

# On the maximum diameter of $k$ -colorable graphs

Éva Czabarka

Department of Mathematics  
University of South Carolina  
South Carolina, U.S.A.

Visiting Prof.

Department of Pure and Applied Mathematics  
University of Johannesburg  
South Africa

czabarka@math.sc.edu

Inne Singgih\*

Department of Mathematical Sciences  
University of Cincinnati  
Ohio, U.S.A.

singgiie@ucmail.uc.edu

László A. Székely†

Department of Mathematics  
University of South Carolina  
South Carolina, U.S.A.

Visiting Prof.

Department of Pure and Applied Mathematics  
University of Johannesburg  
South Africa

szekely@math.sc.edu

Submitted: Apr 20, 2021; Accepted: Aug 24, 2021; Published: Sep 10, 2021

© The authors. Released under the CC BY-ND license (International 4.0).

## Abstract

We show that the diameter of connected  $k$ -colorable graphs with minimum degree  $\geq \delta$  and order  $n$  is at most  $\left(3 - \frac{1}{k-1}\right) \frac{n}{\delta} - 1$ , while for  $k = 3$ , it is at most  $\frac{57n}{23\delta} + O(1)$ .

**Mathematics Subject Classifications:** 05C35, 05C15, 05C12

## 1 Introduction

This paper is concerned with the maximum diameter of connected graphs, namely how it depends on the order and the minimum degree, and possibly on further graph properties.

---

\*Supported in part by the National Science Foundation contract DMS 1600811.

†Supported in part by National Science Foundation contract DMS 1600811.

The history of this problem goes back to the following theorem, which was discovered multiple times [1, 6, 8, 9]:

**Theorem 1.** For a fixed minimum degree  $\delta \geq 2$  and  $n \rightarrow \infty$ , for every  $n$ -vertex connected graph  $G$ , we have  $\text{diam}(G) \leq \frac{3n}{\delta+1} + O(1)$ .

Note that Theorem 1 is sharp (even for  $\delta$ -regular graphs [2]), but all the known constructions have complete subgraphs, whose order increases with  $\delta$ . Erdős, Pach, Pollack, and Tuza [6] conjectured that the upper bound of Theorem 1 can be improved, if we restrict the clique size of the graph:

**Conjecture 2** ([6]). Let  $r, \delta \geq 2$  be fixed integers and let  $G$  be a connected graph of order  $n$  and minimum degree  $\delta$ .

(i) If  $G$  is  $K_{2r}$ -free and  $\delta$  is a multiple of  $(r-1)(3r+2)$  then, as  $n \rightarrow \infty$ ,

$$\begin{aligned} \text{diam}(G) &\leq \frac{2(r-1)(3r+2)}{(2r^2-1)} \cdot \frac{n}{\delta} + O(1) \\ &= \left(3 - \frac{2}{2r-1} - \frac{1}{(2r-1)(2r^2-1)}\right) \frac{n}{\delta} + O(1). \end{aligned}$$

(ii) If  $G$  is  $K_{2r+1}$ -free and  $\delta$  is a multiple of  $3r-1$ , then, as  $n \rightarrow \infty$ ,

$$\text{diam}(G) \leq \frac{3r-1}{r} \cdot \frac{n}{\delta} + O(1) = \left(3 - \frac{2}{2r}\right) \frac{n}{\delta} + O(1).$$

Furthermore, they created examples showing that if Conjecture 2 is true, then it is sharp.

Part (ii) of Conjecture 2 for  $r=1$  was already proved in Erdős *et al.* [6].

Czabarka, Dankelmann and Székely [3] proved a weaker version of Conjecture 2 (ii) for  $r=2$ :

**Theorem 3.** For every connected 4-colorable graph  $G$  of order  $n$  and minimum degree  $\delta \geq 1$ ,  $\text{diam}(G) \leq \frac{5n}{2\delta} - 1$ .

In [4] we gave an unexpected counterexample for Conjecture 2 (i) for every  $r \geq 2$  and  $\delta > 2(r-1)(3r+2)(2r-3)$ . The question whether Conjecture 2 (i) holds in the range  $(r-1)(3r+2) \leq \delta \leq 2(r-1)(3r+2)(2r-3)$  remains open. The counterexample led to the following modification of Conjecture 2, which no longer requires cases for the parity of the order of the excluded complete subgraph:

**Conjecture 4** ([4]). For every  $k \geq 3$  and  $\delta \geq \lceil \frac{3k}{2} \rceil - 1$ , if  $G$  is a  $K_{k+1}$ -free (in a weaker version of the conjecture:  $k$ -colorable) connected graph of order  $n$  and minimum degree at least  $\delta$ ,  $\text{diam}(G) \leq \left(3 - \frac{2}{k}\right) \frac{n}{\delta} + O(1)$  as  $n \rightarrow \infty$ .

In this paper, we follow the restrictive approach of Czabarka, Dankelmann and Székely [3] and work on the weaker version of Conjecture 4. In other words, we use a stronger hypothesis ( $k$ -colorable instead of  $K_{k+1}$ -free) than what Erdős, Pach, Pollack, and Tuza [6] used. In our work on upper bounds on the diameter, we only assume *minimum degree at least  $\delta$* , a weaker assumption than *minimum degree  $\delta$* . Section 2 shows that *some  $k$ -colorable* (in particular 3-colorable) connected graphs realizing the maximum diameter among such graphs with given order and minimum degree have certain *canonical properties*. Hence at proving upper bounds on the diameter, we can assume that those canonical properties hold.

Section 3 gives a linear programming duality approach to the maximum diameter problem. With this approach, proving upper bounds to the diameter boils down to solving a packing problem in a graph, such that a certain value is reached by the objective function. (If a packing with that value is given, the task of checking whether the packing is feasible is trivial.) Using this duality approach for canonical connected  $k$ -colorable graphs, we obtain

**Theorem 5.** *Assume  $k \geq 3$ . If  $G$  is a connected  $k$ -colorable graph of minimum degree at least  $\delta \geq 1$ , then*

$$\text{diam}(G) \leq \frac{3k-4}{k-1} \cdot \frac{n}{\delta} - 1 = \left(3 - \frac{1}{k-1}\right) \frac{n}{\delta} - 1.$$

This corroborates the conjecture of Erdős *et al.* in the sense that the maximum diameter among all graphs investigated in Theorem 5 is  $(3 - \Theta(\frac{1}{k})) \frac{n}{\delta}$ . A corollary of Theorem 5 is the conclusion of Theorem 3, if the graph in question is *3-colorable* (instead of *4-colorable*).

