_n)$ be the sets representing the two parts. If $(A_1, \ldots, A_n, B_1, \ldots, B_n)$ is a reduced k-min-difference representation of G, then*

$$\left| \left(\bigcup A_i \right) \cup \left(\bigcup B_i \right) \right| \leq 4kn.$$

Proof. Suppose that $|A_1| \leq \cdots \leq |A_n|$ and $|B_1| \leq \cdots \leq |B_n|$. Let $A := \bigcup A_i$ and $B := \bigcup B_i$, $S := A \cup B$. Define

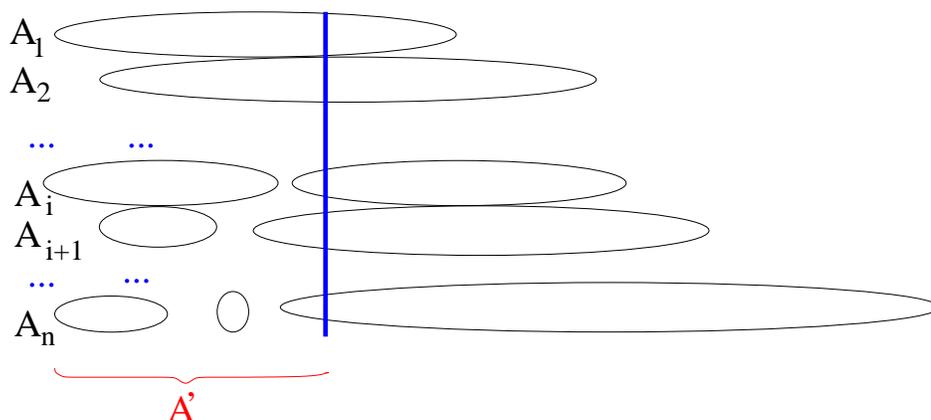$$A' := \bigcup_{i=1}^{n-1} (A_i \setminus A_{i+1}). \tag{1}$$

Figure 1: $|A_1| \leq \cdots \leq |A_n|$ in a min-dif representation when $\{1, 2, \ldots, n\}$ is independent

For each i, the inequality $|A_i \setminus A_{i+1}| \leq |A_{i+1} \setminus A_i|$ follows from the assumption that $|A_i| \leq |A_{i+1}|$. The vertices in each part of G form an independent set, so for each i, we have $|A_i \setminus A_{i+1}| \leq k - 1$. Hence $|A'| \leq (n-1)k$.

If $x \in A_\alpha \setminus A_\beta$ for some $\alpha < \beta$, then there is an index i such that $x \in A_i \setminus A_{i+1}$ and therefore $x \in A'$. In other words, if $x \in A_\alpha \setminus A'$ and $\alpha < \beta$ then $x \in A_\beta$. Therefore the sets $A_i \setminus A'$ form a chain (see Figure 1),

$$A_1 \setminus A' \subseteq A_2 \setminus A' \subseteq \cdots \subseteq A_n \setminus A'.$$

Treat the other part of G analogously: define B' and note the same bound on its size, and note that the sets $B_i \setminus B'$ form a chain.

Let us define $D = S \setminus (A' \cup B')$. We will prove that there are at most $2(n+1)$ sets of the form $A_m \setminus B_\ell$ and $B_p \setminus A_q$, each of cardinality k, covering D. Therefore D contains at most $2(n+1)k$ elements. For each $1 \leq i \leq n$, let us define $\widetilde{A_i} = A_i \cap D$ and $\widetilde{B_i} = B_i \cap D$. Let $\widetilde{A_0} = \widetilde{B_0} = \emptyset$ and $\widetilde{A_{n+1}} = \widetilde{B_{n+1}} = D$. The sets $\widetilde{A_0}, \widetilde{A_1}, \ldots, \widetilde{A_n}, \widetilde{A_{n+1}}$ form a chain, same for $\widetilde{B_0}, \widetilde{B_1} \ldots, \widetilde{B_n}, \widetilde{B_{n+1}}$. The elements of D belong to $(n+1)^2$ atoms (as defined in the beginning of this section), many of them possibly empty, corresponding to the squares in Figure 2.

