Technical Report

TR-2003-011

Distance Graphs on the Integers

by

M. Ferrara, Y. Kohayakawa, V. Rodl

Mathematics and Computer Science

EMORY UNIVERSITY

DISTANCE GRAPHS ON THE INTEGERS

M. FERRARA, Y. KOHAYAKAWA, AND V. RÖDL

Dedicated to the memory of Professor Walter Deuber on the occasion of his 60th birthday.

Abstract. We consider several extremal problems concerning representations of graphs as *distance graphs* on the integers. Given a graph $G = (V, E)$, we wish to find an injective function $\varphi\colon V \to \mathbb{Z}^+ = \{1, 2, \dots\}$ and a set $\mathcal{D} \subset \mathbb{Z}^+$ such that $\{u, v\} \in E$ if and only if $|\varphi(u) - \varphi(v)| \in \mathcal{D}$.

Let $s(n)$ be the smallest N such that any graph G on n vertices admits a representation $(\varphi_G, \mathcal{D}_G)$ such that $\varphi_G(v) \le N$ for all $v \in V(G)$. We show that $s(n) = (1 + o(1))n^2$ as $n \to \infty$. In fact, if we let $s_r(n)$ be the smallest N such that any r-regular graph G on n vertices admits a representation $(\varphi_G, \mathcal{D}_G)$ such that $\varphi_G(v) \le N$ for all $v \in V(G)$, then $s_r(n) = (1 + o(1))n^2$ as $n \to \infty$ for any $r = r(n) \gg \log n$ with rn even for all n.

Given a graph $G = (V, E)$, let $D_{\mathrm{e}}(G)$ be the smallest possible cardinality of a set $\mathcal{D}$ for which there is some $\varphi\colon V \to \mathbb{Z}^+$ so that $(\varphi, \mathcal{D})$ represents G. We show that, for almost all n-vertex graphs G, we have

$$D_{\mathrm{e}}(G) \ge \frac{1}{2}\binom{n}{2} - (1 + o(1))n^{3/2}(\log n)^{1/2},$$

whereas for some n-vertex graph G, we have

$$D_{\mathrm{e}}(G) \ge \binom{n}{2} - n^{3/2}(\log n)^{1/2+o(1)}.$$

Further extremal problems of similar nature are considered.

1. Introduction and statement of the main results

We are interested in representing a graph $G = (V, E)$ by assigning to each vertex $v \in V$ an integer $\varphi(v)$ so that we can then distinguish those pairs of vertices u, v that are edges from those that are not simply by $|\varphi(u) - \varphi(v)|$. In other words, for each graph $G = (V, E)$ we wish to find a

Date: 21 October 2003. Under revision.

1991 *Mathematics Subject Classification.* Primary: 05C62. Secondary: 11B75.

Key words and phrases. Sidon sets, geometric representations of graphs.

The second author was on leave from Instituto de Matemática e Estatística, Universidade de São Paulo, Rua do Matão 1010, 05508–090 São Paulo, Brazil, and was partially supported by MCT/CNPq (ProNEx project Proc. CNPq 664107/1997–4) and by CNPq (Proc. 300334/93–1 and 468516/2000–0). The third author was partially supported by NSF Grant 0071261. The collaboration of the Brazilian author and the US authors is supported by a CNPq/NSF cooperative grant (910064/99–7, 0072064).

function $\varphi\colon V \to \mathbb{Z}^+ = \{1, 2, \dots\}$ and a set $\mathcal{D} \subset \mathbb{Z}^+$ so that $\{u, v\} \in E$ if and only if $|\varphi(u) - \varphi(v)| \in \mathcal{D}$.

We shall investigate numerical parameters associated with such representations $(\varphi, \mathcal{D})$ of G. A basic question is to determine or estimate the smallest possible value for $\max\{\varphi(i)\colon 1 \le i \le n\}$ for a given graph G, when we let φ and $\mathcal{D}$ vary freely. We denote this value by $s(G)$ (see (5)). Let $s(n) = \max_G s(G)$, where the maximum is taken over all n-vertex graphs. One of our results shows that $s(n) = (1 + o(1))n^2$ (see Theorem 4).

We shall also investigate the size of the sets $\mathcal{D}$ that we need, both for individual graphs and when considering all graphs on a fixed number of vertices. Other parameters pertaining to the function φ, which may be of some interest without any graph-theoretical considerations, are also considered (see Lemmas 11 and 12).

1.1. **Definitions.** We shall consider integer functions

$$\varphi\colon [n] \to \mathbb{Z}^+ \tag{1}$$

with domain $[n] = \{1, \dots, n\}$. In what follows, the asymptotics will be with respect to $n \to \infty$. A map φ as in (1) defines an equivalence relation

$$\sim_\varphi \tag{2}$$

on $\binom{[n]}{2}$, the edge set of the complete graph K_n with vertex set $[n]$, given by

$$\{i, j\} \sim_\varphi \{k, \ell\} \iff |\varphi(i) - \varphi(j)| = |\varphi(k) - \varphi(\ell)|. \tag{3}$$

Let $\mathrm{Dist}(\varphi)$ be the set of equivalence classes of $\sim_\varphi$. Note that

(*) $|\mathrm{Dist}(\varphi)|$ *is the number of distinct 'distances', i.e., integers of the form*

$$|\varphi(i) - \varphi(j)| \qquad (i, j \in [n],\ i \ne j), \tag{4}$$

that φ induces.

In what follows, we shall be interested in the numerical parameters $|\mathrm{Dist}(\varphi)|$ and $\max\{\varphi(i)\colon i \in [n]\}$ of the functions $\varphi\colon [n] \to \mathbb{Z}^+$.

We now introduce a central definition.

Definition 1 ($G(\varphi, \mathcal{D})$). *Given $\varphi\colon [n] \to \mathbb{Z}^+$ and $\mathcal{D} \subset \mathbb{Z}^+ = \{1, 2, \dots\}$, we define the graph $G(\varphi, \mathcal{D})$ as the graph whose vertex set is $[n]$, with i and j adjacent if and only if $|\varphi(i) - \varphi(j)| \in \mathcal{D}$.*

If G is isomorphic to $G(\varphi, \mathcal{D})$, then we shall say that $(\varphi, \mathcal{D})$ *spaces* G (or that $(\varphi, \mathcal{D})$ is a *spacing* of G).

Clearly, in the definition of $G(\varphi, \mathcal{D})$ above, we make i adjacent to j if and only if the distance between $\varphi(i)$ and $\varphi(j)$ is an element of a set $\mathcal{D}$ of 'distinguished' distances. Furthermore, it is clear that the edge set of $G(\varphi, \mathcal{D})$ is a union of equivalence classes of the relation $\sim_\varphi$; equivalently, if $\{i, j\} \sim_\varphi \{k, \ell\}$, then $\{i, j\}$ is an edge of $G(\varphi, \mathcal{D})$ if and only if $\{k, \ell\}$ is.

We are interested in representing graphs G as graphs of the form $G(\varphi, \mathcal{D})$. In particular, when representing an n-vertex graph G as some $G(\varphi, \mathcal{D})$, numerical constraints on $\mathcal{D}$ and $\mathrm{im}(\varphi) = \{\varphi(i)\colon i \in [n]\}$ arise. For instance, we shall show that for almost all G, we must allow $|\mathcal{D}|$ to be of order n^2. To prove such constraints, we shall give upper estimates for the number of graphs that are of the form $G(\varphi,\mathcal{D})$, up to isomorphism.

Let us now define some extremal parameters for n-vertex graphs G. In what follows, φ and $\mathcal{D}$ will always stand for a function $\varphi\colon [n]\to\mathbb{Z}^+$ and a set $\mathcal{D}\subset\mathbb{Z}^+$.

Given φ, we shall say that a graph G is a *φ-graph* if G is isomorphic to $G(\varphi,\mathcal{D})$ for some $\mathcal{D}$. Similarly, given $\mathcal{D}$, we shall say that G is a *$\mathcal{D}$-graph* if G is isomorphic to $G(\varphi,\mathcal{D})$ for some φ.

Our first extremal parameter is $s(G)$, the minimal integer N for which there is a function φ with $\mathrm{im}(\varphi)\subset[N]$ such that G is a φ-graph. Formally,

$$\begin{aligned} s(G) &= \min\{N\colon \text{there is } \varphi\colon [n]\to\mathbb{Z}^+ \\ &\qquad\qquad \text{such that } \mathrm{im}(\varphi)\subset[N] \text{ and } G \text{ is a } \varphi\text{-graph}\} \\ &= \min\{\max\mathrm{im}(\varphi)\colon G \text{ is a } \varphi\text{-graph}\}. \end{aligned} \tag{5}$$

Therefore, $s(G)$ is the 'space' that we need in order to obtain a representation $G(\varphi,\mathcal{D})$ of G. We now define $\mu(G)$ to be the maximal cardinality of an equivalence class of $\sim_\varphi$, where we let φ vary among all functions such that G is a φ-graph:

$$\begin{aligned} \mu(G) = \max\{|C|\colon & C \text{ is an equivalence class of } \sim_\varphi, \\ & \text{where } \varphi \text{ is such that } G \text{ is a } \varphi\text{-graph}\}. \end{aligned} \tag{6}$$

Thus, $\mu(G)$ is the maximum 'multiplicity' with which a 'distance' (recall (*)) may occur in a representation $G(\varphi,\mathcal{D})$ of G.