Section 4 applies the inclusion-exclusion (sieve) formula to give upper bounds *locally* for the number of vertices in connected 3-colorable graphs with the *canonical* properties. In Section 5, we define a number of global variables that play a role in the diameter problem, and turn the upper bounds from Section 4 into linear constraints for the global variables. (This approach was motivated by the flag algebra method of Razborov [11].) A linear program of fixed size for the global variables arises, and solving this linear program proves our main positive result:

**Theorem 6.** *For every connected 3-colorable graph  $G$  of order  $n$  and minimum degree at least  $\delta \geq 1$ ,*

$$\text{diam}(G) \leq \frac{57n}{23\delta} + O(1),$$

where the  $O(1)$  term may depend on  $\delta$  but not on  $n$ .

Note that as  $57/23 \approx 2.47826\dots$ , this is an improvement on the  $\frac{5}{2} \cdot \frac{n}{\delta} - 1$  upper bound for 4-colorable graphs of large order in Theorem 3. In Theorem 12, in a *restricted case* that does not include the likely optimal construction in [4], we prove the weaker version of Conjecture 4 for  $k = 3$ .

## 2 Clump Graphs

Let us be given a  $k$ -colorable connected graph  $G$  of order  $n$  and minimum degree at least  $\delta$ . Recall that the eccentricity of a vertex is the largest distance measurable from the vertex in the graph. Let  $x$  be a vertex whose eccentricity is  $\text{diam}(G)$ .

Take a *fixed* good  $k$ -coloring of  $G$ . Let *layer*  $L_i$  denote the set of vertices at distance  $i$  from  $x$ , and a *clump* in  $L_i$  be the set of vertices in  $L_i$  that have the same color. The number of layers is  $\text{diam}(G) + 1$ .

Let  $c(i) \in \{1, 2, \dots, k\}$  denote the number of colors used in layer  $L_i$  by our fixed coloring. We can assume without loss of generality that in  $G$ , two vertices in layer  $L_i$ , which are differently colored, are joined by an edge in  $G$ , and also that two vertices in consecutive layers, which are differently colored, are also joined by an edge in  $G$ . We call this assumption *saturation*. Assuming saturation does not make loss of generality, as adding these edges does not decrease degrees, keeps the fixed good  $k$ -coloring, and does not reduce the diameter, while it makes the graph more structured for our convenience.

From the layered and saturated graph  $G$  above, we create a (*weighted*) *clump graph*  $H$ . Vertices of  $H$  correspond to the clumps of  $G$ . Two vertices of  $H$  are connected by an edge if there is at least one edge between the corresponding clumps in  $G$ .  $H$  is naturally  $k$ -colored and layered, based on the coloring and layering of  $G$ . With a slight abuse of notation, we denote the layers of  $H$  by  $L_i$  as well. We assign as *weights* to each vertex of the clump graph the number of vertices in the corresponding clump in  $G$ .

Given a (positive integer)-weighted graph  $H$ , it defines a graph  $G$ , whose weighted clump graph is  $H$ , by blowing up vertices of  $H$  into as many copies as their weight is. The degree of a vertex  $v$  in  $G$  is the sum of the weights of neighbors of the vertex  $u$  in  $H$ , if  $v$  was created by blowing up  $u$ ;  $\text{diam}(G) = \text{diam}(H)$ ; and the number of vertices in  $G$  is the sum of the weights of all vertices in  $H$ . We use the letters  $X, Y, Z$  to denote three unspecified but different colors from our  $k$  colors.

**Theorem 7.** *Assume  $k \geq 3$ . Let  $G'$  be a  $k$ -colorable connected graph of order  $n$ , diameter  $D$  and minimum degree at least  $\delta$ . Then there is a  $k$ -colored connected graph  $G$  of the same parameters, with layers  $L_0, \dots, L_D$ , which is saturated, and for which the following statements hold for every  $i$  ( $0 \leq i \leq D - 1$ ):*

- (i) *If  $c(i) = 1$ , then  $c(i + 1) \leq k - 1$ .*
- (ii) *The number of colors used to color the set  $L_i \cup L_{i+1}$  is  $\min(k, c(i) + c(i + 1))$ . In particular, when  $c(i) + c(i + 1) \leq k$ , then  $L_i$  and  $L_{i+1}$  do not share any color.*
- (iii) *If  $c(i) = k$ , then  $i \geq 2$  and  $c(i + 1) \geq 2$ .*
- (iv) *If  $|L_i| > c(i)$ , i.e.,  $L_i$  contains two vertices of the same color, then  $i > 0$  and  $c(i) + \max(c(i - 1), c(i + 1)) \geq k$ .*

*Proof.* After having proved a part of the Theorem 7, we will assume that  $G'$  itself already satisfies that property when we complete the proof of the remaining statements. When

we create new  $G'$  graphs, they will still satisfy the already checked parts, in other words, we do not regress to issues that we already resolved. We fix a  $k$ -coloring of  $G'$ , let  $x_0$  be a vertex of eccentricity  $D$  in  $G'$ , and let  $L_0, \dots, L_D$  be the distance layering of  $G'$ . Without loss of generality, we assume that  $G'$  is saturated.

(i) Select  $G = G'$  with the same  $k$ -coloring, The statement follows from the fact that every vertex in  $L_{i+1}$  has a neighbor in  $L_i$ ; therefore if color  $X$  appears in  $L_{i+1}$ , then  $L_i$  has at least one color different from  $X$ .

If (ii) or (iv) is not satisfied in  $G'$ , our general strategy is the following: create a new  $k$ -coloring of the vertices of  $G'$  such that the set of the vertices in any clump of any layer does not change, vertices of different color will remain differently colored, and in the new coloring the already proven statements still hold. We saturate  $G'$  in the new coloring (by adding new edges, if needed) to obtain  $G$ . Now we complete this strategy for (ii), and postpone the proof of (iv) till the end.

If (ii) fails in  $G'$ , consider the smallest  $i$ , such that the set  $L_i \cup L_{i+1}$  contains fewer than  $\min(k, c(i) + c(i + 1))$  colors. As  $L_1$  has a single vertex to which all vertices of  $L_2$  is adjacent to,  $i > 0$ . Observe that there are different colors  $X, Y$  such that  $X$  is used in both of  $L_i$  and  $L_{i+1}$ , while  $Y$  is not used in  $L_i \cup L_{i+1}$ . We define a new coloring by switching colors  $X$  and  $Y$  in all  $L_j$  for all  $j \geq i + 1$ . This is a good coloring, in which  $L_i \cup L_{i+1}$  uses one more color. Repeated application of this procedure yields a  $k$ -coloring where (ii) holds.

The hard part of this theorem is (iii). If  $c(i) = k$ , then by (i)  $i \geq 2$ . If  $c(i + 1) = 1$  (i.e., if (iii) fails), then we will move clumps within  $L_{i-1} \cup L_i$ , and recolor the graph, such that the resulting layered colored graph will have the same required parameters as  $G'$ , creating no violations of (i), (ii), and reducing the number of violations of (iii) in  $G'$ .

Let  $X$  be the color used on  $L_{i+1}$ .

