

# Results on the Wiener profile

Morteza Faghani

Department of Mathematics, Payame Noor University,

P.O. Box 19395-3697, Tehran, Iran

and

Gyula O.H. Katona

MTA Rényi Institute, Budapest, Hungary

## 1 Introduction

Let  $G = (V, E)$  be a simple connected graph (without loops and multiple edges). The distance  $d(u, v)$  of two vertices  $u, v \in V$  is the number of edges in a shortest path between  $u$  and  $v$ . The *Wiener index*  $w(G)$  of  $G$  is the sum of these distances for all distinct pairs  $d(u, v)$ . This concept was introduced by Harry Wiener, who showed that this graph parameter is closely correlated with some chemical parameters, for instance with the boiling points of alkane molecules [11]. As an example, the Wiener indices of a large classes of fullerenes are determined in [6]. The book [7] contains many analogues and generalizations of the Wiener index, all of them have relevance in chemistry.

Let us introduce a new vector-parameter, the *Wiener profile* as the  $n - 1$ -dimensional vector  $(f_1, f_2, \dots, f_{n-1})$  where  $f_k$  is the number of pairs of distinct vertices  $(u, v)$  with distance  $d(u, v) = k$ . It is easy to see that the Wiener index is equal to

$$w(G) = \sum_{i=1}^{n-1} i f_i.$$

The contribution of a pair  $(u, v)$  of vertices (atoms in a molecule) to the Wiener index is  $i$  if their distance is  $d(u, v) = i$ . However, if we want to have a parameter which indicates a certain chemical property then the contribution of a pair  $(u, v)$  could be different from their distance. Suppose that it is some  $\alpha_i$  in case when  $d(u, v) = i$ . Here we supposed that the contribution of the

pair  $(u, v)$  to the property depends only on their distance. But the choice of the numbers  $\alpha_i$  might depend on the chemical property in question. Let  $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_{n-1})$  be the vector of these coefficients. Then the *weighed Wiener index* with these coefficients is

$$w_\alpha = \sum_{i=1}^{n-1} \alpha_i f_i.$$

The choice  $\alpha = (1, 2, \dots, n-1)$  gives back the Wiener index.

We will study the structure of the set of Wiener profiles of graphs on  $n$  vertices. Suppose e.g. that we want to find the molecule containing  $n$  atoms and having a certain chemical property the most or least, then we have to maximize or minimize  $w_\alpha$  under the condition that  $(f_1, f_2, \dots, f_{n-1})$  is the Wiener profile of a graph of  $n$  vertices ( $\alpha$  is fixed!). From the chemical point of view it makes sense only to consider connected graphs, but technically it is easier to handle the set of all graphs on  $n$  vertices. If a graph is not connected disregard the pairs in different components. Take all Wiener profiles of the graphs on at most  $n$  vertices (in other words the subgraphs of the  $n$ -vertex complete graph  $K_n$ .) Denote this set by  $\mathbb{W}_n$ . It is a set of points with integer coordinates in the  $n-1$ -dimensional Euclidean space.

Let us note that *profile vectors* were also studied in Extremal Set Theory (see e.g. [5]). There is, however a huge difference between the structures of the sets of profile vectors there and here. In Extremal Set Theory, if the coordinates of a profile vectors are decreased (remaining in the non-negative range) then the new vector is also a profile vector. This is not true here at all, as the  $n=3$  shows. There are 4 non-isomorphic graphs on three vertices: the complete graph  $K_3$ , the path of two edges, a single edge and the empty graph. Their profiles in this order are  $(3, 3), (2, 1), (1, 0), (0, 0)$ . That is,  $\mathbb{W}_3 = \{(3, 3), (2, 1), (1, 0), (0, 0)\}$ . One can see that  $(3, 2)$  is not a profile though  $(3, 3)$  is.