Our third and fourth parameters concern the *number* of 'distances' that we use in our representations of G. We let

$$D(G) = \min\{|\operatorname{Dist}(\varphi)|\colon G \text{ is a } \varphi\text{-graph}\}. \tag{7}$$

Hence, $D(G)$ is the minimal possible number of 'distances' that are induced by the $\varphi\colon [n]\to\mathbb{Z}^+$ in the representations $G(\varphi,\mathcal{D})$ of G. We may also restrict ourselves to counting distinct 'distances' that occur as *edges* in our representations of G. We let

$$D_{\mathrm{e}}(G) = \min\{|\mathcal{D}|\colon G \text{ is a } \mathcal{D}\text{-graph}\}. \tag{8}$$

Thus, $D_{\mathrm{e}}(G)$ is the minimum number of 'edge lengths' that are required in the representations $G(\varphi,\mathcal{D})$ of G.

Note that, in the definitions above, we may restrict ourselves to injective functions φ, as G has n vertices and we always wish to have G isomorphic to $G(\varphi,\mathcal{D})$.

It is easy to see that $s(K_n)=s(C_n)=s(P_n)=n$, where, as usual, C_n is the cycle on n vertices and P_n is the path on n vertices. However, calculating the spacing number of a graph exactly can be delicate, even for very simple graphs. Note that $K_{m,n}$ denotes the complete bipartite graph with vertex classes of cardinality m and n.

Proposition 2. *For $m \geq n$, we have*

$$s(K_{m,n}) = \begin{cases} 2m & \text{if } m = n \\ 2m-1 & \text{if } m > n. \end{cases} \tag{9}$$

In order to prove the proposition, we must first prove the following.

Proposition 3. *For any $n > 1$, we have $s(K_{1,n}) = 2n - 1$.*

To see that $s(K_{1,n}) \leq 2n - 1$ consider $G(\varphi, \mathcal{D})$, where

$$\mathrm{im}(\varphi) = \{1, 2, 3, 5, 7, \ldots, 2n-1\} \tag{10}$$

and

$$\mathcal{D} = \{1, 3, 5, \ldots, 2n-3\}. \tag{11}$$

Then $(\varphi, \mathcal{D})$ spaces $K_{1,n}$, and hence $s(K_{1,n}) \leq 2n - 1$. The reverse inequality will be proved in Section 2. Proposition 2 follows by monotonicity under taking induced subgraphs.

We now let

$$s(n) = \max_G s(G), \quad \mu(n) = \min_G \mu(G), \quad D(n) = \max_G D(G), \tag{12}$$

and

$$D_{\mathrm{e}}(n) = \max_G D_{\mathrm{e}}(G), \tag{13}$$

where the maxima are taken over all n-vertex graphs G. For positive integers r, we also let

$$s_r(n) = \max_G s(G), \tag{14}$$

where the maximum is taken over all r-regular graphs G on n vertices. Our main concern in this note is to give estimates for the extremal functions $s(n)$, $s_r(n)$, $\mu(n)$, $D(n)$, and $D_{\mathrm{e}}(n)$.

1.2. **Statement of the results.** The following are our results pertaining to the parameters given above. The proofs will be given in Sections 4 and 5.

We start with the extremal functions $s(n)$ and $s_r(n)$.

Theorem 4. *We have*

$$n^2\left(1 - (4 + o(1))\sqrt{\frac{\log n}{n}}\right) \leq s(n) \leq n^2 + O(n^{1.05}). \tag{15}$$

Theorem 5. (*i*) *For every fixed integer $r \geq 2$ and for all n with rn even, we have*

$$s_r(n) \geq (1 + o(1)) \left(\frac{1}{r!}\right)^{2/(r+2)} \left(\frac{r}{\mathrm{e}\sqrt{2}}\right)^{2r/(r+2)} n^{2r/(r+2)}. \tag{16}$$

(*ii*) *Let $r = r(n)$ be an integer function satisfying* $\log n \ll r \leq (n-1)/2$ *and with rn even for all n. Then*

$$s_r(n) = (1 + o(1))n^2. \tag{17}$$

Our result concerning the 'multiplicity' parameter $\mu(n)$ is as follows.

Theorem 6. *For all sufficiently large n, we have*

$$\mu(n) < 2 + 8\log_2 n. \tag{18}$$

The last three results deal with the functions $D(n)$ and $D_{\mathrm{e}}(n)$ and the parameter $D_{\mathrm{e}}(G)$, concerning the number of distinct distances that are involved in our representations of graphs.

Theorem 7. *For any $\varepsilon > 0$, there is $n_0 = n_0(\varepsilon)$ such that if $n \geq n_0$, then*

$$D(n) > \binom{n}{2} - 4n\log_2 n + 2n - \frac{1}{2}\log_2(2\pi n) - \varepsilon. \tag{19}$$

Theorem 8. *For almost all n-vertex graphs G, we have*

$$D_{\mathrm{e}}(G) \geq \frac{1}{2}\binom{n}{2} - (1+o(1))n^{3/2}\sqrt{\log n}. \tag{20}$$

Theorem 9. *For any function $\omega = \omega(n)$ such that $\omega(n) \to \infty$ as $n \to \infty$, we have*

$$D_{\mathrm{e}}(n) \geq \binom{n}{2} - 3\omega n^{3/2}\sqrt{\log n} - 3 \tag{21}$$

for all large enough n.

2. Proof of Proposition 3

We have already observed that $s(K_{1,n}) \leq 2n-1$. To establish the reverse inequality, we need the following claim.

Claim 10. *If $K_{1,n}$ is isomorphic to $G(\varphi, \mathcal{D})$, where $\mathrm{im}(\varphi) \subset [2n-1]$, then the root vertex is associated with neither 1 nor $\max \mathrm{im}(\varphi)$.*

Proof. Assume otherwise; specifically, assume that the root vertex of $K_{1,n}$ is associated with 1. Then there are n distinct differences in $\mathcal{D}$ corresponding to the edges in the graph. However, since 1 is associated with the vertex of degree n in our graph, the largest value, namely $\max \mathrm{im}(\varphi)$, must be associated with a leaf v. Hence there are $n-1$ distinct values not in $\mathcal{D}$ corresponding to the non-edges between v and the other leaves. However this implies that $2n-2 = \big|\mathcal{D}| + |[2n-2] \setminus \mathcal{D}\big| \geq n + n - 1 = 2n-1$, a contradiction. □

We can now prove the proposition.

Proof of Proposition 3. We proceed by induction on n, and note that the claim is easily shown true when $n = 2$ or $n = 3$. Let $(\varphi, \mathcal{D})$ be a spacing of $K_{1,n}$ ($n > 3$) that realizes the spacing number and assume that $\max \mathrm{im}(\varphi) = s(K_{1,n})$ is at most $2n-2$. If x is the vertex of degree n in the graph, then the claim tells us that x is not associated with 1 or $s(K_{1,n})$. Therefore, both 1 and $s(K_{1,n})$ must be leaves. Note that both 2 and $s(K_{1,n}) - 1$ must also be in $\mathrm{im}(\varphi)$ and associated with leaves, lest $s(K_{1,n-1}) \leq 2n-4$ (contradicting the induction assumption), or one of these points is associated with x, contradicting the claim. Now, assume that x is associated with some point p_i in $\mathrm{im}(\varphi)$. By the above, $p_i \notin \{1, 2, s(K_{1,n}) - 1,\ s(K_{1,n})\}$. Assume that there are $l \geq 2$ points associated with leaves in the interval $[1, p_i]$ and $n - l$ points associated with leaves in the interval $[p_i, s(K_{1,n})]$. Then the points from $\mathrm{im}(\varphi)$ in these intervals along with $\mathcal{D}$ comprise spacings of $K_{1,l}$ and $K_{1,n-l}$ respectively. Also, the root vertex of these stars are associated with the largest and smallest values in these intervals. In order to space $K_{1,l}$ in this way, the claim implies $p_i \geq 2l$. However, we then have $s(K_{1,n-l}) \leq s(K_{1,n}) - p_i \leq 2n - 2 - 2l = 2(n - l) - 2$, contradicting the induction assumption. □

As noted above, this also completes the proof of Proposition 2.

3. Auxiliary lemmas

In this section, we give two elementary results from combinatorial number theory.

3.1. **Quasi-Sidon sets.** To establish our bounds on $s(n)$, we shall make use of a classical object from combinatorial number theory. A set

$$A = \{a_1, \dots, a_n\} \tag{22}$$

is called a *Sidon set* if

$$\big|\{|a_i - a_j| \colon i < j\}\big| = \binom{n}{2}, \tag{23}$$

that is, if all of the differences of distinct elements are distinct. It was shown by Erdős and Turán in [9] that if $A \subset [N]$ is an n-element Sidon set, then

$$N \geq n^2 + O(n^{3/2}).$$

We shall examine a similar type of set, wherein the pairwise differences are nearly all distinct.