Assume first that  $L_{i-2}$  contains a color different from  $X$ . Let  $S$  be the set of vertices in  $L_i$  that is colored  $X$ . Move the vertices of  $S$  from  $L_i$  to  $L_{i-1}$  without recoloring them, either merging them into the  $X$ -colored clump of  $L_{i-1}$  or creating one, if no such clump existed in  $L_{i-1}$ . Add new edges to achieve saturation. In the resulting graph, the layer indexed by  $i$  contains  $k - 1$  colors, reducing the number of violations of (iii) in  $G'$ , and not creating any violation of (i) or (ii).

Hence in the following we may assume that  $c(i - 2) = 1$  and  $L_{i-2}$  is colored with color  $X$ . By (i), we have  $c(i - 1) \leq k - 1$ . If  $c(i - 1) < k - 1$ , then there is a color  $Y$  not used in  $L_{i-2} \cup L_{i-1}$ . Recolor  $G'$  by switching colors  $X$  and  $Y$  in  $L_j$  ( $0 \leq j \leq i - 2$ ) and recover saturation. In the new coloring  $L_{i-2}$  has a color different from  $X$ , and we are back to the case we already handled above.

Hence in the rest we may assume that  $c(i - 2) = c(i + 1) = 1$ ,  $c(i) = k$ ,  $c(i - 1) = k - 1$ , and both  $L_{i-2}$  and  $L_{i+1}$  are colored with  $X$ . Let  $Y, Z$  be two arbitrary colors different from  $X$ , and  $S$  be the set of vertices in  $L_{i-1} \cup L_i$  colored with  $X, Y$  or  $Z$ . We will repartition and recolor (only with colors  $X, Y, Z$ ) the vertices in  $S$ , and possibly recolor  $L_{i+1}$  from color  $X$  to color  $Y$ . If we recolor  $L_{i+1}$ , then we exchange the colors  $X$  and  $Y$  in all layers  $L_j$  for  $j \geq i + 2$ . After these steps, we recover saturation in  $G'$ . After the changes, in  $G'$  both  $L_{i-1}$  and  $L_i$  will contain fewer than  $k$  colors, and in the resulting  $k$ -colored graph

the diameter, the order and minimum degree condition do not change, and no instances violating (i) and (ii) will be created, and we reduced the number of violations of (iii). The difficulty is in maintaining the minimum degree condition in  $G'$  along these operations. This is what we check next, and the repartitioning and recoloring of the vertices in  $S$  will depend on some inequalities between certain clump sizes.

If  $y$  is a vertex not in  $L_{i-2} \cup L_{i-1} \cup L_i \cup L_{i+1}$  or  $y$  is not colored with  $X, Y$  or  $Z$  in the graph before the operations, then the neighborhood set of  $y$  does not change. If  $y$  is a vertex in  $L_{i-2} \cup L_{i-1} \cup L_i \cup L_{i+1}$  colored with one of  $X, Y, Z$ , then the symmetric difference between the new and old neighborhood set of  $Y$  is a subset of  $S$ . Therefore we only need to check the minimum degree condition for vertices colored  $X, Y$  or  $Z$  in  $L_{i-2} \cup L_{i-1} \cup L_i \cup L_{i+1}$ , and we have to show that after the operations they have at least as many  $X, Y, Z$  colored neighbors in  $L_{i-2} \cup L_{i-1} \cup L_i \cup L_{i+1}$  as before the operations. For any  $j$  ( $1 \leq j \leq 4$ ), we will denote by  $x_j, y_j$  and  $z_j$  the number of vertices in  $L_{i+j-3}$  colored  $X, Y$  and  $Z$  in  $G'$  respectively, *before the operations*. The  $k \geq 3$  assumption, together with the fact that  $L_{i-1}$  has no color  $X$  by (ii), implies that  $x_1, y_2, z_2, y_3, z_3$  and  $x_4$  are positive.

We have several cases to consider. Initially we have  $|L_{i-2}| = x_1, |L_{i-1}| = y_2 + z_2, |L_i| = x_3 + y_3 + z_3$  and  $L_{i+1} = x_4$ . We will recolor the vertices in the single clump in  $L_{i+1}$  from color  $X$  to color  $Y$  and switch colors  $X$  and  $Y$  in all  $L_j$  for  $j > I + 1$ , change  $L_i$  and  $L_{i-1}$  such that we still have  $|L_{i-1} \cup L_i| = |L_{i-1}| + |L_i| = y_2 + z_2 + x_3 + y_3 + z_3$  by merging some of their clumps into a single clump and adding edges so saturation is maintained.  $L_{i-1}$  will have at most 2 colors,  $Y$  and  $Z$ , and  $L_i$  will have at most 2 colors,  $X$  and  $Y$ . This, it is enough to describe the clumps and their sizes after the change in  $L_{i-1}$  and in  $L_i$ .

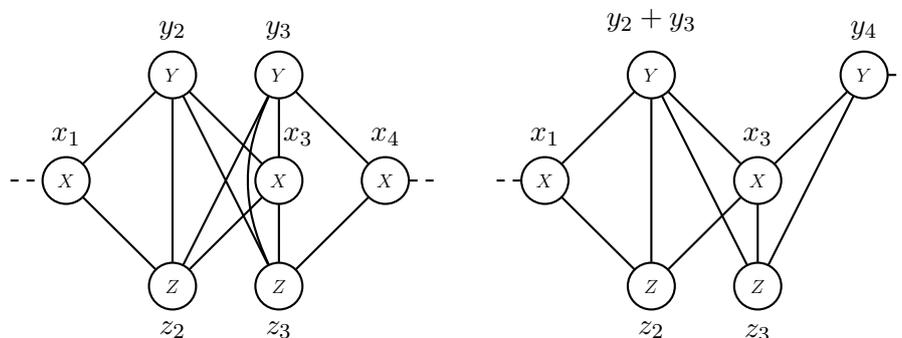


Figure 1: When  $x_3 \geq y_3$ , before and after the operations (left and right).

**1**  $x_3 \geq y_3$  or  $x_3 \geq z_3$ . (See Figure 1)

It suffices to handle the case  $x_3 \geq y_3$ , as the case  $x_3 \geq z_3$  can be handled similarly. After the change we have in  $L_{i-1}$  a clump of size  $y_2 + y_3$  of color  $Y$ , and a clump of size  $z_2$  of color  $Z$  in  $L_{i-1}$ ; also we have in  $L_i$  a clump of size  $x_3$  of color  $X$  and a clump of size  $z_3$  of color  $Z$ .