Our main task is to give necessary and/or sufficient conditions on vectors to make them Wiener profiles. There is one trivial condition. Since  $f_i$  is counting the number of distances  $i$ , their sum counts the total number of distances.

$$\sum_{i=1}^{n-1} f_i = \binom{n}{2}.$$

But there are more complex relations among the coordinates  $f_i$ . We will introduce them below. However, before that, we suggest a new invariant of

graphs that intuitively also has significance in chemistry. In the Wiener index and in the Wiener profile the pairs of vertices with distance  $k$  are considered. The distance is defined by a shortest path between the two vertices. But there might be more paths of the same (shortest) length between the vertices. The molecule is probably more rigid if the vertices are connected with more paths. This motivation makes us to introduce the concept of the *path profile*. Let  $p_k$  be the number of paths of length  $k$  in the graph  $G$  containing  $n$  vertices ( $1 \leq k < n$ ). The path profile is the vector  $(p_1, p_2, \dots, p_{n-1})$ . The *path index* is then

$$\pi(G) = \sum_{i=1}^{n-1} ip_i.$$

Here we will call the attention on some more complex necessary conditions that  $p_k$ 's and  $f_k$ 's must satisfy. Suppose that  $p_k$  is known, fixed. One can intuitively feel that  $p_{k-1}$  cannot be too small. Indeed, the graph contains  $p_k$  paths of length  $k$ , these paths contain many paths of length  $k-1$ . This fact must give a lower bound on the number  $p_{k-1}$ . Let us use a somewhat different approach, considering the set  $E$  of edges. The  $p_k$  paths are its  $k$ -element subsets. Deleting the first or last edge from such a path, a path of length  $k-1$  is obtained. What is the minimum number of these  $k-1$ -element subsets? The literature knows a very closely related problem/results. Let  $\mathcal{A}$  be a family of  $k$ -element sets. The *shadow*  $\sigma(\mathcal{A})$  is a family of those  $k-1$ -element sets which are obtained from the members of  $\mathcal{A}$  deleting one (arbitrary) element. The Shadow Theorem [8], [9] determines the exact minimum of  $|\sigma(\mathcal{A})|$  for given  $k$  and  $|\mathcal{A}|$  (where  $|S|$  denote the size of the set  $S$ ). We will study an analogous problem here.

The *Path Shadow Problem*. Knowing the the number  $p_k$  of paths of length  $k$  in a graph, determine (or estimate) the minimum number of subpaths of length  $k-1$  (or in general of length  $\ell$ ). There are substantial differences between this problem and the traditional Shadow Problem. Here the  $k$ -element subsets of edges are not necessarily paths of length  $k$ , in contrast to the case of the Shadow Problem. On the other hand, in the new problem we cannot delete edges arbitrarily from a path obtaining a shorter path.

The *Distance Shadow Problem* is formally very similar. Given  $f_k$ , the number of pairs of distinct vertices with distance  $k$ , determine the minimum number of distances  $k-1$  in a graph. Our present paper is devoted to these two problems.

## 2 Minimization of path and distance shadows

Given a graph  $G$  let  $f_k = f_k(G)$  be the number of pairs of distinct vertices  $(u, v)$  with distance  $d(u, v) = k$ . Our goal in this section is to find the minimum of  $f_\ell(G)$  for all graphs containing  $f_k$  pairs with distance  $k$  where  $k > \ell$ . Denote the minimum by  $M(f_k, k, \ell)$ . If  $n$ , the number of vertices is known then we can define the minimum  $M(f_k, n, k, \ell)$ . These quantities are the minimum sizes of the distance shadow. The following inequality is obvious.

$$M(f_k, k, \ell) \leq M(f_k, n, k, \ell). \quad (1)$$

But there is an another variant of this problem. The distance  $d(u, v)$  is defined by a (shortest) path. In our modified problem we consider all paths of length  $k$ , not only the ones which define the distance. Let  $N(f_k, k, \ell)$  denote the minimum number of paths of length  $\ell$  in a graph containing exactly  $f_k$  paths of length  $k$ . Similarly, let  $N(f_k, n, k, \ell)$  denote the same under the condition that the number of vertices of the graph is fixed:  $n$ . The following inequality is obvious, again.

$$N(f_k, k, \ell) \leq N(f_k, n, k, \ell). \quad (2)$$

These are the minimum sizes of the path shadows.

### 2.1 The case $\ell = 1$ , paths

The results in this subsection are taken from the literature. We list them for the sake of completeness.

Here the number of distances 1 and the number of paths of length 1, that is the number of edges should be minimized in both cases. It is easier to handle the following inverse problem. The number of edges in the graph is given, determine the maximum number of distances  $k$  or paths of length  $k$  in the graph, either fixing the number of vertices or not.

Let the inverse of  $N(f_k, k, 1)$  be  $R(e, k)$ . This is the maximum number of paths of length  $k$  in a graph containing  $e$  edges. Similarly, let  $R(n, e, k)$  denote the same under the condition that not only the number of edges, but also the number of vertices is given:  $e$  and  $n$ , respectively. It is easy to determine  $R(e, 2)$ .