For each i, j, $1 \leq i, j \leq n+1$, $(i, j) \neq (n+1, n+1)$, the atom $S_{i,j}$ is defined as $(\widetilde{A_i} \setminus \widetilde{A_{i-1}}) \cap (\widetilde{B_j} \setminus \widetilde{B_{j-1}})$. Since the representation is reduced, no elements from the atom $S_{i,j}$ can be left out, so e.g., $S_{1,1} = \emptyset$. It follows that either there are some m and ℓ such that $|A_m \setminus B_\ell| = k$ and the atom $S_{i,j}$ belongs in $A_m \setminus B_\ell$ (here $n \geq m \geq i \geq 1$ and $j > \ell \geq 1$), or there are some p, q such that $|B_p \setminus A_q| = k$ and the atom $S_{i,j}$ is in $B_p \setminus A_q$. Since $\widetilde{|A_m \setminus B_\ell|} \subseteq A_m \setminus B_\ell$, we have $|\widetilde{A_m \setminus B_\ell}| \leq k$ in the first case. Likewise in the second case, $|\widetilde{B_p \setminus A_q}| \leq k$. In Figure 2, the first option corresponds to a rectangle containing the $S_{i,j}$ cell and the upper-right corner, with all the squares in this rectangle together containing only at most k elements. The second
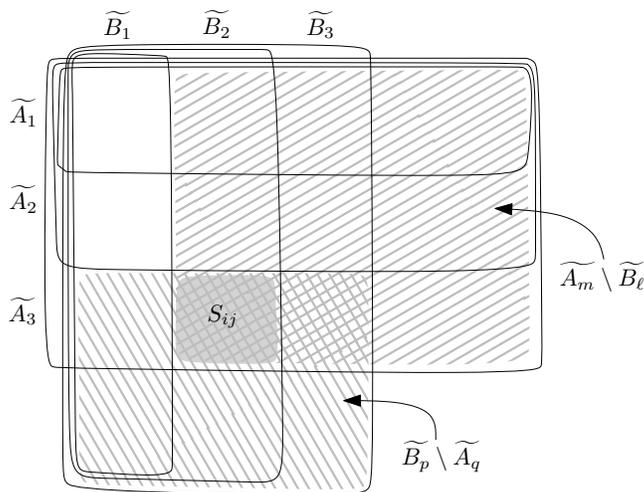
Figure 2: The elements of $S \setminus (A' \cup B')$ split into $(n+1)^2$ atoms.

option corresponds to a similar rectangle with only at most k elements in it, containing the $S_{i,j}$ square and the lower-left corner.

Call a subrectangle $\widetilde{A_m} \setminus \widetilde{B_\ell}$ *critical* if $|A_m \setminus B_\ell| = k$, and similarly $\widetilde{B_p} \setminus \widetilde{A_q}$ is critical if $|B_p \setminus A_q| = k$. Our argument above can be reformulated that every (nonempty) cell $S_{i,j}$ is covered by a critical rectangle. This implies that in each row one can find at most two critical rectangles that cover all non-empty atoms in it. This yields the desired upper bound $|D| \leq 2(n+1)k$.

Finally, altogether $|S| \leq |A'| + |B'| + |D| \leq 4kn$.   □

3.2   Counting reduced matrices

Let S be a set of size $|S| = 4kn$. In this subsection we give an upper bound for the number of sequences $(A_1, \ldots, A_n)$ of subsets of S satisfying the following two properties

(P1) $|A_1| \leq \cdots \leq |A_n|$,
(P2) $|A_i \setminus A_{i+1}| \leq k - 1$ (for all $1 \leq i \leq n - 1$).

Let $\mathcal{M}$ be the 0-1 matrix that has the characteristic vectors of the sets $A_1, \ldots, A_n$ as its rows (in this order). The positions in $\mathcal{M}$ where an entry 1 is directly above an entry 0 will be called *one-zero configurations*, while the positions where a 0 is directly above a 1 will be called *zero-one configurations*. A column in a 0-1 matrix is uniquely determined by the locations of the one-zero configurations and the zero-one configurations unless it is a full 0 or full 1 column. We count the number of possible matrices $\mathcal{M}$ by filling up the $n \times (4kn)$ entries in three steps.

Each one-zero configuration corresponds to an $i < n$ and to an element $x \in A_i \setminus A_{i+1}$. A set $A_{i+1} \setminus A_i$ can be selected in at most

$$\binom{4nk}{0} + \cdots + \binom{4nk}{k-1} < (4en)^k$$

ways ($n > k \geq 1$). Do this for each $i < n$, altogether we have less than $(4en)^{kn}$ ways to write in the one-zero configurations into $\mathcal{M}$.

Select in each column the top 1. If there is no such element in a column we indicate that it is blank, a full zero column. There are at most $n + 1$ outcomes for each column, altogether there are at most $(n + 1)^{4kn}$ possibilities. Fill up with 0's each column above its top 1. Define $A' \subset S$ as in (1), $A' := \bigcup_{i=1}^{n-1}(A_i \setminus A_{i+1})$. We have $|A'| \leq kn$. The columns of $\mathcal{M}$ that correspond to the elements of $S \setminus A'$ have a (possibly empty) string of zeros followed by a string of ones. We almost filled up $\mathcal{M}$ and we can finish this process by selecting the remaining zero-one configurations.

There may be several zero-one configurations in a single column. Each of them has a unique (closest, or smallest indexed) 1 above them. That element 1 is already written in into our still partially filled $\mathcal{M}$, because that element 1 (even if it is the top 1 element) belongs to a unique one-zero configuration. This correspondence is an injection. So there are at most $\sum_i |A_i \setminus A_{i+1}| \leq kn$ zero-one configurations in the columns corresponding to A' which are not yet identified. There are at most n^{kn} ways to select them.