Note that, as claimed,  $|L_{i-1} \cup L_i|$  did not change. We verify the minimum degree

condition. Let  $d(W_i)$  to denote the number of neighbors of a vertex  $w$  from the clump colored  $W$  in layer  $L_i$  among the  $X, Y, Z$  colored vertices of  $L_{i-2} \cup L_{i-1} \cup L_i \cup L_{i+1}$  before the operations, and  $d'(W_i)$  to denote the degree of a vertex  $w'$  from the clump colored  $W$  in layer  $L_i$  among the  $X, Y, Z$  colored vertices of  $L_{i-2} \cup L_{i-1} \cup L_i \cup L_{i+1}$  after the operations. We have:

$$\begin{aligned} d'(X_{i-2}) &= d(X_{i-2}) + y_3 > d(X_{i-2}), \\ d'(Y_{i-1}) &= x_1 + z_2 + x_3 + z_3 = d(Y_{i-1}), \\ d'(Z_{i-1}) &= x_1 + y_2 + z_3 + y_3 = d(Z_{i-1}), \\ d'(X_i) &= y_2 + y_3 + z_2 + z_3 + x_4 > y_2 + y_3 + z_2 + z_3 = d(X_i), \\ d'(Z_i) &= y_2 + y_3 + x_3 + x_4 = d(Z_i), \\ d'(Y_{i+1}) &= x_3 + z_3 = d(X_{i+1}) + (x_3 - y_3) \geq d(X_{i+1}). \end{aligned}$$

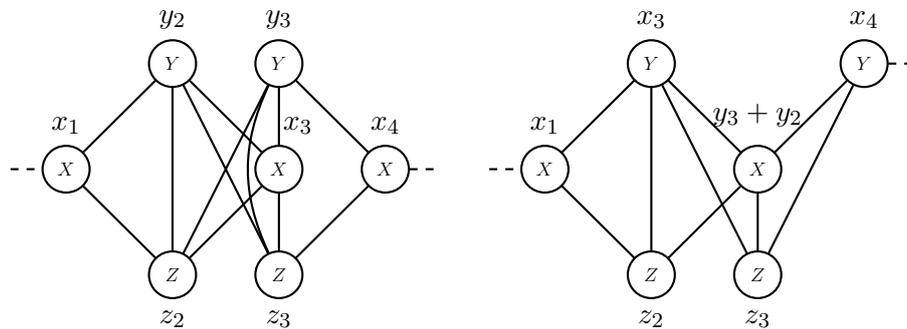


Figure 2: The case when  $x_3 < \min(y_3, z_3)$ , and  $x_3 \geq y_2$ , before and after the operations (left and right).

**2**  $x_3 < y_3$  and  $x_3 < z_3$  and ( $x_3 \geq y_2$  or  $x_3 \geq z_2$ ). (See Figure 2.)

We may assume  $x_3 \geq y_2$ , as  $x_3 \geq z_2$  can be handled similarly. After the change we have in  $L_{i-1}$  a clump of size  $x_3$  of color  $Y$  and a clump of size  $z_2$  of color  $Z$ ; and we have in  $L_i$  a clump of size  $y_3 + y_2$  of color  $X$  and a clump  $z_3$  of color  $Z$ .

Note that  $|L_{i-1} \cup L_i|$  did not change. When we verify the minimum degree condition, we use the notation of Case **1**. We have:

$$\begin{aligned} d'(X_{i-2}) &= d(X_{i-2}), \\ d'(Y_{i-1}) &= x_1 + y_2 + y_3 + z_2 + z_3 > x_1 + z_2 + x_3 + z_3 = d(Y_i), \\ d'(Z_{i-1}) &= x_1 + x_3 + y_2 + y_3 = d(Z_{i-1}), \\ d'(X_i) &= x_3 + x_4 + z_2 + z_3 = d(Y_i), \\ d'(Z_i) &= x_3 + x_4 + y_2 + y_3 = d(Z_i), \\ d'(Y_{i+1}) &= d(X_{i+1}) + y_2. \end{aligned}$$

At this point, we are left with checking the statement for  $x_3 < \min(y_3, z_3, y_2, z_2)$ . We split this into two cases.

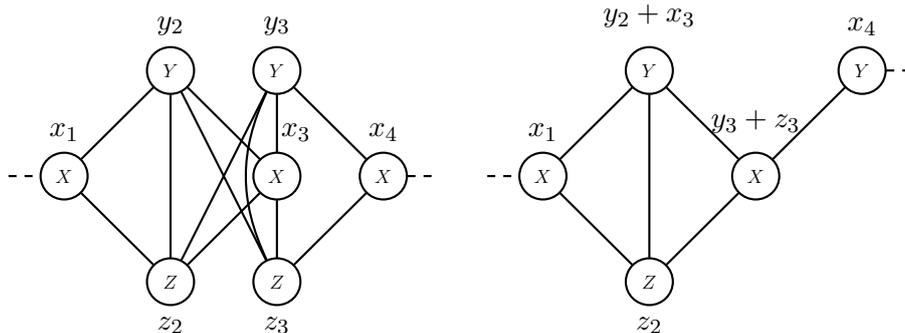


Figure 3: The case  $x_3 \leq \max(y_2, y_3, z_2, z_3)$  and  $z_2 \geq y_3$ , before and after the operations.

3  $x_3 < \min(y_3, z_3, y_2, z_2)$  and  $z_2 \geq y_3$ .

See Figure 3 for the changes. When we check the degrees, we use the notation introduced in Case 1:

$$\begin{aligned}
 d'(X_{i-2}) &= d(X_{i-2}) + x_3, \\
 d'(Y_{i-1}) &= x_1 + z_2 + y_3 + z_3 > x_1 + z_2 + x_3 + z_3 = d(Y_{i-1}), \\
 d'(Z_{i-1}) &= x_1 + y_2 + x_3 + y_3 + z_3 > x_1 + y_2 + y_3 + x_3 = d(Z_{i-1}), \\
 d'(X_i) &= z_2 + y_2 + x_3 + x_4 \geq y_2 + y_3 + x_3 + x_4 = d(Z_i), \\
 d'(Y_{i+1}) &= d(X_{i+1}).
 \end{aligned}$$

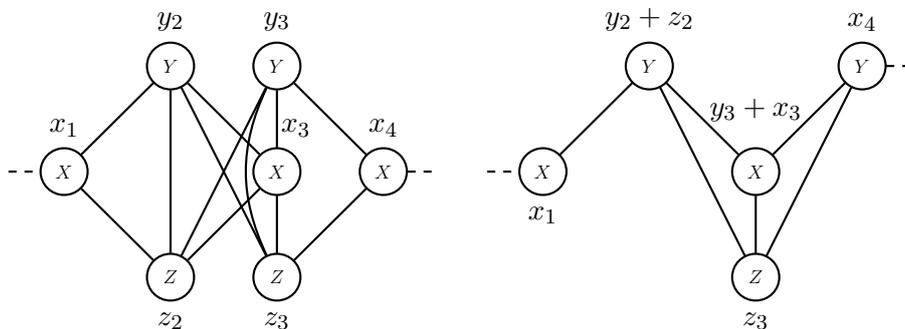


Figure 4: The case  $x_3 \leq \max(y_2, y_3, z_2, z_3)$  and  $z_2 < y_3$ , before and after the operations.