**Proposition 1**  $R(e, 2) = \binom{e}{2}$ .

**Proof.** The star (one "central" vertex contained in all  $e$  edges) gives the above value, proving  $R(e, 2) \geq \binom{e}{2}$ . On the other hand a path of length two consists of two edges, therefore one cannot have more of them than  $\binom{e}{2}$ .  $\square$

It is more difficult to determine  $R(n, e, 2)$ . Paper [1] contains an almost complete solution. In order to be able to formulate its main statement some definitions are needed. A *quasi-click* consists of a click (complete graph) and an additional vertex adjacent to a subset of the vertices of the click. (If there is no additional vertex then the quasi-click is simply a click.) A *quasi-star* is the complement of a quasi-click.

**Theorem 1** [1]  $R(n, e, 2)$  is equal to the number of paths of length two either in a quasi-star or in a quasi-click with  $e$  edges. Moreover the extremal graph is a quasi-star if  $e \leq \frac{1}{2}\binom{n}{2} - \frac{n}{2}$  and a quasi-click if  $e \geq \frac{1}{2}\binom{n}{2} + \frac{n}{2}$ .

One can see from this special case that the determination of  $R(e, k)$  is easier than that of  $R(n, e, k)$ . However in the case of  $k = 3$  the difference is not so much.

**Theorem 2** [3] If  $10 \leq \binom{a}{2} \leq e < \binom{a+1}{2}$  holds then  $R(e, 3) \leq 2e(e - a)\frac{a-2}{a}$  with equality for the complete graph on  $a$  vertices.

Here we do not have the exact value for all  $e$ , but this theorem gives a very good asymptotic estimate for  $R(e, 3)$ . Moreover, the best (asymptotic) construction does not depend on  $n$ . Therefore the asymptotic solution (that is sharp for infinitely many  $e$ ) gives the asymptotically sharp solution for  $R(n, e, 3)$ , too.

Based on these results one can guess that the odd and even  $k$ 's behave differently.

**Theorem 3** ([2], see also [3]) If  $k$  is odd then both  $R(n, e, k)$  and  $R(e, k)$  are asymptotically equal to  $2^{\frac{k-1}{2}} e^{\frac{k+1}{2}}$  and the asymptotically sharp construction is a complete graph.

**Theorem 4** ([3]) If  $k$  is even then  $R(e, k)$  is asymptotically equal to  $c_k e^{\frac{k}{2}+1}$ , where  $c_k$  is a constant depending only on  $k$ .

If  $k = 4$ , then  $R(e, 4)$  is basically known.

**Theorem 5** ([4]) *If  $e$  is large enough then*

$$R(e, 4) = \begin{cases} \frac{e^3}{8} - \frac{3e^2}{4} + e & \text{if } e \text{ is even,} \\ \frac{e^3}{8} - \frac{7e^2}{8} + \frac{15e}{8} - \frac{9}{8} & \text{if } e \text{ is odd.} \end{cases}$$

*The extremal construction is a complete bipartite graph with sizes  $\lfloor \frac{e}{2} \rfloor$  and 2.*

Very little is known on  $R(n, e, k)$ , but  $R(n, e, 4)$  is asymptotically determined.

**Theorem 6** ([10])  *$R(n, e, 4)$  is asymptotically equal to the number of paths either in a quasi-star or in a quasi-click.*

## 2.2 The case $\ell = 1$ , distances

We consider the inverse again. The inverse of  $M(f_k, k, 1)$  is  $S(e, k)$ : this the maximum number of distances  $k$  in a graph with  $e$  edges, while  $S(n, e, k)$  is the maximum number of distances  $k$  in a graph with  $n$  vertices and  $e$  edges. Of course

$$S(e, k) \leq R(e, k) \text{ and } S(n, e, k) \leq R(n, e, k) \quad (3)$$

holds. The inequalities are usually very sharp, since, e.g. the number of distances is at most quadratic in  $n$ , while the number of paths can be much larger.