Since (for $n > k \geq 1$)

$$(4en)^{kn} \times (n + 1)^{4kn} \times n^{kn} < n^{6kn + O(kn/\log n)} = e^{6kn \log n + O(kn)},$$

we obtain the following

Claim 7. *Altogether, there are $e^{O(kn \log n)}$ ways to fill $\mathcal{M}$ with entries in $\{0, 1\}$ according to the rules* (P1) *and* (P2).

3.3 Proofs of the lower bounds

Proof of the lower bound in Theorem 3. Let $G = G(V_1, V_2)$ be a bipartite graph with both parts of size n and suppose that G belongs to $\mathcal{G}(n, n, k, \min)$. By Lemma 6 we may suppose that G has a reduced k-min-difference representation $(A_1, \ldots, A_n, B_1, \ldots, B_n)$ such that each representing set is a subset of S, where $|S| = 4kn$. There are permutations π and $\rho \in S_n$ which rearrange the sets according their sizes $|A_{\pi(1)}| \leq \cdots \leq |A_{\pi(n)}|$ and $|B_{\rho(1)}| \leq \cdots \leq |B_{\rho(n)}|$. Consider the $V_i \times S$ matrices, $\mathcal{M}_i$, $i = 1, 2$ whose i'th row is the 0-1 characteristic vector of $A_{\pi(i)}$ and $B_{\rho(i)}$, respectively. The permutations π, ρ and the matrices $\mathcal{M}_1$, $\mathcal{M}_2$ completely describe G. The matrices $\mathcal{M}_i$ satisfies properties (P1) and (P2), so Claim 7 yields the following upper bound for the number of such fourtuples $(\pi, \rho, \mathcal{M}_1, \mathcal{M}_2)$

$$|\mathcal{G}(n, n, k, \min)| \leq \# \text{ of } (\pi, \rho, \mathcal{M}_1, \mathcal{M}_2)\text{'s} \leq (n!)^2 n^{(12 + o(1))kn} = e^{O(kn \log n)}. \tag{2}$$

Here the right hand side is $o(2^{n^2})$ if $k \leq 0.057n/(\log n)$ implying that $f_{\min}(G) > 0.057n/(\log n)$ for almost all the 2^{n^2} bipartite graphs.   □

Proof of the lower bound for the random bipartite graph.
Recall that in a random graph $G \in \mathcal{G}(n,p)$, each of the $\binom{n}{2}$ edges occurs independently with probability p. Similarly, $\mathcal{G}(n,n,p)$ denotes the class of graphs $\mathcal{G}(n,n)$ with the probability of a given graph $G \in \mathcal{G}(n,n)$ is

$$p^{e(G)}(1-p)^{n^2-e(G)}.$$

Here the right hand side is at most $(\max\{p, 1-p\})^{n^2}$. This implies that for any class of graphs $\mathcal{A} \subset \mathcal{G}(n,n)$ the probability $\Pr(G \in \mathcal{A})$ is at most $|\mathcal{A}|$ times this upper bound. If the class of graphs $\mathcal{A}$ is too small, namely

$$|\mathcal{A}| = o\left((\min\{\frac{1}{p}, \frac{1}{1-p}\})^{n^2}\right),$$

then for $\mathcal{G}(n,n,p)$ one has

$$\Pr(G \in \mathcal{A}) \to 0. \tag{3}$$

Taking $\mathcal{A} := \mathcal{G}(n,n,k,\min)$ with a sufficiently small k, we obtain

Corollary 8. *For constant $0 < p < 1$ there exists a constant $c = c_{1,\min}(p) > 0$ such that the following holds for $G \in \mathcal{G}(n,n,p)$ with high probability as $n \to \infty$*

$$c\frac{n}{\log n} < f_{\min}(G).$$

4   Maximum and average difference representations

Boros, Gurvich and Meshulam [5] also defined *k-max-difference* representations and *k-average-difference* representations of a graph G in a natural way, that is, the vertices i and j are adjacent if and only if for the corresponding sets A_i, A_j we have $\max\{|A_i \setminus A_j|, |A_j \setminus A_i|\} \geq k$ and $(|A_i \setminus A_j| + |A_j \setminus A_i|)/2 \geq k$, respectively. Analogously to $f_{\min}$ we can define $f_{\max}(G)$ and $f_{\text{avg}}(G)$. Since for every graph G a Kneser representation is a min-dif, average-difference, and max-difference representation as well we get

$$f_{\min}(G), f_{\text{avg}}(G), f_{\max}(G) \leq f_{\text{Kneser}}(G) \leq n-1. \tag{4}$$

Let $\mathcal{G}(n,k,\max)$, $\mathcal{G}(n,n,k,\max)$, $(\mathcal{G}(n,k,\text{avg})$, $\mathcal{G}(n,n,k,\text{avg}))$ denote the family of graphs $G \in \mathcal{G}(n)$ and in $\mathcal{G}(n,n)$ with n labeled vertices V or with partite sets V_1 and V_2, $|V_1| = |V_2| = n$, respectively, such that $f_{\max}(G) \leq k$ ($f_{\text{avg}}(G) \leq k$, respectively).