4  $x_3 < \min(y_3, z_3, y_2, z_2)$  and  $z_2 < y_3$ .

First note that the clump colored  $X$  in  $L_i$  can simply be moved into  $L_{i-1}$  keeping the minimum degree condition. After this move on the left side of Figure 4, we see the mirror image of the left side of Figure 3, just the numbers are different. The operations for this case, which are the “mirror image”, of the operations in the previous case, is shown in Figure 4. Because of the symmetry, we do not delve into the details.

This concludes the proof of (iii)

Finally, to prove part (iv), take the least  $i$ , such that  $|L_i| > c(i)$ , but  $c(i) + \max(c(i-1), c(i+1)) < k$ . As  $|L_0| = c(0) = 1$ , this means  $i > 0$ . First we show that we may assume that  $L_{i-1} \cup L_i \cup L_{i+1}$  misses some color  $X$ . Indeed, if all colors appear in  $L_{i-1} \cup L_i \cup L_{i+1}$ , let  $X$  be a color not used in  $L_{i-1} \cup L_i$  and  $Y$  be a color not used in  $L_i \cup L_{i+1}$ . Create a new coloring of  $G$  by switching the colors  $X, Y$  in  $L_j$  for all  $j \geq i + 1$ . After the switch,  $X$  is missing from  $L_{i-1} \cup L_i \cup L_{i+1}$ .

Since  $|L_i| > c(i)$ , there are two vertices  $x, y$  in  $L_i$  that are colored the same. Recolor  $x$  with color  $X$ . This is a valid coloring, in which  $L_i$  contains one more color than before. Repeating this procedure produces a coloring, in which  $|L_i| = c_i$  or  $c(i) + \max(c(i-1), c(i+1)) = k$ . Repeating this procedure recursively for the next least  $i$ , we can eliminate one after the other the  $i$ 's that fail (iv), not creating any instances where the first three statements would fail.  $\square$

The main use of the previous theorem is that we can reduce our problem to graphs that satisfy the conclusion of the theorem, i.e., to graphs whose clump graphs are canonical:

**Definition 8.** We call a  $k$ -colored weighted clump graph  $H$  *canonical*, if there is a graph  $G$  satisfying the conclusions of Theorem 7, such that the clump graph of  $G$  is  $H$ . In other words,  $H$  has  $D + 1$  layers  $L_0, L_1, \dots, L_D$ , where  $D = \text{diam}(H)$ , and for each  $1 \leq i < D$  we have

- (i) If  $|L_i| = 1$ , then  $|L_{i+1}| \leq k - 1$ .
- (ii) The number of colors used to color the set  $L_i \cup L_{i+1}$  is  $\min(k, |L_i| + |L_{i+1}|)$ . In particular, when  $|L_i| + |L_{i+1}| \leq k$ , then  $L_i$  and  $L_{i+1}$  do not share any color.
- (iii) If  $|L_i| = k$ , then  $i \geq 2$  and  $|L_{i+1}| \geq 2$ .
- (iv) If  $L_i$  has a weight that is bigger than 1, then  $i > 0$  and  $|L_i| + \max(|L_{i-1}|, |L_{i+1}|) \geq k$ .

Note that (ii) implies that the edges missing between  $L_i$  and  $L_{i+1}$  form a matching of size  $\max(0, |L_i| + |L_{i+1}| - k)$ . In particular, when  $|L_i \cup L_{i+1}| \leq k$  then all edges between  $L_i$  and  $L_{i+1}$  are present.

We will use property (i) of canonical graphs in the proof of Theorem 5, and properties (i), (ii) and (iii) in the following Corollary that will be needed for the proof of Theorem 6.

**Corollary 9.** *In a 3-colored canonical clump graph, the following sets of colors are possible in two consecutive layers:*

$$X|Y, X|YZ, YZ|X, XY|XZ, XY|XYZ, XYZ|XY, XYZ|XYZ. \quad (1)$$

### 3 Duality

In this Section we assume that  $k$  is fixed. Now we look differently at our diameter problem: how small  $n$  can be, such that a connected  $k$ -colorable graph of order  $n$ , minimum degree at least  $\delta$ , and diameter  $D$  exists? Let  $\mathcal{H}_{k,\delta,D}$  denote the family of canonical clump graphs (of diameter  $D$ ) that arises from connected  $k$ -colorable graphs  $G$  with diameter  $D$  and minimum degree at least  $\delta$ , when the order of  $G$  is unspecified. Fix an  $H \in \mathcal{H}_{k,\delta,D}$ , and consider the following packing problem for  $H$ : assign non-negative real dual weights  $u(y) \geq 0$  to  $y \in V(H)$ , and

$$\text{Maximize } \delta \cdot \sum_{y \in V(H)} u(y),$$

subject to condition

$$\forall x \in V(H) \quad \sum_{y \in V(H): xy \in E(H)} u(y) \leq 1. \quad (2)$$

**Theorem 10.** *Fix an integer  $k \geq 2$ . Assume that there exist constants  $\tilde{u} > 0, C$ , such that for all positive integers  $D$  and  $\delta$ , and all  $H \in \mathcal{H}_{k,\delta,D}$ , in the linear program (2) the optimum is at least*

$$\tilde{u}\delta D + C\delta. \quad (3)$$

*Then, for any  $k$ -colorable connected graphs  $G$  on  $n$  vertices and with minimum degree  $\delta$  we have*

$$\text{diam}(G) \leq \frac{1}{\tilde{u}} \cdot \frac{n}{\delta} - \frac{C}{\tilde{u}}. \quad (4)$$

*Proof.* Any  $H \in \mathcal{H}_{k,\delta,D}$  is the clump graph of saturated graph  $G$  with diameter  $D$ .  $G$  can be reconstructed by assigning  $w(x) \geq 1$  integer weights for all vertices of  $H$ , such that we assign 1 to the vertex in  $L_0$ . Now  $n = |V(G)| = \sum_{x \in V(H)} w(x)$ .

Consider the optimization problem

$$\text{Minimize } \sum_{x \in V(H)} w(x),$$

subject to condition

$$\forall y \in V(H) \quad \sum_{x \in V(H): xy \in E(H)} w(x) \geq \delta. \quad (5)$$

We use the trivial inequality of the duality of linear programming [5]: namely, for any  $u$  and  $w$  feasible solutions, by (5) and (2), we have:

$$\begin{aligned} \delta \sum_{y \in V(H)} u(y) &\leq \sum_{y \in V(H)} u(y) \sum_{x \in V(H): xy \in E(H)} w(x) \\ &= \sum_{x \in V(H)} w(x) \sum_{y \in V(H): xy \in E(H)} u(y) \leq \sum_{x \in V(H)} w(x). \end{aligned}$$

As the objective function reaches  $\delta \sum_{y \in V(H)} u(y) \geq \tilde{u}\delta D + C\delta$ , the theorem follows.  $\square$

*Proof of Theorem 5.* Assume  $k \geq 3$ . Consider a  $k$ -colorable canonical clump graph  $H$  with layers  $L_0, \dots, L_D$ . We are going to find a good packing  $u$  on the vertices of  $H$  as required to use Theorem 10. The dual weighting  $u$  will take at most  $2k - 2$  different values, and every layer will get the same total dual weight.