**Proposition 2**  $S(e, 2) = \binom{e}{2}$ .

**Proof.** The statement follows from Proposition 1, (3) and the fact that the star with  $e$  rays contains this many distances 2.  $\square$

**Proposition 3**

$$S(n, e, 2) = \begin{cases} \binom{e}{2} & \text{if } e \leq n - 1, \\ \binom{n}{2} - e & \text{if } n \leq e. \end{cases}$$

**Proof.** If  $e \leq n - 1$  then one cannot have more than  $\binom{e}{2}$  paths of length 2, therefore this is an upper bound for the number of distances, as well. The star of  $e$  edges gives equality. On the other hand, if  $n \leq e$  then the distance between the endpoints of an edge is one, cannot be two. Hence the total number of distances two cannot be more than  $\binom{n}{2} - e$ . The star with  $n - 1$  rays and  $e - (n - 1)$  other edges gives the equality.  $\square$

**Theorem 7** For fixed  $k$  and large  $n$   $S(e, k)$  is asymptotically equal to  $\binom{e}{2}$ .

**Proof.** Let us see that the number of finite distances in a graph with  $e$  edges cannot exceed  $\binom{e+1}{2}$ . If the graph is connected then the number of vertices is at most  $e + 1$ , therefore the number of distances is at most  $\binom{e+1}{2}$ . Suppose now that the numbers of edges in the  $r$  components are  $e_1, e_2, \dots, e_r$  where  $e_1 + e_2 + \dots + e_r = e$ . The  $i$ 's component has at most  $e_i + 1$  vertices therefore at most  $\binom{e_i+1}{2}$  edges. The total number edges is at most

$$\sum_{i=1}^r \binom{e_i + 1}{2}.$$

This is really at most  $\binom{e+1}{2}$  as the following lines show:

$$\begin{aligned} \sum_{i=1}^r \binom{e_i + 1}{2} &= \frac{1}{2} \sum_{i=1}^r e_i^2 + \frac{1}{2} \sum_{i=1}^r e_i \leq \frac{1}{2} \left( \sum_{i=1}^r e_i \right)^2 + \frac{1}{2} \sum_{i=1}^r e_i = \\ &= \binom{\sum_{i=1}^r e_i + 1}{2} = \binom{e + 1}{2}. \end{aligned}$$

Since  $\binom{e+1}{2}$  and  $\binom{e}{2}$  are asymptotically equal, the latter one is really an asymptotical upper bound.

Below we will construct asymptotically optimal graphs, distinguishing cases according to  $k$ .

Suppose first that  $k$  is even. Let  $T(t, k/2)$  be a *complete  $t$ -ary tree of depth  $k$* . This is a tree with a root of degree  $t$ , all other vertices have degrees  $t + 1$  and 1. The latter one are the leaves. The distance of the root and a leaf is  $k/2$ . The number of edges is

$$e = t + t^2 + \dots + t^{k/2} = t \frac{t^{k/2} - 1}{t - 1} \sim t^{k/2} \quad (4)$$

where it is supposed in the asymptotical calculation that  $k$  is fixed and  $t$  is large. Denote the vertices of distance 1 from the root by  $v_1, \dots, v_t$ , while the set of leaves connected to the root through  $v_i$  is denoted by  $L_i$ . It is easy to see that the distance of the vertices  $a \in L_i$  and  $b \in L_j$  ( $i \neq j$ ) in the tree  $T(t, k/2)$  is exactly  $k$ . Of course  $|L_i| = t^{k/2-1}$ . Hence the number of pairs of vertices with distance  $k$  is at least (in fact exactly)

$$\binom{t}{2} (t^{k/2-1})^2 \sim \frac{t^k}{2}.$$

Using (4) we really have that the number of distances  $k$  is asymptotically  $\frac{e^2}{2} \sim \binom{e}{2}$  if  $k$  is fixed and  $t$  is large. This proves the statement for even  $k$ .

Suppose now that  $k > 3$  is odd. Define a graph  $G(t, k)$  in the following way. It will consist of a complete graph  $K_t$  and  $t$  copies of the rooted tree  $T(t, (k-1)/2)$ , one copy attached to every vertex of  $K_t$ . The set of leaves of the copy attached to the  $i$ th vertex is denoted by  $L_i$ . The number of edges of  $G(t, k)$  is

$$\binom{t}{2} + t(t + \dots + t^{(k-1)/2}) \sim t^{(k+1)/2}. \quad (5)$$

The distance of the vertices  $a \in L_i$  and  $b \in L_j (i \neq j)$  in  $G(t, k)$  is exactly  $k$ . Of course  $|L_i| = t^{(k-1)/2}$ . Hence the number of pairs of vertices with distance  $k$  is at least (in fact exactly)

$$\binom{t}{2} (t^{(k-1)/2})^2 \sim \frac{t^{k+1}}{2}.$$

Using (5) we obtain that the number of distances is  $\frac{e^2}{2} \sim \binom{e}{2}$ , as stated.