Lemma 11. *Let $0 < \varepsilon = \varepsilon(n) \leq 1$ be such that $\varepsilon n \to \infty$ as $n \to \infty$. Suppose $\varphi\colon [n] \to \mathbb{Z}^+$ is an injective function such that*

$$|\operatorname{Dist}(\varphi)| \geq (1 - \varepsilon)\binom{n}{2} \tag{24}$$

and

$$\mathrm{im}(\varphi) \subset [N]. \tag{25}$$

Then

$$N \geq \big(1 - (2 + o(1))\sqrt{\varepsilon}\big)n^2. \tag{26}$$

Proof. We follow [9] very closely. Let $\varepsilon = \varepsilon(n)$ and $\varphi\colon [n] \to \mathbb{Z}^+$ be as in the statement of our lemma. Let $\mathrm{im}(\varphi) = \{x_i \colon 1 \leq i \leq n\}$, where

$$x_1 < \cdots < x_n. \tag{27}$$

Let $m = \lfloor \varepsilon^{1/2} n^2 \rfloor$. Let us consider the following $N + m$ intervals of integers, each of cardinality m:

$$I_1 = [-m + 1, 1), I_2 = [-m + 2, 2), \dots, I_{N+m} = [N, N + m). \tag{28}$$

Let B_k be the number of x_i that belong to the interval $I_k = [-m+k, k)$. Let us count the triples $(x_i, x_j; I_k)$ for which we have $x_i < x_j$ and $x_i,\ x_j \in I_k$. Let T be the number of such triples.

We have

$$T = \sum_{1\le k\le N+m} \binom{B_k}{2} \ge \frac{1}{2}(N+m)\frac{nm}{N+m}\left(\frac{nm}{N+m} - 1\right), \qquad (29)$$

where the inequality follows from convexity. Let us now estimate T in an alternative way.

Pick $L \ge (1-\varepsilon)\binom{n}{2}$ pairs $(x_{i_\lambda}, x_{j_\lambda})$ $(1 \le \lambda \le L)$ with $x_{i_\lambda} < x_{j_\lambda}$ for all λ and all the differences

$$d_\lambda = x_{j_\lambda} - x_{i_\lambda} \qquad (1 \le \lambda \le L) \qquad (30)$$

distinct. Fix a pair $(x_{i_\lambda}, x_{j_\lambda})$. If $d_\lambda = x_{j_\lambda} - x_{i_\lambda} \ge m$, then clearly there is no k for which both x_{i_λ} and $x_{j_\lambda} \in I_k$. If $d_\lambda < m$, then there are $m - d_\lambda$ intervals I_k with both x_{i_λ} and $x_{j_\lambda} \in I_k$. Given that all the differences in these L pairs are distinct, we deduce that the number of triples of the form $(x_{i_\lambda}, x_{j_\lambda}; I_k)$ $(1 \le \lambda \le L)$ that we are interested in is

$$\le \sum_{1\le d<m} (m-d) = \binom{m}{2}. \qquad (31)$$

Now we have to take into account the triples $(x_i, x_j; I_k)$ for the pairs (x_i, x_j) that are not one of the $(x_{i_\lambda}, x_{j_\lambda})$ $(1 \le \lambda \le L)$. We use a crude bound for such pairs, of which there are at most $\varepsilon\binom{n}{2}$: each of the pairs occur in at most $m-1$ intervals I_k each. Therefore, we have

$$T \le \binom{m}{2} + \varepsilon\binom{n}{2}(m-1) \le \frac{1}{2}(m^2 + \varepsilon n^2 m). \qquad (32)$$

Comparing (29) and (32), we have

$$n\left(\frac{nm}{N+m} - 1\right) \le m + \varepsilon n^2, \qquad (33)$$

which is equivalent to

$$n^2 m \le (N+m)(m + \varepsilon n^2 + n). \qquad (34)$$

Since $\varepsilon n \to \infty$ as $n \to \infty$, we deduce from (34) that

$$\begin{aligned} N &\ge \frac{n^2 m}{m + (1+o(1))\varepsilon n^2} - m \ge n^2 \frac{1}{1 + (1+o(1))\varepsilon n^2/m} - \varepsilon^{1/2} n^2 \\ &\ge n^2\left(1 - (1+o(1))\varepsilon^{1/2}\right) - \varepsilon^{1/2} n^2, \end{aligned} \qquad (35)$$

and (26) follows. □

3.2. **The number of equivalence relations $\sim_\varphi$.** Recall that each $\varphi\colon [n] \to \mathbb{Z}^+$ defines an equivalence relation $\sim_\varphi$ on $\binom{[n]}{2}$. Let

$$T(n) = \big|\{\sim_\varphi\colon \varphi \in (\mathbb{Z}^+)^{[n]}\}\big|. \tag{36}$$

The upper bound for $T(n)$ given in Lemma 12 below will be crucial when estimating $D(n)$ and $D_{\rm e}(n)$. The reader may find it interesting to observe how small $T(n)$ is, in comparison with the total number of equivalence relations on $\binom{[n]}{2}$, which is given by the Bell number

$$b_{\binom{n}{2}} = \binom{n}{2}^{(1+o(1))\binom{n}{2}} = n^{(1+o(1))n^2} \tag{37}$$

(see, e.g., [11]).

Lemma 12. *For all $n \geq 3$, we have*

$$T(n) \leq \binom{2\binom{\binom{n}{2}}{2}}{n} \leq (1+o(1))\left(\frac{{\rm e}n^3}{4}\right)^n. \tag{38}$$

Proof. Let $\varphi_\lambda\colon [n] \to \mathbb{Z}^+$ $(\lambda \in \Lambda)$ be $T(n)$ maps with all the $\sim_{\varphi_\lambda}$ distinct. Moreover, let p be a sufficiently large prime; any choice with

$$p > \max_{\lambda\in\Lambda} (\max \operatorname{im}(\varphi_\lambda))^2 \tag{39}$$

will do. We may consider the maps φ_λ as maps into $\mathbb{F}_p$, the field of order p, in the obvious way:

$$\varphi_\lambda\colon [n] \to \mathbb{F}_p. \tag{40}$$

For convenience, let

$$x_{\lambda,i} = \varphi_\lambda(i) \quad (\lambda \in \Lambda,\ i \in [n]) \tag{41}$$

and let

$$\mathbf{x}_\lambda = (x_{\lambda,i})_{1\leq i\leq n}. \tag{42}$$

Let

$$\mathcal{R} = \binom{\binom{[n]}{2}}{2} = \binom{E(K^n)}{2}, \tag{43}$$

and consider the linear polynomials

$$f_\varrho(x_1,\dots,x_n) \in \mathbb{F}_p[x_1,\dots,x_n] \qquad (\varrho \in \mathcal{R}), \tag{44}$$

where, if $\varrho = \big\{\{i,j\},\{k,\ell\}\big\} \in \mathcal{R}$, then

$$f_\varrho(x_1,\dots,x_n) = (x_i - x_j)^2 - (x_k - x_\ell)^2. \tag{45}$$

We let $\mathbf{x} = (x_1,\dots,x_n)$ and $\mathbf{f}(\mathbf{x}) = (f_\varrho(\mathbf{x}))_{\varrho\in\mathcal{R}}$.

Given $\mathbf{z} = (z_\varrho)_{\varrho\in\mathcal{R}} \in \mathbb{F}_p^{\mathcal{R}}$, let the *zero pattern* $Z(\mathbf{z})$ of $\mathbf{z}$ be the $\{0,*\}$-vector $(s_\varrho)_{\varrho\in\mathcal{R}}$, where

$$s_\varrho = \begin{cases} 0 & \text{if } z_\varrho = 0 \\ * & \text{otherwise.} \end{cases} \tag{46}$$

We now consider the zero patterns $Z(\mathbf{f}(\mathbf{x}^*))$ of the vectors $\mathbf{f}(\mathbf{x}^*)$, where we let $\mathbf{x}^*$ vary in $\mathbb{F}_p^n$. Theorem 1.3 in [14] tells that

$$\begin{aligned}\big|\{Z(\mathbf{f}(\mathbf{x}^*))\colon \mathbf{x}^*\in\mathbb{F}_p^n\}\big| &\leq \binom{2|\mathcal{R}|}{n}\\ &= \binom{2\binom{\binom{n}{2}}{2}}{n} \leq \binom{n^4/4}{n} \leq (1+o(1))\left(\frac{\mathrm{e}n^3}{4}\right)^n. \qquad (47)\end{aligned}$$

It now suffices to relate zero patterns with the equivalence relations $\sim_{\varphi_\lambda}$. Recall (3), (39), (41), and (42), and observe that, for any $\sim_{\varphi_\lambda}$ and any $\varrho = \big\{\{i,j\},\{k,\ell\}\big\}\in\mathcal{R}$, we have

$$\begin{aligned}\{i,j\}\sim_{\varphi_\lambda}\{k,\ell\} &\iff |\varphi_\lambda(i)-\varphi_\lambda(j)| = |\varphi_\lambda(k)-\varphi_\lambda(\ell)|\\ &\iff (x_{\lambda,i}-x_{\lambda,j})^2 = (x_{\lambda,k}-x_{\lambda,\ell})^2 \iff \mathbf{f}_\varrho(\mathbf{x}_\lambda)=0 \text{ in } \mathbb{F}_p. \qquad (48)\end{aligned}$$

(Note that the 'if' part of the last 'if and only if' follows from (39).) By the definition of zero patterns, the equivalences in (48) tell us that the relations $\sim_{\varphi_\lambda}$ and $\sim_{\varphi_{\lambda'}}$ coincide if and only if $Z(\mathbf{f}(\mathbf{x}_\lambda)) = Z(\mathbf{f}(\mathbf{x}_{\lambda'}))$. This fact and (47) imply (38), and Lemma 12 is proved. □

4. Proofs of Theorems 4, 5, and 6

In this section we shall prove our results concerning the extremal functions $s(n)$, $s_r(n)$, and $\mu(n)$.