Let  $i$  be an integer,  $0 \leq i \leq D$ . Set  $L_{-1} = L_{D+1} = \emptyset$ .

If  $|L_i| \leq k - 1$ , then assign the dual weight  $\frac{k-1}{(3k-4)|L_i|}$  to every  $v \in L_i$ . This makes the total dual weight of  $L_i$  exactly  $\frac{k-1}{3k-4}$ , and the dual weight of every  $v \in L_i$  at least  $\frac{1}{3k-4}$ .

If  $|L_i| = k$ , let  $X_i$  be the (possibly empty) set of vertices in  $L_i$  connected to every vertex in  $L_{i-1} \cup L_{i+1}$ , and set  $Y_i = L_i \setminus X_i$ . By definition, none of the colors in  $L_{i-1}$  appear in  $X_i$ , so  $|X_i| \leq k - |L_{i-1}|$ . Since  $H$  is canonical, by Definition 8 (i) we have  $|L_{i-1}| \geq 2$ . Therefore  $0 \leq |X_i| \leq k - 2$ , and  $Y_i \neq \emptyset$ .

Set the dual weight of every  $v \in X_i$  to  $\frac{1}{3k-4}$ , and the dual weight of every  $v \in Y_i$  to  $\frac{1}{3k-4} - \frac{1}{(3k-4)(k-|X_i|)}$ . As  $|Y_i| = k - |X_i|$ , the total dual weight of  $L_i$  is

$$\frac{|X_i|}{3k-4} + (k - |X_i|) \left( \frac{1}{3k-4} - \frac{1}{(3k-4)(k-|X_i|)} \right) = \frac{k-1}{3k-4}.$$

Moreover, as  $k - |X_i| \geq 2$ , the dual weight of  $v \in Y_i$  is at least  $\frac{1}{2(3k-4)}$ .

Now take vertex  $x$  of  $H$ . Then  $x \in L_j$  for some  $0 \leq j \leq D$ . We are going to check that the neighbors of  $x$  have a total dual weight of at most 1.

If  $|L_j| \leq k - 1$ , or ( $|L_j| = k$  and  $x \in X_j$ ), then the weight of  $x$  is at least  $\frac{1}{3k-4}$ . Since the open neighborhood of  $x$  is a subset of  $(L_{j-1} \cup L_j \cup L_{j+1}) \setminus \{x\}$ , the sum of the weight of its neighbors is at most  $\frac{3(k-1)}{3k-4} - \frac{1}{3k-4} = 1$ .

If  $|L_j| = k$  and  $x \in Y_j$ , then there is a  $y \in L_{j-1} \cup L_{j+1}$  such that the open neighborhood of  $x$  is contained by  $(L_{j-1} \cup L_j \cup L_{j+1}) \setminus \{x, y\}$ . As the sum of the weights of  $x$  and  $y$  is at least  $\frac{1}{3k-4}$ , the total weight of the neighbors of  $x$  at at most  $\frac{3(k-1)}{3k-4} - \frac{1}{3k-4} = 1$ .

The total dual weight of the vertices in  $H$  is  $\frac{k-1}{3k-4}(D+1)$ . Now Theorem 5 follows from Theorem 10 with  $\tilde{u} = C = \frac{k-1}{3k-4}$ .  $\square$

Using  $k = 3$ , we get a weaker version of Theorem 3:

**Corollary 11.** *If  $G$  is connected 3-colorable graph of order  $n$  and minimum degree  $\delta \geq 1$ ,*

$$\text{diam}(G) \leq \frac{5n}{2\delta} - 1.$$

## 4 Inclusion-Exclusion (Sieve) for the case $k = 3$

Let us be given a 3-colorable saturated connected graph  $G$  of order  $n$  and minimum degree at least  $\delta$ , which maximizes the diameter  $D$  among such graphs. By Theorem 7, we may assume without loss of generality that the clump graph of  $G$  is canonical. Furthermore, Corollary 9 tells what kind of color sets can be in consecutive layers. We often use these facts without explicit reference in the future. Let  $\ell_i = |L_i|$  denote the number of vertices in the  $i^{\text{th}}$  layer of  $G$ . As we are about to prove Theorem 6, we can assume without loss

of generality that  $\ell_i \leq 3\delta$ . Indeed, if  $G$  does not satisfy this inequality, eliminate vertices from clumps with excess above  $\delta$ , to obtain the graph  $G'$  on  $n'$  vertices.  $G'$  still satisfies the conditions of Theorem 6, and therefore its conclusion with  $n'$  replacing  $n$ . Hence  $G$  also satisfies the conclusion of Theorem 6. We are going to build lower bounds for the sum of a couple of consecutive  $\ell_i$ 's, from which we derive lower bounds for  $n$ . The key tool is the inclusion-exclusion formula for the size of the union of the open neighborhoods of some vertices. Note that a vertex in  $L_i$  can have neighbors only in  $L_{i-1}, L_i, L_{i+1}$ . We denote the open neighborhood of vertex  $z$  by  $N(z)$ . In Subsection 4.1 we do this approach when the vertices are taken from different clumps from the same  $L_i$ , in Subsection 4.2 we do this for vertices taken from two consecutive layers. Recall that  $c(i)$  denotes the number of clumps in  $L_i$ . Let  $\mathcal{S} = \{i : c(i) = 1\}$  be the set of *singles*. We use the notation  $x_i, y_i, z_i$  to represent the number of vertices in the clumps with color  $X_i, Y_i, Z_i$ , respectively. Here  $X_i, Y_i, Z_i$  can be any of the colors  $A, B, C$  of the 3-coloring, but they must be different colors. For the ease of computation we introduce  $L_{-1} = L_{D+1} = \emptyset$ , so  $\ell_{-1} = \ell_{D+1} = 0$ .

#### 4.1 Sieve for neighborhoods of vertices from one layer

Here we assume  $0 \leq i \leq D$ . We will only need the cases when  $c(i) \neq 3$

Case 1.  $c(i) = 1$ . We obviously have  $\ell_{i-1} + \ell_{i+1} \geq \delta$ , which we will express as

$$2\ell_{i-1} + 2\ell_i + 2\ell_{i+1} \geq 2\delta + 2\ell_i. \quad (6)$$

Case 2.  $c(i) = 2$  We have  $2\ell_{i-1} + \ell_i + 2\ell_{i+1} \geq 2\delta$  from the fact that vertices from either color in the  $i^{\text{th}}$  layer have at least  $\delta$  neighbors. We will express this as

$$2\ell_{i-1} + 2\ell_i + 2\ell_{i+1} \geq 2\delta + \ell_i. \quad (7)$$

#### 4.2 Sieve by two consecutive layers

Now we assume  $0 \leq i < D$ , so  $i + 1 \leq D$ .