In the case of  $k = 3$  the construction above should be slightly modified. Here we attach a star with  $t^2$  edges to each vertex of the complete graph  $K_t$ . The endpoints of different stars have distance 3. The number of edges is

$$e = \binom{t}{2} + t \cdot t^2 \sim t^3. \quad (6)$$

The number of pairs with distance 3 is

$$\binom{t}{2} t^2 \cdot t^2 \sim \frac{t^6}{2} \sim \frac{e^2}{2},$$

where (6) was used.

Our constructions shows the statement for an infinite sequence of values of  $e$ , namely for the values of form (4), (5) and (6) in the respective cases. This would be sufficient if we knew that  $S(e, k)$  is a monotone function of  $e$ . However it is absolutely non-trivial that adding an edge to a graph this increases the distances  $k$ . Here, however only the monotony of our construction is needed.

Suppose that  $k$  is even and

$$t + t^2 + \dots + t^{k/2} < e < (t+1) + (t+1)^2 + \dots + (t+1)^{k/2}.$$

Add

$$e - (t + t^2 + \dots + t^{k/2})$$

edges to  $T(t, k/2)$  without forming a cycle. No distance  $k$  can be destroyed in this way in  $T(t, k/2)$ . Therefore the number of distances  $k$  in this new graph is at least as much as in  $T(t, k/2)$ . This proves that the number of distances  $k$  can asymptotically be  $\frac{t^k}{2}$  when  $e$  is asymptotically at most  $(t+1)^{k/2} \sim t^{k/2}$ .

The other cases of  $k$  can be similarly settled.  $\square$

If  $k$  is very close to  $e$  then there is a hope to get the exact value of  $S(e, k)$ .

**Conjecture 1** *If  $\ell \geq 0$  is fixed,  $e$  is large enough then  $S(e, e - \ell) = \lfloor \frac{\ell}{2} \rfloor \lceil \frac{\ell}{2} \rceil$ .*

The construction giving this value is two stars with  $\lfloor \frac{\ell}{2} \rfloor$  and  $\lceil \frac{\ell}{2} \rceil$  edges connected with a path of  $e - \ell$  edges. One easily prove the conjecture for  $\ell = 0, 1, 2, 3$ .

### 2.3 The case $\ell > 1$ , paths

Here we consider the inverse problem, again. Let  $U(p, \ell, k)$  ( $\ell < k$ ) be the largest number of paths of length  $k$  in a graph containing  $p$  paths of length  $\ell$ . The most important case is  $\ell = k - 1$ . Having a good upper estimate on  $U(p, k - 1, k)$  (and  $U(p, k - 2, k - 1)$ ) we get one for  $U(p, k - 2, k)$ , and so on. Make our notation shorter:  $U(p, k) = U(p, k - 1, k)$ .

**Proposition 4**  *$U(p, k)$  is asymptotically lowerbounded by  $2^{\frac{1}{k}} p^{\frac{k+1}{k}}$ .*

**Proof.** The complete graph  $K_n$  serves as a construction. The number of paths of length  $k - 1$  is  $p = \frac{n(n-1)\dots(n-k+1)}{2}$  (choosing an ordered sequence of vertices of length  $k$ , we have to divide the product by 2 because every path is counted starting from both ends). Hence  $p \sim \frac{n^k}{2}$  that is  $n \sim (2p)^{\frac{1}{k}}$ . The number of paths of length  $k$  is  $\frac{n(n-1)\dots(n-k)}{2} \sim \frac{n^{k+1}}{2} \sim \frac{(2p)^{\frac{k+1}{k}}}{2} = 2^{\frac{1}{k}} p^{\frac{k+1}{k}}$ .  $\square$

Observe that this upper bound is too weak if  $k = 2$ . It gives  $\sqrt{2}p^{\frac{3}{2}}$  while we know by Proposition 1 that  $U(p, 2) = R(e, 2) = \binom{e}{2} = \binom{p}{2}$ . However we believe that the bound is asymptotically sharp starting from  $k = 3$ .