It was proved in [5] that $f_{\max}$ and $f_{\min}$ are not bounded by a constant, for a matching of size t one has $f_{\max}(tK_2) = \Theta(\log t)$ and $f_{\text{avg}}(tK_2) = \Theta(\log t)$. (It turns out that $f_{\min}(tK_2) = 1$.) The proof of Theorem 3 can be easily adapted for these parameters for $\mathcal{G}(n,n,p)$ as well. Even more, we can handle the general case $G \in \mathcal{G}(n,p)$, too.

Corollary 9. *For constant $0 < p < 1$ there exists a constant $c = c(p) > 0$ such that the following holds for $G \in \mathcal{G}(n, n, p)$ with high probability as $n \to \infty$*

$$c\frac{n}{\log n} < f_{\mathrm{avg}}(G),\ f_{\mathrm{max}}(G). \tag{5}$$

Similarly for $G \in \mathcal{G}(n, p)$ with high probability we have

$$c\frac{n}{\log n} < f_{\mathrm{avg}}(G),\ f_{\mathrm{max}}(G). \tag{6}$$

These lower bounds together with the upper bounds from Corollary 14 below imply that for almost all n vertex graphs, and for almost all bipartite graphs on $n + n$ vertices, $f_{\mathrm{avg}}(G)$ and $f_{\mathrm{max}}(G)$ are $\Theta(n/(\log n))$.

Sketch of the proof. If $G \in \mathcal{G}(n, n, k, \mathrm{max})$ (and if $G \in \mathcal{G}(n, n, k, \mathrm{avg})$) and $(A_1, \ldots, A_n, B_1, \ldots, B_n)$ is a k-max-difference (k-average-difference) representation then

$$|A_i \triangle A_j| \le 2k - 2 \ (\le 2k - 1) \tag{7}$$

holds for each pair i, j. If the representation is reduced, then we obtain (without the tricky proof of Lemma 6) that $|\bigcup_i A_i| < 2kn$, and the same holds for $|\bigcup_i B_i|$, too. The conditions of Claim 7 are satisfied implying

$$|\mathcal{G}(n, n, k, \mathrm{max})|, |\mathcal{G}(n, n, k, \mathrm{avg})| = e^{O(kn \log n)}.$$

We complete the proof of (5) applying (3) as it was done at the end of the previous Section.

Consider a graph $G \in \mathcal{G}(n, k, \mathrm{max})$ and let $(A_1, \ldots, A_n)$ be a reduced k-max-difference representation. (The case of k-average-difference representation can be handled in the same way, and the details are left to the reader). The only additional observation we need is that since (7) holds for each non-edge $\{i, j\}$, we have $|A_i \triangle A_j| \le 4k - 4$ for *all* pairs of vertices whenever $\mathrm{diam}(\overline{G}) \le 2$. Thus for every reduced representation (in case of $\mathrm{diam}(\overline{G}) \le 2$) one has $|\bigcup_i A_i| \le (4k - 4)n$. Also, $|A_i \setminus A_j| \le 2k - 2$ for $|A_i| \le |A_j|$. Then the conditions of Claim 7 are fulfilled (with $2k - 2$ instead of k) implying the following version of (2)

$$|\mathcal{G}(n, n, k, \mathrm{min}) \setminus \mathcal{G}_2(n)| = e^{O(kn \log n)},$$

where $\mathcal{G}_2(n)$ denotes the class of graphs with $G \in \mathcal{G}(n), \mathrm{diam}(\overline{G}) > 2$.

We complete the proof of (6) by applying (3) and the fact that

$$\mathrm{diam}(G) \ge 2$$

holds with high probability for $G \in \mathcal{G}(n, p)$. $\square$

5   Clique covers of the edge sets of graphs

We need the following version of Chernoff's inequality (see, e.g., [1]). Let $Y_1, \ldots, Y_n$ be mutually independent random variables with $E[Y_i] \le 0$ and all $|Y_i| \le 1$. Let $a \ge 0$. Then

$$\Pr[Y_1 + \cdots + Y_n > a] < e^{-a^2/(2n)}. \tag{8}$$

A finite *linear space* is a pair $(P, \mathcal{L})$ consisting of a set P of elements (called *points*) and a set $\mathcal{L}$ of subsets of P (called *lines*) satisfying the following two properties.

(L1)   Any two distinct points $x, y \in P$ belong to exactly one line $L = L(x, y) \in \mathcal{L}$.