Case 1.  $i \in \mathcal{S}, i + 1 \in \mathcal{S}$ . We have

$$\ell_{i-1} + \ell_i + \ell_{i+1} + \ell_{i+2} \geq 2\delta. \quad (8)$$

Case 2.  $i \in \mathcal{S}, i + 1 \notin \mathcal{S}$ . By Corollary 9 we have  $c(i + 1) = 2$ . Apply (6) to  $L_i$  to obtain  $2\ell_{i-1} + 2\ell_{i+1} \geq 2\delta$ , apply (7) to  $L_{i+1}$  to obtain  $2\ell_i + \ell_{i+1} + 2\ell_{i+2} \geq 2\delta$ , and average into

$$\ell_{i-1} + \ell_i + \frac{3}{2}\ell_{i+1} + \ell_{i+2} \geq 2\delta. \quad (9)$$

Case 3.  $i \notin \mathcal{S}, i + 1 \in \mathcal{S}$ . Like in Case 2, we obtain

$$\ell_{i-1} + \frac{3}{2}\ell_i + \ell_{i+1} + \ell_{i+2} \geq 2\delta. \quad (10)$$

Case 4.  $i \notin \mathcal{S}, i + 1 \notin \mathcal{S}$ . In this case  $L_i$  and  $L_{i+1}$  must share a color, and by Corollary 9 their union must use all 3 colors. We can assume without loss of generality that none of  $X_i, Y_i, X_{i+1}, Z_{i+1}$  is empty. Take  $x_i \in X_i, y_i \in Y_i, x_{i+1} \in X_{i+1}, z_{i+1} \in Z_{i+1}$ . Considering the neighborhood of  $x_i$ , we have

$$\delta \leq \ell_{i-1} + |Y_i| + |Z_i| + |Y_{i+1}| + |Z_{i+1}|, \quad (11)$$

considering the neighborhood of  $x_{i+1}$ , we have

$$\delta \leq \ell_{i+2} + |Y_i| + |Z_i| + |Y_{i+1}| + |Z_{i+1}|, \quad (12)$$

considering the neighborhood of  $y_i$ , we have

$$\delta \leq \ell_{i-1} + |X_i| + |Z_i| + |X_{i+1}| + |Z_{i+1}|, \quad (13)$$

considering the neighborhood of  $z_{i+1}$ , we have

$$\delta \leq \ell_{i+2} + |X_i| + |Y_i| + |X_{i+1}| + |Y_{i+1}|. \quad (14)$$

Weighting (11) and (12) with  $1/3$ , (13) and (14) with  $2/3$ , and summing them up, we obtain

$$\ell_{i-1} + \frac{4}{3}[\ell_i + \ell_{i+1}] + \ell_{i+2} \geq 2\delta. \quad (15)$$

Adding up (8), (9), (10), (15) for  $i = 1, \dots, D - 1$ , we obtain

$$4n + \sum_{\substack{(i,j):i \notin \mathcal{S} \\ j \notin \mathcal{S}, |i-j|=1}} \frac{1}{3}\ell_i + \sum_{i:i+1 \in \mathcal{S}, i \notin \mathcal{S}} \frac{1}{2}\ell_i + \sum_{i:i-1 \in \mathcal{S}, i \notin \mathcal{S}} \frac{1}{2}\ell_i \geq 2D\delta + O(\delta). \quad (16)$$

The  $O(\delta)$  error term arises from the fact that certain  $\ell_i$  terms, at the front and at the end, do not arise four times, as many times they are counted in  $4n$ .

### 4.3 Sieve for neighborhoods of vertices from three consecutive layers

We are going to give lower bounds to

$$2(\ell_{i-2} + \ell_{i-1} + \ell_i + \ell_{i+1} + \ell_{i+2}) = 2|L_{i-2} \cup L_{i-1} \cup L_i \cup L_{i+1} \cup L_{i+2}| \quad (17)$$

using inclusion-exclusion, based on a case analysis of the color content of  $L_{i-1}, L_i, L_{i+1}$ .

Case 1.  $i - 1 \notin \mathcal{S}, i \notin \mathcal{S}, i + 1 \notin \mathcal{S}$ . In particular by Corollary 9 we have that for  $j \in \{i - 1, i\}$ ,  $L_j$  and  $L_{j+1}$  share a color and their union contains all three colors. The analysis boils down to two subcases:

Subcase 1.1.  $L_{i-1}$  and  $L_{i+1}$  share at least two colors. We may assume in this case that none of  $X_{i-1}, Y_{i-1}, X_i, Z_i, X_{i+1}, Y_{i+1}$  is empty.

Take  $y_{i-1} \in Y_{i-1}, x_{i-1} \in X_{i-1}, z_i \in Z_i, x_{i+1} \in X_{i+1}$  and  $y_{i+1} \in Y_{i+1}$ . Using inclusion-exclusion we have

$$|N(y_{i-1}) \cup N(z_i) \cup N(x_{i+1})| \geq 3\delta - (|X_{i-1}| + \ell_i + |Y_{i+1}|).$$

and

$$|N(x_{i-1}) \cup N(z_i) \cup N(y_{i+1})| \geq 3\delta - (|Y_{i-1}| + \ell_i + |X_{i+1}|).$$

Combining the two inequalities above we obtain

$$2(\ell_{i-2} + \ell_{i-1} + \ell_i + \ell_{i+1} + \ell_{i+2}) \geq 6\delta - 2\ell_i - \ell_{i-1} - \ell_{i+1}. \quad (18)$$

Subcase 1.2.  $L_{i-1}$  and  $L_{i+1}$  share only one color. We may assume that  $L_{i-1} = X_{i-1} \cup Y_{i-1}; L_{i+1} = Y_{i+1} \cup Z_{i+1}$  where none of  $X_{i-1}, Y_{i-1}, Y_{i+1}, Z_{i+1}, X_i, Z_i$  is empty.

Let  $x_{i-1} \in X_{i-1}, y_{i-1} \in Y_{i-1}, x_i \in X_i, z_i \in Z_i, y_{i+1} \in Y_{i+1}, z_{i+1} \in Z_{i+1}$ . Inclusion-exclusion gives

$$|N(x_{i-1}) \cup N(z_i) \cup N(y_{i+1})| \geq 3\delta - (|Y_{i-1}| + \ell_i + |Z_{i+1}|),$$

and

$$|N(y_{i-1}) \cup N(x_i) \cup N(z_{i+1})| \geq 3\delta - (|Z_{i-1}| + \ell_i + |Y_{i+1}|).$$

We obtain (18), like in the previous subcase.