**Conjecture 2**  $U(p, 3) \sim 2^{\frac{1}{3}} p^{\frac{4}{3}}$ .

**Theorem 8** *Conjecture 1 is true for regular graphs.*

**Proof.** Let  $n$  denote the number of vertices of the regular graph of degree  $d$ . The number of paths of length 2 with middle vertex at a given vertex is  $\binom{d}{2}$ . The total number of paths of length 2 is

$$p = n \binom{d}{2}. \quad (7)$$

The number of paths of length 3 with a fixed edge in the middle is  $(d-1)^2$ , since we have  $(d-1)$  ways to continue the path at each end of the fixed edge. The total number of paths of length 3 is obtained by multiplying this with the number of edges. The number of edges is  $\frac{nd}{2}$ . Hence the number of paths of length 3 is

$$\frac{nd}{2}(d-1)^2. \quad (8)$$

We know by Proposition 4 that  $2^{\frac{1}{3}}p^{\frac{4}{3}}$  is an asymptotic lower bound. Here we will prove that it is also an upper bound for (8) that is for the number of paths of length 3:

$$\frac{nd}{2}(d-1)^2 \leq 2^{\frac{1}{3}}p^{\frac{4}{3}}. \quad (9)$$

Use (7) on the right hand side.

$$\frac{nd}{2}(d-1)^2 \leq 2^{\frac{1}{3}}p^{\frac{4}{3}} = 2^{\frac{1}{3}} \left( n \binom{d}{2} \right)^{\frac{4}{3}}.$$

Writing out the details this becomes

$$nd(d-1)^2 \leq n^{\frac{4}{3}}d^{\frac{4}{3}}(d-1)^{\frac{4}{3}}.$$

Divide both sides by  $nd(d-1)^{\frac{4}{3}}$ . We have arrived to

$$(d-1)^{\frac{2}{3}} \leq n^{\frac{1}{3}}d^{\frac{1}{3}}.$$

This is true, since the degree cannot exceed the number of vertices. (9) and the theorem are proved.  $\square$

## 2.4 The case $\ell > 1$ , distances

Let  $T(p, \ell, k)$  ( $\ell < k$ ) be the largest number of distances of length  $k$  in a graph containing  $p$  distances of length  $\ell$ . The most important case is  $\ell = k - 1$ .

Having a good upper estimate on  $T(p, k - 1, k)$  (and  $T(p, k - 2, k - 1)$ ) we get one for  $T(p, k - 2, k)$ , and so on. Make our notation shorter:  $T(p, k) = T(p, k - 1, k)$ .

Of course  $T(p, 2) = S(p, 2)$  and Proposition 2 gives the exact solution. The number of distances 2 is a quadratic function of the number of distances 1 (edges). We believe that this cannot happen for larger  $k$ 's. Actually, it is difficult to construct a graph in which the number of distances 3 is much larger than the number of distances 2.

**Theorem 9**

$$\frac{1}{36}(1 - o(1))p \log p \leq T(p, 3)$$

where  $\log$  means logarithm of basis 2.

The proof will be divided into lemmas.

Let the vertices of a graph  $G^{d\Box}$  be the sequences of length  $d$  formed from the elements  $-3, -2, -1, 0, 1, 2, 3$ . Two vertices are adjacent if the corresponding sequences are equal in all but one place where the difference of the values is 1 or -1. This is actually a  $7 \times 7 \times \dots \times 7$  "grid" containing  $7^d$  vertices. Let  $n(d, t)$  denote the number of vertices having distance  $t$  from the origin  $(0, 0, \dots, 0)$ . The case  $t = 1$  is trivial.

$$n(d, 1) = 2d. \tag{10}$$

**Lemma 1**  $n(d, 2) = 2d^2$ .

**Proof.** Let  $H$  denote the set of vertices of distance 2 from the origin in  $G^{(d+1)\Box}$ . Divide  $H$  into subsets according to their first coordinates:  $H(-2), H(-1), H(0), H(1), H(2)$ . Then

$$|H| = |H(-2)| + |H(-1)| + |H(0)| + |H(1)| + |H(2)|. \tag{11}$$

The elements of  $H(0)$  have distance 2 in the rest, that is in  $G^{d\Box}$ . Hence we have

$$|H(0)| = n(d, 2). \tag{12}$$

The elements of  $H(1)$  have a value 1 in the first coordinate. Therefore their rest have distance 1 from the origin. Hence we have

$$|H(1)| = n(d, 1). \tag{13}$$

The same holds for  $H(-1)$ :