4.1. Proofs of Theorems 4 and 5.

We start with the spacing parameters $s(n)$ and $s_r(n)$.

Proof of Theorem 4. We shall show that

$$\left(1-(4+o(1))\sqrt{\frac{\log n}{n}}\right)n^2 \leq s(n) \leq n^2 + O(n^{1.05}). \qquad (49)$$

We prove the upper bound using Sidon sets. Let G be a graph with vertex set $\{v_1,\dots,v_n\}$ and let $A=\{a_1,\dots,a_n\}$ be a Sidon set. Then $(\varphi,\mathcal{D})$ spaces G, where $\varphi\colon[n]\to\mathbb{Z}^+$ is any function such that $\mathrm{im}(\varphi)=A$ and

$$\mathcal{D} = \big\{|a_i-a_j|\colon \{v_i,v_j\}\in E(G)\big\}. \qquad (50)$$

Thus, in order to estimate $s(n)$ from above, we wish to minimize the maximum element an n-element Sidon set. The following was proved by Chowla [7].

Lemma 13. *Let q be a prime power. Then there is a Sidon set $A\subset[q^2]$ with exactly q elements.*

It was shown by Baker, Harman, and Pintz [3] that, for any sufficiently large n, there is always a prime in the interval $[n,\ n+n^{.525}]$. Thus

$$s(n) \leq (n+n^{.525})^2, \qquad (51)$$

and the upper bound in (49) follows. We now turn to the lower bound on $s(n)$.

Let δ be a fixed positive real, and let

$$\varepsilon = \varepsilon(n) = \frac{4 \log n}{n-1}. \tag{52}$$

Suppose for a contradiction that

$$s(n) \leq N_0 = \left\lfloor n^2 \left(1 - (4+\delta)\sqrt{\frac{\log n}{n}} \right) \right\rfloor. \tag{53}$$

Let us invoke Lemma 11, with ε given in (52). Lemma 11 tells us that if $\varphi\colon [n] \to \mathbb{Z}^+$ is such that

$$|\operatorname{Dist}(\varphi)| \geq (1-\varepsilon)\binom{n}{2} \tag{54}$$

and $\operatorname{im}(\varphi) \subset [N]$, then

$$N \geq n^2\bigl(1 - (2+o(1))\sqrt{\varepsilon}\bigr) = n^2 \left(1 - (4+o(1))\sqrt{\frac{\log n}{n}} \right). \tag{55}$$

Comparing (53) and (55), we see that, for all $n \geq n_0(\delta)$, the $\varphi\colon [n] \to \mathbb{Z}^+$ that are such that $\operatorname{im}(\varphi) \subset [N_0]$ must all be such that (54) fails, that is,

$$|\operatorname{Dist}(\varphi)| < (1-\varepsilon)\binom{n}{2}. \tag{56}$$

Note that (53) tells us that if we consider the graphs $G(\varphi, \mathcal{D})$ with $\varphi\colon [n] \to \mathbb{Z}^+$ such that $\operatorname{im}(\varphi) \subset [N_0]$ and $\mathcal{D} \subset \mathbb{Z}^+$ arbitrary, then we obtain all n-vertex graphs up to isomorphism.

Remark 18(i) and (56), coupled with the fact that there are at least $2^{\binom{n}{2}}/n!$ graphs on n vertices up to isomorphism, tell us that

$$\frac{N_0^n}{n!} 2^{(1-\varepsilon)\binom{n}{2}} \geq \binom{N_0}{n} 2^{(1-\varepsilon)\binom{n}{2}} \geq \frac{1}{n!} 2^{\binom{n}{2}}. \tag{57}$$

Inequalities (52) and (57) give that $N_0 \geq n^2$, which contradicts (53) for all $n \geq n_0(\delta)$. This contradiction completes the proof of Theorem 4. □

The proof of Theorem 5 is similar to the proof of Theorem 4. The following estimates for the number of r-regular graphs will be used now. We first quote a classical result of Bender and Canfield [4] (see also [5, Corollary II.17]), in the case in which r is a fixed integer.

Theorem 14. *For any fixed integer $r \geq 2$, the number $R(n,r)$ of labelled, n-vertex r-regular graphs satisfies*

$$R(n,r) \geq (\sqrt{2} + o(1))\mathrm{e}^{-(r^2-1)/4} \left(\frac{r^{r/2}}{\mathrm{e}^{r/2} r!} \right)^n n^{rn/2} \tag{58}$$

as $n \to \infty$ with rn even.

For recent results concerning the number of r-regular graphs with $r = r(n) \to \infty$ as $n \to \infty$, see McKay and Wormald [13], where Theorem 14 is extended to $r = o(\sqrt{n})$. For a wider range of r, we shall make use of the following fact, which may be deduced from the well known theorems of Egorychev and Falikman [8, 10] and Brégman [6]. (A result stronger than Fact 15 may be proved with the same method; here, we only state a result that is enough for our purposes.)

Fact 15. *If $r = r(n)$ satisfies $\log n \ll r \le n/2$ and rn is even for all n, then the number $R(n, r)$ of labelled, r-regular graphs on n vertices satisfies*

$$R(n,r) \ge \binom{\binom{n}{2}}{rn/2}(1-o(1))^{rn}. \tag{59}$$

Note that Fact 15 implies that the number of r-regular graphs on n vertices is the same as the number of graphs with n vertices and $rn/2$ edges, up to a factor of the form $\exp\{-o(rn)\}$. We shall only give a brief sketch of the proof of Fact 15 here. (For an alternative proof, the reader is referred to Shamir and Upfal [15].)

Proof of Fact 15 (Sketch). For simplicity, we suppose that n and r are both even. First we examine the number $\mathrm{Bip}(n/2, r/2)$ of $r/2$-regular spanning subgraphs of $K_{n/2,n/2}$. In a standard way, using the theorems of Egorychev [8] and Falikman [10] that give a lower bound for the number of perfect matchings in regular bipartite graphs (van der Waerden's conjecture) and a result of Brégman [6] that gives an upper bound for such a number (Minc's conjecture), one may show that

$$\mathrm{Bip}(n/2, r/2) \ge \binom{n/2}{r/2}^{n/2}(1-o(1))^{rn}. \tag{60}$$

Next, one observes that

$$R(n,r) \ge \prod_{j\ge 1}\left[\mathrm{Bip}\left(\frac{n}{2^j}, \frac{r}{2^j}\right)\right]^{2^{j-1}} \tag{61}$$

(for simplicity, divisibility conditions are ignored in (61)). One may derive (59) from (60) and (61). □

We may now prove Theorem 5.

Proof of Theorem 5. We first prove (i). Note that if $r = 2$, the bound is trivial. Fix an integer $r \ge 3$. In what follows, rn is supposed to be even.

The number of graphs that are of the form $G(\varphi, \mathcal{D})$ with $\mathcal{D} \subset \mathbb{Z}^+$ and with $\varphi\colon [n] \to \mathbb{Z}^+$ such that $\mathrm{im}(\varphi) \subset [N]$ is

$$\le \binom{N}{n} \sum_{1\le j\le rn/2} \binom{N-1}{j}, \tag{62}$$

as we have $\binom{N}{n}$ choices for $\mathrm{im}(\varphi)$ and we may suppose $\mathcal{D} \subset [N-1]$.