(L2)   Any line has at least two points.

In other words, the edge set of the complete graph $K(P)$ has a clique decomposition into the complete graphs $K(L)$, $L \in \mathcal{L}$.

Lemma 10. *For every positive integer n there exists a linear space $\mathcal{L} = \mathcal{L}_n$ with lines $L_1, \ldots, L_m$ such that $m = n + o(n)$, every edge has size $(1 + o(1))\sqrt{n}$, and every point belongs to $(1 + o(1))\sqrt{n}$ lines.*

Proof. (Folklore). If $n = q^2$ where $q > 1$ is a power of a prime then we can take the $q^2 + q$ lines of an affine geometry $AG(2, q)$. Each line has exactly $q = \sqrt{n}$ points and each point belongs to $q + 1$ lines. In general, one can consider the smallest power of prime q with $n \le q^2$ (we have $q = (1 + o(1))\sqrt{n}$) and take a random n-set $P \subset \mathbb{F}_q^2$ and the lines defined as $P \cap L$, $L \in \mathcal{L}(AG(2, q))$. □

5.1   Thickness of clique covers

The *clique cover number* $\theta_1(G)$ of a graph G is the minimum number of cliques required to cover the edges of graph G. Frieze and Reed [9] proved that for p constant, $0 < p < 1$, there exist constants $c_i' = c_i'(p) > 0$, $i = 1, 2$ such that for $G \in \mathcal{G}(n, p)$ with high probability

$$c_1' \frac{n^2}{(\log n)^2} < \theta_1(G) < c_2' \frac{n^2}{(\log n)^2}.$$

They note that 'a simple use of a martingale tail inequality shows that θ_1 is close to its mean with very high probability'. We only need the following consequence concerning the expected value.

$$E(\theta_1(G)) < c_3' \frac{n^2}{(\log n)^2}. \tag{9}$$

The *thickness* θ_0 of a clique cover $\mathcal{C} := \{C_1, \ldots, C_m\}$ of G is the maximum degree of the hypergraph $\mathcal{C}$, i.e., $\theta_0(\mathcal{C}) := \max_{v \in V(G)} \deg_{\mathcal{C}}(v)$. The minimum thickness among the clique covers of G is denoted by $\theta_0(G)$.

A clique cover $\mathcal{C}$ corresponds to a set representation $v \mapsto A_v$ in a natural way $A_v := \{C_i : v \in C_i\}$ with the property that A_u and A_v are disjoint if and only if $\{u, v\}$ is a non-edge of G.

The size of the largest A_v is the thickness of $\mathcal{C}$. For $k > \theta_0(\mathcal{C})$ one can add $k - |A_v|$ distinct extra elements to A_v (for each $v \in V(G)$), thus obtaining a Kneser representation of rank k of the complement of G, $\overline{G}$. This relation can be reversed, yielding

$$\theta_0(G) \leq f_{\text{Kneser}}(\overline{G}) \leq \theta_0(G) + 1. \tag{10}$$

Theorem 11. *For constant $0 < p < 1$ there exist constants $c_i = c_i(p) > 0$, $i = 1, 2$ such that for $G \in \mathcal{G}(n, p)$ with high probability*

$$c_1 \frac{n}{\log n} < \theta_0(G) < c_2 \frac{n}{\log n}.$$

Proof. The lower bound is easy. The maximum degree $\Delta(G)$ of $G \in \mathcal{G}(n, p)$ with high probability satisfies $\Delta \approx np$. As usual we write $a_n \approx b_n$ as n tends to infinity. Also for the unproved but well-known statements concerning the random graphs see the monograph [12]. The size of the largest clique $\omega = \omega(G)$ with high probability satisfies $\omega \approx 2 \log n / (\log(1/p))$. Since $\theta_0 \geq \Delta / (\omega - 1)$ we may choose $c_1 \approx p(\log(1/p))/2$.

The upper bound probably can be proved by analyzing and redoing the clever proof of Frieze and Reed concerning $\theta_1(G_{n,p})$. Probably their randomized algorithm yields the upper bound for the thickness, too. Although there are steps in their proof where they remove from G (as cliques of size 2) an edge set of size $O(n^{31/16})$ and one needs to show that these edges are well-distributed. However, one can easily deduce the upper bound for $\theta_0(G_{n,p})$ directly only from Equation (9).

Given n, fix a linear hypergraph $\mathcal{L} = \mathcal{L}_n$ with point set $[n]$ and hyperedges $L_1, \ldots, L_m$ provided by Lemma 10. We have $m = n + o(n)$, $\ell_i := |L_i| = (1 + o(1))\sqrt{n}$, and every point v belongs to $b_v = (1 + o(1))\sqrt{n}$ lines. Build the random graph $G \in \mathcal{G}(n, p)$ in m steps by taking a $G_i \in \mathcal{G}(L_i, p)$. Let $\mathcal{C}_i$ be a clique cover of G_i with $\theta_1(G_i)$ members, $\mathcal{C} = \cup_{1 \leq i \leq m} \mathcal{C}_i$.