$$|H(-1)| = n(d, 1). \quad (14)$$

Finally, if the first coordinate is 2 or  $-2$  then the other  $d$  coordinates must be all 0. Hence we have

$$|H(2)| = |H(-2)| = 1. \quad (15)$$

The equations (10)-(15) give

$$n(d+1, 2) = n(d, 2) + 4d + 2. \quad (16)$$

Now use induction on  $d$  to prove the statement of the lemma. For  $d = 1$ , it is trivial that there are two vertices of distance 2 from the point 0. Suppose now that the statement holds for  $d$  and prove it for  $d + 1$ . By (16) and the inductual hypothesis we can write

$$n(d+1, 2) = 2d^2 + 4d + 2 = 2(d+1)^2.$$

□

**Lemma 2**  $n(d, 3) = \frac{4}{3}d^3 + \frac{2}{3}d$ .

**Proof.** The logic of the previous proof will be used again. Let  $H$  denote the set of vertices of distance 3 from the origin in  $G^{(d+1)\square}$ . Divide  $H$  into subsets according to their first coordinates:  $H(-3), H(-2), H(-1), H(0), H(1), H(2), H(3)$ . Then

$$|H| = |H(-3)| + |H(-2)| + |H(-1)| + |H(0)| + |H(1)| + |H(2)| + |H(3)|. \quad (17)$$

The elements of  $H(0)$  have distance 3 in the rest, that is in  $G^{d\square}$ . Hence we have

$$|H(0)| = n(d, 3). \quad (18)$$

The elements of  $H(1)$  have a value 1 in the first coordinate. Therefore their rest have distance 2 from the origin. Hence we have

$$|H(1)| = n(d, 2). \quad (19)$$

The same holds for  $H(-1)$ :

$$|H(-1)| = n(d, 2). \quad (20)$$

If the first coordinate of a vertex is 2 then the rest must have distance 1 from the origin. This proves

$$|H(2)| = n(d, 1) \quad (21)$$

and

$$|H(-2)| = n(d, 1). \quad (22)$$

Finally, if the first coordinate is 3 or  $-3$  then the other  $d$  coordinates must be all 0. Hence we have

$$|H(3)| = |H(-3)| = 1. \quad (23)$$

The equations (17)-(23) give

$$n(d+1, 3) = n(d, 3) + 2n(d, 2) + 4d + 2 \quad (24)$$

and, using the statement of Lemma 1,

$$n(d+1, 3) = n(d, 3) + 4d^2 + 4d + 2. \quad (25)$$

Now use induction on  $d$  to prove the statement of the lemma. For  $d = 1$ , it is trivial that there are two vertices of distance 3 from the point 0. Suppose now that the statement holds for  $d$  and prove it for  $d + 1$ . By (25) and the inductional hypothesis we can write

$$n(d+1, 3) = \frac{4}{3}d^3 + \frac{2}{3}d + 4d^2 + 4d + 2 = \frac{4}{3}(d+1)^3 + \frac{2}{3}(d+1).$$

□

**Proof of Theorem 9.** Modify the graph  $G^{d\Box}$  at the beginning of our proof. Let the vertices of the graph  $Z_8^{d\Box}$  be the sequences of length  $d$  formed from the elements  $-3, -2, -1, 0, 1, 2, 3, 4$  and consider these integers modulo 8. Two vertices are adjacent if the corresponding sequences are equal in all but one place where the difference of the values is 1 or  $-1 \pmod{8}$ . This is actually a  $8 \times 8 \times \dots \times 8$  "cyclic grid" containing  $8^d$  vertices. The *3-neighborhood* of a vertex  $v$  in a graph is the subgraph spanned by the set of vertices of distance at most 3 from  $v$ . Observe that the 3-neighborhoods of the origins in  $G^{d\Box}$  and  $Z_8^{d\Box}$  are isomorphic. Moreover the 3-neighborhoods of two distinct vertices in  $Z_8^{d\Box}$  are also isomorphic. Hence it follows that the number of vertices with distance 3 (resp. 2) from a vertex of  $Z_8^{d\Box}$  is the same as the number of vertices with distance 3 (resp. 2) from the origin of  $G^{d\Box}$ .