If $s_r(n) \le N$, then the quantity in (58) divided by $n!$ has to be at most the quantity in (62). This gives that

$$(\sqrt{2}+o(1))\mathrm{e}^{-(r^2-1)/4}\left(\frac{r^{r/2}}{\mathrm{e}^{r/2}r!}\right)^n\frac{n^{rn/2}}{n!} \le \binom{N}{n}\sum_{1\le j\le rn/2}\binom{N-1}{j}. \tag{63}$$

We first observe that $N \ge rn$, for otherwise we would get that

$$(\sqrt{2}+o(1))\mathrm{e}^{-(r^2-1)/4}\left(\frac{r^{r/2}}{\mathrm{e}^{r/2}r!}\right)^n\frac{n^{rn/2}}{n!} \le \binom{rn}{n}2^{rn}, \tag{64}$$

which is equivalent to

$$\frac{r^{r/2}}{e^{r/2}r!}n^{r/2} \le (rn)2^r. \tag{65}$$

This is clearly false for $n \ge n_0(r)$, so we may therefore assume that $N > rn$. In this case, the quantity in (62) is at least

$$\frac{1}{2}rn\binom{N}{n}\binom{N}{rn/2}. \tag{66}$$

It must then be the case that

$$(\sqrt{2}+o(1))\mathrm{e}^{-(r^2-1)/4}\left(\frac{r^{r/2}}{\mathrm{e}^{r/2}r!}\right)^n\frac{n^{rn/2}}{n!} \le \frac{1}{2}rn\binom{N}{n}\binom{N}{rn/2}. \tag{67}$$

Taking the nth root, we obtain

$$(1+o(1))\frac{r^{r/2}}{\mathrm{e}^{r/2}r!}\mathrm{e}n^{r/2-1} \le \frac{\mathrm{e}N}{n}\left(\frac{2\mathrm{e}N}{rn}\right)^{r/2}, \tag{68}$$

whence

$$(1+o(1))\frac{r^r}{2^{r/2}\mathrm{e}^r r!}n^r \le N^{(r+2)/2}. \tag{69}$$

Therefore

$$N \ge (1+o(1))\left(\frac{1}{r!}\right)^{2/(r+2)}\left(\frac{r}{\mathrm{e}\sqrt{2}}\right)^{2r/(r+2)}n^{2r/(r+2)}, \tag{70}$$

and (16) follows. This completes the proof of (i).

Let us now turn to the proof of (ii). Let $r = r(n)$ as in the statement of the theorem be given. Clearly, it follows from the upper bound in (15) that $s_r(n) \le (1+o(1))n^2$. We need to prove the reverse inequality.

Assume that, for some fixed $\varepsilon > 0$,

$$s_r(n) \le (1-\varepsilon)n^2 \text{ for arbitrarily large } n. \tag{71}$$

We shall derive a contradiction from this assumption, and this will complete the proof of (17).

Lemma 11 tells us that if $\varphi\colon [n] \to \mathbb{Z}^+$ is such that

$$\operatorname{im}\varphi \subset [(1-\varepsilon)n^2], \tag{72}$$

then there exists some $\sigma > 0$ for which we have

$$|\operatorname{Dist}(\varphi)| < (1-\sigma)\binom{n}{2}. \tag{73}$$

Thus, the total number of φ-graphs with φ satisfying (72) is, up to isomorphism, at most

$$\binom{n^2}{n} \sum_{j \le rn/2} \binom{(1-\sigma)\binom{n}{2}}{j}. \tag{74}$$

We shall now compare (59) and (74). It will be convenient to consider separately two ranges for r.

Case 1. $r \le (1-\sigma)(n-1)/2$

In this case, we bound the quantity in (74) from above by

$$\binom{n^2}{n} \frac{rn}{2} \binom{(1-\sigma)\binom{n}{2}}{rn/2}. \tag{75}$$

Using that $r \gg \log n$, we see that the quantity in (75) is smaller than (59) for $n \ge n_0(\sigma)$. This contradicts (71), completing the proof in this case.

Case 2. $(1-\sigma)(n-1)/2 < r \le (n-1)/2$

In this case, we bound the quantity in (74) from above by

$$2^{(1-\sigma)\binom{n}{2}}(1+o(1))^{n^2}. \tag{76}$$

However, examining (59), one sees that in this case the number of r-regular graphs is at least

$$\binom{\binom{n}{2}}{\lfloor((1-\sigma)/2)\binom{n}{2}\rfloor}(1-o(1))^{n^2}. \tag{77}$$

We shall estimate (77) using the following well known fact, which is a consequence of Stirling's formula.

Fact 16. *Let $0 < \alpha < 1$ be fixed and let m be a positive integer. Then*

$$\binom{m}{\lfloor \alpha m \rfloor} = \left[\left(\frac{1}{\alpha}\right)^{\alpha}\left(\frac{1}{1-\alpha}\right)^{1-\alpha}\right]^{m}(1+o(1))^{m}, \tag{78}$$

where $o(1) \to 0$ as $m \to \infty$.

Applying Fact 16 to (77), we have that the number of r-regular graphs is at least

$$\left[\left(\frac{2}{1-\sigma}\right)^{(1-\sigma)/2}\left(\frac{2}{(1+\sigma)}\right)^{(1+\sigma)/2}\right]^{\binom{n}{2}}(1-o(1))^{n^2}. \tag{79}$$

As in the first case, assumption (71) tells us that the quantity in (76) should be at least as large as the quantity in (79), that is,

$$\left[\left(\frac{2}{1-\sigma}\right)^{(1-\sigma)/2}\left(\frac{2}{(1+\sigma)}\right)^{(1+\sigma)/2}\right]^{\binom{n}{2}} \leq 2^{(1-\sigma)\binom{n}{2}}(1+o(1))^{n^2}. \tag{80}$$

We shall now see that (80) does not hold if $n \geq n_0(\sigma)$.

Keeping in mind that $\sigma > 0$ is a constant and n tends to infinity, inequality (80) simplifies to

$$\left(\frac{1-\sigma}{2}\right)\log_2\frac{2}{1-\sigma} + \left(\frac{1+\sigma}{2}\right)\log_2\frac{2}{1+\sigma} \leq 1-\sigma. \tag{81}$$

Equivalently,

$$\left(\frac{1-\sigma}{2}\right)(1-\log_2(1-\sigma)) + \left(\frac{1+\sigma}{2}\right)(1-\log_2(1+\sigma)) \leq 1-\sigma; \tag{82}$$

that is,

$$(1-\sigma)\log_2(1-\sigma) + (1+\sigma)\log_2(1+\sigma) - 2\sigma \geq 0. \tag{83}$$

Let

$$f(\sigma) = (1-\sigma)\log_2(1-\sigma) + (1+\sigma)\log_2(1+\sigma) - 2\sigma. \tag{84}$$

Note that $f(0) = 0$ and

$$f'(\sigma) = \log_2\left(\frac{1+\sigma}{1-\sigma}\right) - 2 < 0. \tag{85}$$

This implies that f is decreasing for $\sigma \in (0, \frac{1}{2}]$. Hence, for any small $\sigma > 0$, the inequalities given in (80)–(83) fail for $n \geq n_0(\sigma)$. Thus, as in Case 1, assumption (71) cannot hold. This completes the proof of Case 2, and hence (*ii*) of Theorem 5 is proved. □

4.2. **Proof of Theorem 6.** We now consider the 'multiplicity' parameter $\mu(n)$.

Proof of Theorem 6. We wish to show that there exists a graph G such that for any $\varphi\colon [n] \to \mathbb{Z}^+$ such that G is a φ-graph, no element of $\mathrm{Dist}(\varphi)$ has cardinality $2 + 8\log_2 n$ or more.

Suppose $\varphi\colon [n] \to \mathbb{Z}^+$ is an injective function such that $\sim_\varphi$ has an equivalence class $C \subset \binom{[n]}{2}$ of cardinality $\mu \geq 2 + 8\log_2 n$. It is clear that C spans a disjoint union of paths on $[n]$. Therefore, there is a matching $M \subset C$ with $|M| \geq |C|/2$. Let $\nu = |M| \geq 1 + 4\log_2 n$. Considering this matching, one sees that there are 2ν distinct integers

$$x_1 < y_1, x_2 < y_2, \ldots, x_\nu < y_\nu \tag{86}$$

in $\mathrm{im}(\varphi)$ with all the differences $y_i - x_i$ $(1 \leq i \leq \nu)$ equal.

Clearly, by the definition of $G(\varphi, \mathcal{D})$, regardless of $\mathcal{D}$, the subgraphs of $G(\varphi, \mathcal{D})$ induced by the x_i $(1 \leq i \leq \nu)$ and the y_i $(1 \leq i \leq \nu)$ are isomorphic. To prove our theorem, namely, to verify (18), it suffices to prove the following fact.

Fact 17. *For all large enough n, there is an n-vertex graph G that does not contain two vertex disjoint, isomorphic induced subgraphs of order $\nu \geq 1 + 4\log_2 n$.*

Proof. Consider the binomial random graph $G = G(n,p)$ with $p = 1/2$, and let $\nu = \nu(n) \geq 1 + 4\log_2 n$ be given. Let $X_\nu = X_\nu(G)$ be the number of pairs (H, H') of induced subgraphs of G of order ν with H isomorphic to H'. Then

$$\mathbb{P}(X_\nu > 0) \leq \mathbb{E}(X_\nu) < n^{2\nu} 2^{-\binom{\nu}{2}} = \left(n^2 2^{-(\nu-1)/2}\right)^\nu \leq 1, \tag{87}$$

and therefore a graph as required does exist. □

As observed above, Fact 17 concludes the proof of Theorem 6. □

5. Proofs of Theorems 7, 8, and 9

Let us now prove our results concerning the number of distances in our representations of graphs. Before we proceed, we make the following remark on the number of graphs that may be 'generated' from a given φ.