Let $X_i(v)$ denote the thickness of $\mathcal{C}_i$ at the point $v \in L_i$. We consider $X_i(v)$ as a random variable, whose distribution is depending only on ℓ_i and p. For every L_i, we have

$$\sum_{v; v \in L_i} X_i(v) = \sum_{C \in \mathcal{C}_i} |C| \leq \theta_1(G_i)\omega(G_i).$$

Here with *very* high probability $\omega_i = \omega(G_i)$ satisfies $\omega_i \approx 2 \log \ell_i / (\log(1/p))$. Then (9) implies that

$$E(\sum_{v \in L_i} X_i(v)) \leq c_3' \frac{\ell_i^2}{(\log \ell_i)^2} \times (1 + o(1)) \frac{2 \log \ell_i}{\log(1/p)}.$$

Since the distributions of $X_i(u)$ and $X_i(v)$ are identical (for $u, v \in L_i$), and there are ℓ_i terms on the left hand side, we obtain that

$$E(X_i(v)) \leq (1 + o(1)) \frac{2c_3'}{\log(1/p)} \times \frac{\ell_i}{\log \ell_i} < c \frac{\sqrt{n}}{\log n}.$$

Here we chose $c > 4c_3'/(\log(1/p))$.

Let $X(v)$ be the thickness of $\mathcal{C}$ at v. We have $X(v) = \sum_{L_i \ni v} X_i(v)$, where this is a sum of $b := b_v = (1 + o(1))\sqrt{n}$ mutually independent random variables, and each term is non-negative and is bounded by $\ell = \max_i \ell_i = (1 + o(1))\sqrt{n}$. Define b independent random variables

$$Y_i := \frac{1}{\ell}\left(X_i(v) - c\frac{\sqrt{n}}{\log n}\right)$$

for each i with $L_i \ni v$. We can apply Chernoff's inequality (8) for any real $a > 0$

$$\Pr\left[\left(\sum_{L_i \ni v} Y_i\right) > a\right] < e^{-a^2/(2b)}.$$

Substituting $a := \sqrt{4b \log n}$ the right hand side is $1/n^2$ and we get

$$\Pr\left[X(v) > c\frac{b\sqrt{n}}{\log n} + \ell\sqrt{4b \log n}\right] < \frac{1}{n^2}.$$

Since this is true for all $v \in [n]$, we obtain that (for large enough n) for any $c_2 > c$

$$\Pr\left[X(v) < c_2\frac{n}{\log n} \text{ for all } v\right] > 1 - \frac{1}{n},$$

completing the proof of the upper bound for $\theta_0(G)$. $\qquad\square$

Since the complement of a random graph $G \in \mathcal{G}(n, p)$ is a random graph from $\mathcal{G}(n, 1-p)$, Theorem 11 and (10) imply that

Corollary 12. *For constant $0 < p < 1$, there exist constants $c_{i,\text{Kneser}} = c_{i,\text{Kneser}}(p) > 0$, $i = 1, 2$ such that for $G \in \mathcal{G}(n, p)$ with high probability*

$$c_{1,\text{Kneser}}\frac{n}{\log n} < f_{\text{Kneser}}(G) < c_{2,\text{Kneser}}\frac{n}{\log n}.$$

$\qquad\square$

One can also prove a similar upper bound for the random bipartite graph.

Corollary 13. *For constant $0 < p < 1$ there exist a constant $c_{3,\text{Kneser}} = c_{3,\text{Kneser}}(p) > 0$ such that for $G \in \mathcal{G}(n, n, p)$ with high probability*

$$f_{\text{Kneser}}(G) < c_{3,\text{Kneser}}\frac{n}{\log n}.$$

Proof. Let $G \in \mathcal{G}(n, n, p)$ a random bipartite graph with partite sets $|A| = |B| = n$. Consider a random graph $G_A \in \mathcal{G}(A, p)$ and $G_B \in \mathcal{G}(B, p)$, their union is $H = G \cup G_A \cup G_B$. We can

consider $\overline{H}$ as a member of $\mathcal{G}(A \cup B, 1-p)$. Since $\overline{G}$ can be obtained from $\overline{H}$ by adding two complete graphs K_A and K_B, we obtain

$$f_{\text{Kneser}}(G) - 1 \leq \theta_0(\overline{G}) \leq \theta_0(\overline{H}) + 1 < 1 + c_2(1-p)\frac{2n}{\log(2n)},$$

where the last inequality holds with high probability according to Theorem 11.   □

Recall (4), that for every graph G, $f_{\min}(G)$, $f_{\text{avg}}(G)$, $f_{\max}(G) \leq f_{\text{Kneser}}(G)$ holds. These and the above two Corollaries imply the following upper bounds.