Summarising, the number of vertices with distance 3 (resp. 2) from a vertex of  $Z_8^{d\Box}$  is  $n(d, 3)$  (resp.  $n(d, 2)$ ). Therefore the total number of distances 2 in  $Z_8^{d\Box}$

$$p = \frac{1}{2}8^d n(d, 2).$$

By Lemma 1 this is equal to

$$p = \frac{1}{2}8^d \cdot 2d^2 = 8^d \cdot d^2. \quad (26)$$

Similarly, the total number of distances 3 in  $Z_8^{d\Box}$

$$\frac{1}{2}8^d n(d, 3)$$

and by Lemma 2 this is equal to

$$8^d \left( \frac{2}{3}d^3 + \frac{1}{3}d \right). \quad (27)$$

Here (26) and (27) lead to

$$8^d \left( \frac{2}{3}d^3 + \frac{1}{3}d \right) \leq T(8^d \cdot d^2, 3). \quad (28)$$

Equation (26) gives us

$$p \log p = 8^d d^2 (3d + 2 \log d).$$

Compare this with the left hand side of (28). One can see that

$$\frac{2}{9}(1 - o(1))p \log p = \frac{2}{9}(1 - o(1))8^d d^2 (3d + 2 \log d) \leq 8^d \left( \frac{2}{3}d^3 + \frac{1}{3}d \right). \quad (29)$$

The inequalities (28) and (29) prove the statement of the theorem in a stronger form, but only for some special values of  $p$ , which form an exponential sequence.

Let us prove now the weaker inequality for the intermediate values of  $p$ . Suppose

$$p_1 = 8^d d^2 < p < 8^{d+1} (d+1)^2 = p_2. \quad (30)$$

Add  $p - 8^d d^2$  paths of length 2 to  $Z_8^{d\Box}$  in such a way that they do not form any cycle neither with  $Z_8^{d\Box}$  nor with each other. The so obtained graph  $G$  contains  $p$  pairs with distance 2 and at least (27) pairs with distance 3.

Observe that

$$\frac{1}{8}(1 - o(1))p_2 \log p_2 = p_1 \log p_1. \quad (31)$$

Formulas (30) and (31) imply

$$\frac{1}{36}(1 - o(1))p \log p \leq \frac{2}{9}(1 - o(1))p_1 \log p_1.$$

Combining this with (29) the desired inequality is obtained.  $\square$

**Open Problem 1** Find a non-trivial upper bound on  $T(p, 3)$ .

## References

- [1] R. Ahlswede, G.O.H. Katona, Graphs with maximal number of adjacent pairs of edges, *Acta Math. Acad. Sci. Hungar.* **32**(1978) 097-120.
- [2] N. Alon, On the number of subgraphs of prescribed type of graphs with a given number of edges, *Israel J. Math.* **38**(1981) 116-130.
- [3] B. Bollobás and A. Sarkar, Paths in graphs, *Studia Sci. Math. Hungar.* **38**(2001) 115-137.
- [4] B. Bollobás and A. Sarkar, Paths of length four, *Discrete Math.* **265**(2003) 357-363.
- [5] Peter L. Erdős, P. Frankl and G.O.H. Katona, Extremal hypergraph problems and convex hulls, *Combinatorica* **5**(1985) 11-26.
- [6] Hongbo Hua, Morteza Faghani, Ali Reza Ashrafi, The Wiener and Wiener Polarity Indices of a Class of Fullerenes with Exactly  $12n$  Carbon Atoms, *MATCH Commun. Math. Comput. Chem.* **71**(2014) 361-372.
- [7] Dušanka Janežić, Ante Miličević, Sonja Nikolić, Nenad Trinajstić, *Graph-Theoretical Matrices in Chemistry*, CRC Press, 2015.
- [8] G. Katona, A theorem on finite sets, *Theory of Graphs*, Proc. Coll. held at Tihany, 1966, Akadémiai Kiadó, pp. 187-207.

- [9] J.B. Kruskal, The number of simplices in a complex, *Mathematical Optimization Techniques*, University of Calif. Press, (Berkeley and Los Angeles, 1963), pp. 251-278.
- [10] Dániel T. Nagy, On the Number of 4-edge Paths in Graphs With Given Edge Dencity, *Combinatorics, Probability and Computing*, appeared online: DOI: <https://doi.org/10.1017/S0963548316000389>, see also in arXiv:1601.01211 [math.CO].
- [11] Wiener, H., Structural determination of paraffin boiling points, *Journal of the American Chemical Society* **1**(1947)(69): 1720, doi:10.1021/ja01193a005.