Remark 18. (*i*) *Fix $\varphi\colon [n] \to \mathbb{Z}^+$. Then*

$$\left|\{G(\varphi, \mathcal{D})\colon \mathcal{D} \subset \mathbb{Z}^+\}\right| \leq 2^{|\operatorname{Dist}(\varphi)|}. \tag{88}$$

(*ii*) *Fix $\varphi\colon [n] \to \mathbb{Z}^+$ and let D be an integer. Then*

$$\left|\{G(\varphi, \mathcal{D})\colon |\mathcal{D}| \leq D\}\right| \leq \sum_{j=0}^{D} \binom{|\operatorname{Dist}(\varphi)|}{j} \leq \sum_{j=0}^{D} \binom{\binom{n}{2}}{j}. \tag{89}$$

5.1. **Proofs of Theorems 7 and 8.** We start with our result concerning $D(n)$.

Proof of Theorem 7. We wish to show that for any $\varepsilon > 0$ and any $n \geq n_0(\varepsilon)$, there exists a graph G on n vertices such that whenever G is a φ-graph we have

$$|\operatorname{Dist}(\varphi)| > \binom{n}{2} - 4n\log_2 n + 2n - \frac{1}{2}\log_2(2\pi n) - \varepsilon.$$

Let $D = \binom{n}{2} - 4n\log_2 n + 2n - (1/2)\log_2(2\pi n) - \varepsilon$, where $\varepsilon > 0$ is an arbitrary fixed constant, and suppose for a contradiction that $D(n) \leq D$. Note that if $\varphi\colon [n] \to \mathbb{Z}^+$ is such that $|\operatorname{Dist}(\varphi)| \leq D$, then, by Remark 18($i$), we have

$$\left|\{G(\varphi, \mathcal{D})\colon \mathcal{D} \subset \mathbb{Z}^+\}\right| \leq 2^{|\operatorname{Dist}(\varphi)|} \leq \frac{1}{2^\varepsilon\sqrt{2\pi n}} 2^{\binom{n}{2}} n^{-4n} 4^n. \tag{90}$$

By Lemma 12, as we let $\varphi\colon [n] \to \mathbb{Z}^+$ vary, we have

$$T(n) \leq (1 + o(1))\left(\frac{\mathrm{e}n^3}{4}\right)^n \tag{91}$$

distinct $\sim_\varphi$. By (90) and (91), if we only consider $\varphi\colon [n] \to \mathbb{Z}^+$ with $|\operatorname{Dist}(\varphi)| \le D$, we are able to produce

$$\le (1+o(1))\left(\frac{\mathrm{e}n^3}{4}\right)^n \left(\frac{4}{n^4}\right)^n \frac{2^{\binom{n}{2}}}{2^\varepsilon\sqrt{2\pi n}} = (1+o(1))\left(\frac{\mathrm{e}}{n}\right)^n \frac{2^{\binom{n}{2}}}{2^\varepsilon\sqrt{2\pi n}} \tag{92}$$

non-isomorphic graphs. On the other hand, using Stirling's formula $n! = (1+o(1))(n/\mathrm{e})^n\sqrt{2\pi n}$, we see that there are

$$\ge \frac{1}{n!}2^{\binom{n}{2}} \ge \frac{1+o(1)}{\sqrt{2\pi n}}\left(\frac{\mathrm{e}}{n}\right)^n 2^{\binom{n}{2}} \tag{93}$$

graphs on n vertices up to isomorphism. According to our assumption that $D(n) \le D$, the quantity in (92) must be at least the quantity in (93), which is not the case for $n \ge n_0(\varepsilon)$. This completes the proof of Theorem 7. □

Proof of Theorem 8. We shall prove that for almost all graphs G on n vertices, if G is a $\mathcal{D}$-graph, then

$$|\mathcal{D}| \ge \frac{1}{2}\binom{n}{2} - (1+o(1))n^{3/2}\sqrt{\log n}. \tag{94}$$

The proof of this result follows the same steps as the proof of Theorem 7, except that we shall make use of Remark 18(*ii*) instead of Remark 18(*i*).

For any given k, let us consider all $\mathcal{D}$-graphs for all $\mathcal{D} \subset \mathbb{Z}^+$ such that

$$|\mathcal{D}| \le D = \frac{1}{2}\binom{n}{2} - k. \tag{95}$$

By Remark 18(*ii*) and Lemma 12, the number of such graphs is, up to isomorphism,

$$\begin{aligned}\le T(n)\sum_{j=0}^{D}\binom{\binom{n}{2}}{j} &\le (1+o(1))\left(\frac{\mathrm{e}n^3}{4}\right)^n \sum_{j=0}^{D}\binom{\binom{n}{2}}{j} \\ &\le (1+o(1))\left(\frac{\mathrm{e}n^3}{4}\right)^n 2^{\binom{n}{2}} \exp\left\{-2k^2\binom{n}{2}^{-1}\right\},\end{aligned} \tag{96}$$

where in the last inequality we used that

$$\sum_{j=0}^{\lceil N/2-a\rceil}\binom{N}{j} \le 2^N \mathrm{e}^{-2a^2/N} \tag{97}$$

for all integers $N \ge 1$ and all real a (see, e.g., [1, Theorem A.1]). From (96), we see that if, for some $\varepsilon > 0$, an ε-fraction of the graphs on n vertices are such that $D_{\mathrm{e}}(G) \le \frac{1}{2}\binom{n}{2} - k$, then we must have

$$\frac{\varepsilon}{\sqrt{2\pi n}}\left(\frac{\mathrm{e}}{n}\right)^n 2^{\binom{n}{2}} \le (1+o(1))\left(\frac{\mathrm{e}n^3}{4}\right)^n 2^{\binom{n}{2}} \exp\left\{-2k^2\binom{n}{2}^{-1}\right\}, \tag{98}$$

which tells us that

$$\frac{\varepsilon}{\sqrt{2\pi n}}\left(\frac{4}{n^4}\right)^n \leq (1+o(1))\exp\left\{-2k^2\binom{n}{2}^{-1}\right\}. \tag{99}$$

Therefore, for large enough n, we have, say,

$$\frac{2}{\varepsilon}\sqrt{\pi n}\left(\frac{n^4}{4}\right)^n \geq \exp\left\{2k^2\binom{n}{2}^{-1}\right\}, \tag{100}$$

and hence

$$2k^2 \leq \binom{n}{2}\left(4n\log n - n\log 4 + \frac{1}{2}\log(\pi n) + \log\frac{2}{\varepsilon}\right), \tag{101}$$

which tells us that

$$k \leq (1+o(1))n^{3/2}\sqrt{\log n}. \tag{102}$$

As $\varepsilon > 0$ was arbitrary, this concludes the proof. □

5.2. **Proof of Theorem 9.** The proof of our result on $D_{\mathrm{e}}(n)$, related to the number of 'edge lengths' in our representations of graphs, will be somewhat more technical.

5.2.1. *Preliminary lemmas.* In this section, we shall state and prove two simple inequalities related to a certain enumeration problem for an algebra of sets. Suppose we have a set X of cardinality N, and suppose we have an equivalence relation $\sim$ defined on X. We are interested in the sets that may be obtained by taking unions of equivalence classes of $\sim$. In fact, we are interested in such unions that have a given cardinality T. In what follows, we shall have $N = \binom{n}{2}$ and T quite close to $\binom{n}{2}$.

Both the statement and the proof of our first lemma are purely numerical, and hence we postpone its combinatorial interpretation (see the discussion after the proof of Lemma 19).

Let a pair of non-negative integers $\mathbf{b} = (b_1, b_2)$ such that

$$b_1 + 2b_2 = \binom{n}{2} \tag{103}$$

be given. For all non-negative integers T and real $g \geq 0$, we let

$$c(\mathbf{b}; T, g) = c(b_1, b_2; T, g) = \sum \binom{b_1}{t_1}\binom{b_2}{t_2}, \tag{104}$$

where the sum is over all integers t_1 and t_2 such that $t_1+2t_2 = T$, and $t_2 \geq g$. Note that, in (104), we may ignore the terms with $t_2 > b_2$. In what follows, we always suppose that $g \leq t_2 \leq b_2$.

Lemma 19. *Let $\omega = \omega(n) \to \infty$ as $n \to \infty$, and suppose $\omega = o(\log\log n)$. For any $\mathbf{b} = (b_1, b_2)$ as above, if $g \geq \omega n^{3/2}\sqrt{\log n}$ and*

$$T = \binom{n}{2} - \left\lceil \omega n^{3/2}\sqrt{\log n}\right\rceil, \tag{105}$$

then

$$c(b_1, b_2; T, g) \le n^{-\omega n} \binom{\binom{n}{2}}{T} \tag{106}$$

for all large enough n.