Corollary 14. *For constant $0 < p < 1$ the following holds for $G \in \mathcal{G}(n,p)$ with high probability as $n \to \infty$*

$$f_{\min}(G),\ f_{\text{avg}}(G),\ f_{\max}(G) < c_{2,\text{Kneser}}(p)\frac{n}{\log n},$$

and similarly for $G \in \mathcal{G}(n,n,p)$

$$f_{\min}(G),\ f_{\text{avg}}(G),\ f_{\max}(G) < c_{3,\text{Kneser}}(p)\frac{n}{\log n}.$$

6   Prague dimension

The *Prague dimension* (it is also called *product dimension*) $f_{\text{Pra}}(G)$ of a graph G is the smallest integer k such that one can find vertex distinguishing good colorings $\varphi_1, \ldots, \varphi_k :$ $V(G) \to \mathbb{N}$. This means that $\varphi_i(u) \neq \varphi_i(v)$ for every edge $uv \in E(G)$ and $1 \leq i \leq k$ but for every non-edge $\{u,v\}$, there exists an i with $\varphi_i(u) = \varphi_i(v)$, moreover the vectors $(\varphi_1(u), \varphi_2(u), \ldots, \varphi_k(u))$ and $(\varphi_1(v), \varphi_2(v), \ldots, \varphi_k(v))$ are distict for $u \neq v$. Two vertices are adjacent if and only if their labels disagree in every φ_i. As Hamburger, Por, and Walsh [11] observed, the Kneser rank never exceeds the Prague dimension, so one can extend (4) as follows. For every graph G

$$f_{\min}(G),\ f_{\text{avg}}(G),\ f_{\max}(G) \leq f_{\text{Kneser}}(G) \leq f_{\text{Pra}}(G). \tag{11}$$

The determination of $f_{\text{Pra}}(G)$ is usually a notoriously difficult task. The results of Lovász, Nešetřil, and Pultr [14] were among the first (non-trivial) applications of the algebraic method. Hamburger, Por, and Walsh [11] observed that there are graphs where the difference of $f_{\text{Pra}}(G) - f_{\text{Kneser}}(G)$ is arbitrarily large, even for Kneser graphs $\text{Kn}(s,k)$. Poljak, Pultr, and Rödl [16] proved that $f_{\text{Pra}}(\text{Kn}(s,k)) = \Theta(\log \log s)$ (as k is fixed and $s \to \infty$) while $f_{\text{Kneser}}(\text{Kn}(s,k)) = k$ for all $s \geq 2k > 0$. Still we think that for most graphs these parameters have the same order of magnitude.

Conjecture 15. *For a constant probability $0 < p < 1$ there exists a constant $c_{2,\text{Pra}} = c_{2,\text{Pra}}(p) > 0$, such that for $G \in \mathcal{G}(n,p)$ with high probability*

$$f_{\text{Pra}}(G) < c_{2,\text{Pra}}\frac{n}{\log n}.$$

A matching lower bound $c_{1,\text{Kneser}}(n/(\log n)) < f_{\text{Pra}}(G)$ (with high probability) follows from (11) and Corollary 12. We think the same order of magnitude holds for the case when G is bipartite.

Conjecture 16. *For a constant probability $0 < p < 1$ there exists a constant $c_{3,\text{Pra}} = c_{3,\text{Pra}}(p) > 0$, such that for $G \in \mathcal{G}(n,n,p)$ with high probability*

$$f_{\text{Pra}}(G) < c_{3,\text{Pra}}\frac{n}{\log n}.$$

6.1   Prague dimension and clique coverings of graphs

The *chromatic index* $\theta'_0(\mathcal{C})$ of a clique cover $\mathcal{C} := \{C_1, \ldots, C_m\}$ of the graph G is the chromatic index of the hypergraph $\mathcal{C}$, i.e., $\theta'_0(\mathcal{C})$ is the smallest k that one can decompose the clique cover into k parts, $\mathcal{C} = \mathcal{C}_1 \cup \cdots \cup \mathcal{C}_k$ such that the members of each $\mathcal{C}_i$ are pairwise (vertex)disjoint. The minimum chromatic index among the clique covers of G is denoted by $\theta'_0(G)$. In other words, $E(G)$ can be covered by k subgraphs with complete graph components. Obviously, the thickness is a lower bound $\theta_0(G) \leq \theta'_0(G)$. Here the left hand side is at most $O(n/(\log n))$ for almost all graphs by Theorem 11. We think that the Frieze–Reed [9] method can be applied to find the correct order of magnitude of θ'_0, too.