Proof. We start by splitting the sum in (104) into two parts. Let us write $\sum_1$ for the sum over all integers t_1 and t_2 such that $t_1 + 2t_2 = T$, $t_2 \ge g$, and, moreover,

$$\binom{b_2}{t_2} \ge 2n^{\omega n}. \tag{107}$$

Let $\sum_2$ denote the sum over the remaining pairs (t_1, t_2) in (104). Therefore,

$$c(b_1, b_2; T, g) = \sum\nolimits_1 \binom{b_1}{t_1}\binom{b_2}{t_2} + \sum\nolimits_2 \binom{b_1}{t_1}\binom{b_2}{t_2}. \tag{108}$$

The first sum in (108) may be easily estimated from above, since

$$\begin{aligned} 2n^{\omega n} \sum\nolimits_1 \binom{b_1}{t_1}\binom{b_2}{t_2} &\le \sum\nolimits_1 \binom{b_1}{t_1}\binom{b_2}{t_2}^2 \\ &\le \sum\nolimits_1 \binom{b_1}{t_1}\binom{2b_2}{2t_2} \le \binom{\binom{n}{2}}{T}. \end{aligned} \tag{109}$$

Let us now consider the second sum in (108); that is, let us now consider t_1 and t_2 for which we have $\binom{b_2}{t_2} < 2n^{\omega n}$. We shall estimate the terms

$$\binom{b_1}{t_1}\binom{b_2}{t_2} < 2n^{\omega n}\binom{b_1}{t_1} \tag{110}$$

by bounding $\binom{b_1}{t_1}$. Let us first derive some conclusions from the fact that $\binom{b_2}{t_2} < 2n^{\omega n}$. Suppose $b_2 \ge 2t_2$. Then

$$2^{t_2} \le \left(\frac{b_2}{t_2}\right)^{t_2} \le \binom{b_2}{t_2} < 2n^{\omega n}, \tag{111}$$

and hence $g \le t_2 < 1 + \omega n \log_2 n$, which is a contradiction for large enough n. Therefore, we have $b_2 = t_2 + s$, where $0 \le s < t_2$. But then

$$2^s \le \left(\frac{b_2}{s}\right)^s \le \binom{b_2}{s} = \binom{b_2}{t_2} < 2n^{\omega n}, \tag{112}$$

whence

$$s < 1 + \omega n \log_2 n. \tag{113}$$

For convenience, put $N = \binom{n}{2}$. Now observe that

$$\begin{aligned} \binom{b_1}{t_1} = \binom{N - 2b_2}{T - 2t_2} &= \binom{N}{T} \frac{(N-T)_{2(b_2 - t_2)}(T)_{2t_2}}{(N)_{2b_2}} \\ &= \binom{N}{T} \frac{(N-T)_{2s}}{(N)_{2s}} \frac{(T)_{2t_2}}{(N-2s)_{2t_2}}, \end{aligned} \tag{114}$$

which, by (105) and (113), is

$$\begin{aligned}
&\leq \binom{N}{T}\left(\frac{N-T}{N}\right)^{2s}\left(\frac{T}{N-2s}\right)^{2t_2}\\
&\quad\leq \binom{N}{T}\left(\frac{\lceil \omega n^{3/2}\sqrt{\log n}\rceil}{N}\right)^{2s}\left(\frac{N-\omega n^{3/2}\sqrt{\log n}}{N-2(1+\omega n\log_2 n)}\right)^{2t_2}\\
&\quad\leq \binom{N}{T}\left(\frac{1-2\omega\sqrt{(\log n)/n}}{1-(3\omega\log_2 n)/n}\right)^{2b_2} \leq \binom{N}{T}\left(1-\omega\sqrt{\frac{\log n}{n}}\right)^{2b_2}. \qquad (115)
\end{aligned}$$

Using that $b_2 \geq t_2 \geq g \geq \omega n^{3/2}\sqrt{\log n}$, we see that the quantity in (115) is

$$\leq \binom{N}{T}\exp\left(-2\omega^2 n\log n\right) \leq n^{-2\omega^2 n}\binom{N}{T}. \qquad (116)$$

Therefore,

$$\sum\nolimits_2 \binom{b_1}{t_1}\binom{b_2}{t_2} < 2Tn^{-2\omega^2 n+\omega n}\binom{\binom{n}{2}}{T} < \frac{1}{2}n^{-\omega n}\binom{\binom{n}{2}}{T}. \qquad (117)$$

Putting together (108), (109), and (117), we get that

$$c(b_1,b_2;T,g) = \sum\nolimits_1 \binom{b_1}{t_1}\binom{b_2}{t_2} + \sum\nolimits_2 \binom{b_1}{t_1}\binom{b_2}{t_2} \leq n^{-\omega n}\binom{\binom{n}{2}}{T}, \qquad (118)$$

as required. □

We may interpret Lemma 19 as follows. Suppose we have a partition of $\binom{[n]}{2}$ into b_1 blocks of cardinality 1 and b_2 blocks of cardinality 2. Let us call a set $E \subset \binom{[n]}{2}$ compatible with this partition if E is a union of blocks of this partition. We are interested in compatible, T-element subsets E of $\binom{[n]}{2}$, with the further restriction that we now describe.

For each such set E, let us say that we pay $i-1$ pence for every block of cardinality i that we use in the decomposition of E as a union of blocks. In the function $c(\mathbf{b};T,g)$ defined in (104), we are only counting those sets E for which we pay at least g pence, and Lemma 19 tells us that not many sets E cost at least g.

In our next lemma, we consider a general partition π of $\binom{[n]}{2}$, with b_i blocks of cardinality i ($i \geq 1$). Lemma 20 tells us that the number of T-element subsets E of $\binom{[n]}{2}$ with cost $\geq 2g$, compatible with π, is dominated by the number of T-element subsets E of $\binom{[n]}{2}$ with cost $\geq g$, compatible with π', a certain partition with blocks of cardinality 1 and 2 only.

Let us describe the setup that we consider in Lemma 20, which is, in fact, a purely numerical result. Suppose we have a sequence $\mathbf{b} = (b_1, b_2, \dots)$ of non-negative integers such that

$$b_1 + 2b_2 + 3b_3 + \cdots = \binom{n}{2}. \qquad (119)$$

For all non-negative integers T and real $g \geq 0$, we let

$$C(\mathbf{b}; T, g) = C(b_1, b_2, \dots; T, g) = \sum \prod_{i \geq 1} \binom{b_i}{t_i}, \tag{120}$$

where the sum is over all sequences $\mathbf{t} = (t_1, t_2, \dots)$ of non-negative integers such that

$$t_1 + 2t_2 + 3t_3 + \cdots = T \tag{121}$$

and

$$t_2 + 2t_3 + 3t_4 + \cdots = \sum_{i \geq 2} (i-1) t_i \geq g. \tag{122}$$

Given $\mathbf{b} = (b_1, b_2, \dots)$ as above, let b'_1 and b'_2 be given by

$$b'_1 = b_1 + b_3 + b_5 + \cdots = \sum_{k \geq 0} b_{2k+1} \tag{123}$$

and

$$b'_2 = b_2 + b_3 + 2b_4 + 2b_5 + \cdots = \sum_{k \geq 2} \left\lfloor \frac{k}{2} \right\rfloor b_k. \tag{124}$$

Note that, then, we have

$$b'_1 + 2b'_2 = \binom{n}{2}. \tag{125}$$

Lemma 20. *Let* $\mathbf{b}$*,* T*, and* g *be as above. Then*

$$C(\mathbf{b}; T, 2g) \leq c(b'_1, b'_2; T, g). \tag{126}$$

Proof. Consider a partition π of $\binom{[n]}{2}$ into b_i blocks of size i for all $i \geq 1$. Then π induces another partition π' of $\binom{[n]}{2}$ into b'_1 blocks of size 1 and b'_2 blocks of size 2, where b'_1 and b'_2 are as in (123) and (124). We construct π' from π by arbitrarily decomposing every block of size $2k$ into k blocks of size 2, and arbitrarily decomposing every block of size $2k+1$ into one singleton block and k blocks of size 2.

Consider E, any T-element subset of $\binom{[n]}{2}$ that is a union of blocks from π. Then, clearly, E is a T-element subset composed solely of blocks from π'. It remains to show that if E has 'cost' at least $2g$ in π, then G has cost at least g in π'.

Consider any block with $2k+1$ elements contained in E. In π, such a block would cost $2k$ pence. The corresponding blocks in π' would consist of a singleton block and k blocks of size 2, and hence cost only k pence in total. Similarly, a block of size $2k$ in π would cost $2k-1$ pence, while the corresponding k blocks of size 2 in π' would also cost k pence in total. This implies that the cost of any block in π is reduced by at most half in π'. Thus, if the total cost of E is at least $2g$ pence in π, then E would cost at least g pence in π', and the result holds. □

5.2.2. *The proof.* Having dealt with the technical lemmas in Section 5.2.1, the proof of Theorem 9 will be straightforward. Recall that we wish to show that for any sufficiently large n and $\omega = \omega(n)$ such that $\omega(n) \to \infty$ as $n \to \infty$, there exists a graph G on n vertices such that whenever G is a $\mathcal{D}$-graph, then

$$|\mathcal{D}| \geq \binom{n}{2} - 3\omega n^{3/2}\sqrt{\log n} - 3.$$

Proof of Theorem 9. Let $\omega = \omega(n)$ as in the statement of our theorem be given.

Fix $\varphi\colon [n] \to \mathbb{Z}^+$, and consider the equivalence relation $\sim_\varphi$ defined on $\binom{[n]}{2}$. Let $\mathbf{b} = (b_1, b_2, \dots)$, where b_i $(i \geq 1)$ is the number of blocks, i.e. equivalence classes, of $\sim_\varphi$ with cardinality i.

Let us write $\mathcal{G}(n,T)$ for the class of n-vertex graphs with T edges, up to isomorphism, where

$$T = \binom{n}{2} - \left\lceil \omega n^{3/2}\sqrt{\log n} \right\rceil. \tag{127}$$

We have

$$|\mathcal{G}(n,T)| \geq \frac{1}{n!}\binom{\binom{n}{2}}{T}. \tag{128}$$

Let

$$g = \left\lceil \omega n^{3/2}\sqrt{\log n} \right\rceil \tag{129}$$

and observe that the number of graphs in $\mathcal{G}(n,T)$ that are isomorphic to some $G(\varphi,\mathcal{D})$ $(\mathcal{D} \subset \mathbb{Z}^+)$ with

$$T - |\mathcal{D}| \geq 2g = 2\left\lceil \omega n^{3/2}\sqrt{\log n} \right\rceil \tag{130}$$

is

$$\leq C(\mathbf{b}; T, 2g) \leq n^{-\omega n}\binom{\binom{n}{2}}{T}, \tag{131}$$

where the second inequality in (131) follows from Lemmas 19 and 20.

Recall that we are considering a fixed $\varphi\colon [n] \to \mathbb{Z}^+$. However, as we consider all $\varphi\colon [n] \to \mathbb{Z}^+$, we only have

$$T(n) \leq (1 + o(1))\left(\frac{\mathrm{e}n^3}{4}\right)^n \tag{132}$$

distinct equivalence relations (see (36) and Lemma 12). Therefore, using (131) and (132), we see that the number of graphs in $\mathcal{G}(n,T)$ that are $\mathcal{D}$-graphs with $\mathcal{D}$ satisfying (130) is

$$\leq T(n)C(\mathbf{b}; T, 2g) \leq (1 + o(1))\left(\frac{\mathrm{e}n^3}{4}\right)^n n^{-\omega n}\binom{\binom{n}{2}}{T}, \tag{133}$$

which is a great deal smaller than the lower bound given in (128). This shows that there is a graph G_0 in $\mathcal{G}(n,T)$ such that if G_0 is isomorphic to $G(\varphi,\mathcal{D})$ for some $\varphi\colon[n]\to\mathbb{Z}^+$ and $\mathcal{D}\subset\mathbb{Z}^+$, then $T-|\mathcal{D}|<2g$. This graph G_0 shows that

$$\begin{aligned} D_{\mathrm{e}}(n) &= \max_G D_{\mathrm{e}}(G) > T-2g \\ &= T-2\left\lceil \omega n^{3/2}\sqrt{\log n}\right\rceil > \binom{n}{2} - 3\omega n^{3/2}\sqrt{\log n} - 3, \qquad (134) \end{aligned}$$

as required. □

6. Concluding remarks

Theorems 8 and 9 may be generalized to representations of graphs in high-dimensional Euclidean spaces. More precisely, we consider the graphs $G(\varphi,\mathcal{D})$, but now with $\varphi\colon[n]\to\mathbb{R}^d$ and $\mathcal{D}\subset\mathbb{R}$. Let

$$\begin{aligned} D_{\mathrm{e}}^{\mathbb{R}}(d,G) = \min\{|\mathcal{D}|\colon \mathcal{D}\subset\mathbb{R} &\text{ for which } \exists\varphi\in\left(\mathbb{R}^d\right)^{[n]} \\ &\text{such that } G \text{ is isomorphic to } G(\varphi,\mathcal{D})\}, \qquad (135) \end{aligned}$$

Note that, even for $d=2$, it becomes easier to represent graphs in the form $G(\varphi,\mathcal{D})$ in $\mathbb{R}^d$: an immediate argument shows that any tree may be drawn on the plane in such a way that all edges are straight lines of length 1, and no two non-adjacent vertices are at distance 1. Therefore, $D_{\mathrm{e}}^{\mathbb{R}}(2,T)=1$. We mention that bounded degree graphs may be represented in bounded dimension: as proved in [12], if a graph G has maximum degree Δ, then $D_{\mathrm{e}}^{\mathbb{R}}(2\Delta,G)=1$.

Let us now put $D_{\mathrm{e}}^{\mathbb{R}}(d,n)=\max_G D_{\mathrm{e}}^{\mathbb{R}}(d,G)$, where the maximum is taken over all n-vertex graphs.

As long as $d=o(n/\log n)$, a bound similar to the one in Theorem 9 holds for $D_{\mathrm{e}}^{\mathbb{R}}(d,n)$. Moreover, one may also prove a bound similar to the one in Theorem 8 for $D_{\mathrm{e}}^{\mathbb{R}}(d,G)$ for almost all n-vertex graphs G. To prove these results, one may use a well known result of Warren [16].

In addition to these higher-dimensional remarks, one may consider several problems. First, is it possible to pair the lower bounds on $D(n)$ or $D_{\mathrm{e}}(n)$ with similar upper bounds? Can one give a non-trivial bound from below for $\mu(n)$?

It seems that a challenging problem is to find a non-trivial upper bound on $s_r(n)$ or, in fact, to bound the spacing numbers of any broad class of graphs from above. In [2], a problem of representing trees in a similar manner is considered, and remains open.

It was demonstrated above (via $K_{1,n}$) that determining the spacing number of even very simple graphs may be somewhat delicate. It would be of some interest to find exactly the spacing numbers of more graphs.

6.1. **Acknowledgements.** The authors would like to thank John Schmitt and Brian Wagner for their help in proving Proposition 2. We would also like to extend our thanks to B. Bollobás, A. Kostochka and N. Linial for their helpful questions and input. We are greatly appreciative for the input of N. Wormald on the subject of enumerating regular graphs. Finally, many thanks to the referee for several excellent suggestions.

References

1. Noga Alon and Joel H. Spencer, *The probabilistic method*, John Wiley & Sons Inc., New York, 1992, With an appendix by Paul Erdős, A Wiley-Interscience Publication. MR **93h:**60002
2. L. Babai and V.T. Sós, *Sidon sets in groups and induced subgraphs of Cayley graphs*, European Journal of Combinatorics **6** (1985), no. 2, 101–114.
3. R.C. Baker, G. Harman, and J. Pintz, *The difference between consecutive primes, II.*, Proc. London Math. Soc. **83** (2001), no. 3, 532–562.
4. E. Bender and E.R. Canfield, *The asymptotic number of labeled graphs with given degree sequences*, J. Combinatorial Theory Ser. A **24** (1978), no. 3, 296–307. MR 58 #21793
5. B. Bollobás, *Random graphs*, Academic Press Inc. [Harcourt Brace Jovanovich Publishers], London, 1985. MR **87f:**05152
6. L. M. Brègman, *Certain properties of nonnegative matrices and their permanents*, Dokl. Akad. Nauk SSSR **211** (1973), 27–30. MR 48 #6130
7. S. Chowla, *Solution of a problem of Erdős and Turán*, Proc. Nat. Acad. Sci India. Sect A **14** (1944), 1–2.
8. G. P. Egoryčev, *Reshenie problemy van-der-Vardena dlya permanentov*, Akad. Nauk SSSR Sibirsk. Otdel. Inst. Fiz., Krasnoyarsk, 1980, Preprint IFSO-13 M. MR **82e:**15006
9. P. Erdős and P. Turán, *On a problem of Sidon in additive number theory, and on some related problems*, J. London Math. Soc. **16** (1941), 212–215.
10. D. I. Falikman, *Proof of the van der Waerden conjecture on the permanent of a doubly stochastic matrix*, Mat. Zametki **29** (1981), no. 6, 931–938, 957. MR **82k:**15007
11. Ronald L. Graham, Donald E. Knuth, and Oren Patashnik, *Concrete mathematics*, second ed., Addison-Wesley Publishing Company, Reading, MA, 1994, A foundation for computer science. MR **97d:**68003
12. H. Maehara and V. Rödl, *On the dimension to represent a graph by a unit distance graph*, Graphs and Combinatorics **6** (1990), 365–367.
13. Brendan D. McKay and Nicholas C. Wormald, *Asymptotic enumeration by degree sequence of graphs with degrees* $o(n^{1/2})$, Combinatorica **11** (1991), no. 4, 369–382. MR **93e:**05002
14. L. Rónyai, L. Babai, and M.K. Ganapathy, *On the number of zero-patterns of a sequence of polynomials*, J. Amer. Math. Soc. **14** (2001), no. 3, 717–735. MR **2002f:**11026
15. E. Shamir and E. Upfal, *Large regular factors in random graphs*, Convexity and graph theory (Jerusalem 1981) (Amsterdam–New York), North Holland Math. Stud., vol. 87, North Holland, 1984, pp. 271–282.
16. Hugh E. Warren, *Lower bounds for approximation by nonlinear manifolds*, Trans. Amer. Math. Soc. **133** (1968), 167–178. MR 37 #1871

Department of Mathematics and Computer Science, Emory University, Atlanta, GA 30322, USA

E-mail address: {mferrar|yoshi|rodl}@mathcs.emory.edu