Conjecture 17. *For p constant, $0 < p < 1$, there exists a constants $c_4 = c_4(p) > 0$ such that for $G \in \mathcal{G}(n,p)$ with high probability*

$$\theta'_0(G) < c_4\frac{n}{\log n}.$$

One can observe that (similarly as f_{Kneser} and θ_0 are related, see (10)) there is a remarkable simple connection between Prague dimension and θ'_0.

$$\theta'_0(G) \leq f_{\text{Pra}}(\overline{G}) \leq \theta'_0(G) + 1.$$

So Conjectures 15 and 17 are in fact equivalent, and Conjecture 17 also implies Conjecture 16.

7   Conclusion

We have considered five graph functions $f_{\min}(G)$, $f_{\text{avg}}(G)$, $f_{\max}(G)$, $f_{\text{Kneser}}(G)$, and $f_{\text{Pra}}(G)$, which are *hereditary* (monotone for induced subgraphs) and two random graph models $\mathcal{G}(n,p)$ and $\mathcal{G}(n,n,p)$. We gave an upper bound for the order of magnitude for eight of the possible ten problems, and we have also have conjectures for the missing two upper bounds (Conjectures 15 and 16). We also established matching lower bounds in seven cases, which also gave probably the best lower bound in two more cases (concerning f_{Pra}). All of these 19 estimates were $\Theta(n/(\log n))$. In the last case (in Corollary 4) we have a weaker bound, so it is natural to ask that

Problem 18. *Is it true that for any fixed $0 < p < 1$ for $G \in \mathcal{G}(n,p)$ with high probability one has $\Omega(n/(\log n)) \leq f_{\min}(G)$?*

Let us remark that if G is a complement of a triangle-free graph then the Kneser rank and Prague dimension is $\Delta(\overline{G})$ or $\Delta(\overline{G}) + 1$. So it can be $\Omega(n)$. For example, $f_{\text{Kneser}}(\overline{K_{1,n-1}}) = n - 1$. No such results are known for $f_{\min}$.

Problem 19. *What is the maximum of $f_{\min}(G)$ over the set of n-vertex graphs G? Is it true that $f_\rho(G) = o(n)$ for every $\rho \in \{\min, \text{avg}, \max\}$ and $G \in \mathcal{G}(n) \cup \mathcal{G}(n,n)$?*

References

[1] N. Alon and J. H. Spencer, *The probabilistic method*, Wiley Series in Discrete Mathematics and Optimization, John Wiley & Sons, Inc., Hoboken, NJ, fourth ed., 2016.

[2] J. Balogh and B. Bollobás, *Unavoidable traces of set systems*, Combinatorica, 25 (2005), pp. 633–643.

[3] J. Balogh and N. Prince, *Minimum difference representations of graphs*, Graphs Combin., 25 (2009), pp. 647–655.

[4] E. Boros, R. Collado, V. Gurvich, and A. Kelmans, *On 2-representable graphs*, manuscript, DIMACS REU Report, July, (2000).

[5] E. Boros, V. Gurvich, and R. Meshulam, *Difference graphs*, Discrete Math., 276 (2004), pp. 59–64. 6th International Conference on Graph Theory.

[6] M. S. Chung and D. B. West, *The p-intersection number of a complete bipartite graph and orthogonal double coverings of a clique*, Combinatorica, 14 (1994), pp. 453–461.

[7] N. Eaton and V. Rödl, *Graphs of small dimensions*, Combinatorica, 16 (1996), pp. 59–85.

[8] P. Erdős, A. W. Goodman, and L. Pósa, *The representation of a graph by set intersections*, Canad. J. Math., 18 (1966), pp. 106–112.

[9] A. Frieze and B. Reed, *Covering the edges of a random graph by cliques*, Combinatorica, 15 (1995), pp. 489–497.

[10] Z. Füredi, *Competition graphs and clique dimensions*, Random Structures Algorithms, 1 (1990), pp. 183–189.

[11] P. Hamburger, A. Por, and M. Walsh, *Kneser representations of graphs*, SIAM J. Discrete Math., 23 (2009), pp. 1071–1081.

[12] S. Janson, T. Łuczak, and A. Rucinski, *Random graphs*, Wiley-Interscience Series in Discrete Mathematics and Optimization, Wiley-Interscience, New York, 2000.

[13] J. Körner and A. Monti, *Compact representations of the intersection structure of families of finite sets*, SIAM J. Discrete Math., 14 (2001), pp. 181–192.

[14] L. Lovász, J. Nešetřil, and A. Pultr, *On a product dimension of graphs*, J. Combin. Theory Ser. B, 29 (1980), pp. 47–67.

[15] T. A. McKee and F. R. McMorris, *Topics in intersection graph theory*, SIAM Monographs on Discrete Mathematics and Applications, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1999.

[16] S. Poljak, A. Pultr, and V. Rödl, *On the dimension of Kneser graphs*, in Algebraic methods in graph theory, Vol. I, II (Szeged, 1978), vol. 25 of Colloq. Math. Soc. János Bolyai, North-Holland, Amsterdam-New York, 1981, pp. 631–646.