Case 2.  $i-1 \notin \mathcal{S}, i \in \mathcal{S}, i+1 \notin \mathcal{S}$ . We may assume  $L_i = Z_i$ , and for  $j \in \{i-1, i+1\}$   $L_j = X_j \cup Y_j$ , where none of  $X_{i+1}, Y_{i+1}, X_{i-1}, Y_{i-1}, Z_i$  is empty. This can be handled like Subcase 1.1 to obtain (18).

Case 3.  $i-1 \in \mathcal{S}, i \in \mathcal{S}, i+1 \in \mathcal{S}$ . We can assume  $L_{i-1} = X_{i-1}, L_i = Y_i, L_{i+1} = Z_{i+1}$  (in case  $L_{i+1} = X_{i+1}$ , switch colors  $X$  and  $Z$  in layers  $L_j$  for  $j \geq i+1$ ). Select  $x_{i-1} \in X_{i-1}, y_i \in Y_i, z_{i+1} \in Z_{i+1}$ , and apply inclusion-exclusion for  $|N(x_{i-1}) \cup N(y_i) \cup N(z_{i+1})|$  to obtain

$$2(\ell_{i-2} + \ell_{i-1} + \ell_i + \ell_{i+1} + \ell_{i+2}) \geq 6\delta - 2\ell_i.$$

Case 4.  $i-1 \in \mathcal{S}, i \in \mathcal{S}, i+1 \notin \mathcal{S}$ . As the clump graph is canonical,  $c(i+1) = 2$ . Hence we can assume  $L_{i-1} = X_{i-1}, L_i = Y_i, L_{i+1} = X_{i+1} \cup Z_{i+1}$ . Applying inclusion-exclusion for the neighborhoods of representative elements, we obtain

$$|N(x_{i-1}) \cup N(y_i) \cup N(x_{i+1})| \geq 3\delta - \ell_i - |Z_{i+1}|$$

and

$$|N(x_{i-1}) \cup N(y_i) \cup N(z_{i+1})| \geq 3\delta - \ell_i - |X_{i+1}|.$$

Combining the last two displayed formulae, we obtain

$$2(\ell_{i-2} + \ell_{i-1} + \ell_i + \ell_{i+1} + \ell_{i+2}) \geq 6\delta - 2\ell_i - \ell_{i+1},$$

which is even stronger than (18).

Case 5.  $i-1 \notin \mathcal{S}, i \in \mathcal{S}, i+1 \in \mathcal{S}$ . This is a mirror image of Case 4, so we have

$$2(\ell_{i-2} + \ell_{i-1} + \ell_i + \ell_{i+1} + \ell_{i+2}) \geq 6\delta - \ell_{i-1} - 2\ell_i.$$

Case 6.  $i-1 \in \mathcal{S}, i \notin \mathcal{S}, i+1 \in \mathcal{S}$ . We may assume  $X_{i-1} = L_{i-1}, Y_i \cup Z_i = L_i, X_{i+1} = L_{i+1}$ , where  $X_{i-1}, Y_i, Z_i, X_{i+1}$  are nonempty. Select  $x_{i-1} \in X_{i-1}, y_i \in Y_i, z_i \in Z_i, x_{i+1} \in X_{i+1}$ . Clearly

$$\begin{aligned} & |L_{i-2} \cup L_{i-1} \cup L_i \cup L_{i+1} \cup L_{i+2}| \geq \\ & \geq |N(x_{i-1}) \cup N(x_{i+1})| + |N(y_i) \cup N(z_i)| - |(N(x_{i-1}) \cup N(x_{i+1})) \cap (N(y_i) \cup N(z_i))| \geq \\ & \geq (2\delta - \ell_i) + (2\delta - \ell_{i-1} - \ell_{i+1}) - \ell_i = 4\delta - 2\ell_i - \ell_{i-1} - \ell_{i+1}. \end{aligned}$$

We conclude

$$2(\ell_{i-2} + \ell_{i-1} + \ell_i + \ell_{i+1} + \ell_{i+2}) \geq 8\delta - 4\ell_i - 2\ell_{i-1} - 2\ell_{i+1}. \quad (19)$$

Case 7.  $i-1 \in \mathcal{S}, i \notin \mathcal{S}, i+1 \notin \mathcal{S}$ . We may assume  $X_{i-1} = L_{i-1}, Y_i \cup Z_i = L_i$ , where  $X_{i-1}, Y_i, Z_i$  and  $X_{i+1}$  are nonempty. Select  $x_{i-1} \in X_{i-1}, y_i \in Y_i, z_i \in Z_i, x_{i+1} \in X_{i+1}$ . Clearly

$$\begin{aligned} |L_{i-2} \cup L_{i-1} \cup L_i \cup L_{i+1} \cup L_{i+2}| & \geq |N(x_{i-1}) \cup N(x_{i+1}) \cup N(y_i) \cup N(z_i)| \\ & \geq (2\delta - \ell_i) + (2\delta - \ell_{i-1} - \ell_{i+1}) - \ell_i \\ & = 4\delta - 2\ell_i - \ell_{i-1} - \ell_{i+1}. \end{aligned}$$

We conclude (19) again.

Case 8.  $i-1 \notin \mathcal{S}, i \notin \mathcal{S}, i+1 \in \mathcal{S}$ . As this is the mirror image of Case 7', we arrive at the same conclusion (19), as the conclusion is symmetric.

For  $1 \leq i \leq D-1$ , we call a triplet of consecutive layers  $(i-1, i, i+1)$  *singular*, if  $i \notin \mathcal{S}$  and  $(i+1 \in \mathcal{S}$  or  $i-1 \in \mathcal{S})$ . Let  $s$  denote the number of singular triplets. Summing up the lower bounds to (17) obtained in the 8 cases, we have

$$\begin{aligned} 10n & \geq 6\delta D - 2n + O(\delta) - \sum_{\substack{i:i \notin \mathcal{S} \\ i-1, i+1 \notin \mathcal{S}}} (\ell_{i-1} + \ell_{i+1}) - \sum_{\substack{i:i \in \mathcal{S} \\ i-1, i+1 \notin \mathcal{S}}} (\ell_{i-1} + \ell_{i+1}) \\ & - \sum_{\substack{i \in \mathcal{S} \\ i-1 \in \mathcal{S}, i+1 \notin \mathcal{S}}} \ell_{i+1} - \sum_{\substack{i \in \mathcal{S} \\ i-1 \notin \mathcal{S}, i+1 \in \mathcal{S}}} \ell_{i-1} - \sum_{\substack{i:i \notin \mathcal{S} \\ i-1 \in \mathcal{S}, i+1 \in \mathcal{S}}} (-2\delta + 2\ell_i + \ell_{i-1} + \ell_{i+1}) \\ & = 6\delta D - 2n + 2s\delta + O(\delta) \\ & - \sum_{\substack{i:i \notin \mathcal{S} \\ i+1, i+2 \notin \mathcal{S}}} \ell_i - \sum_{\substack{i:i \notin \mathcal{S} \\ i-1, i-2 \notin \mathcal{S}}} \ell_i - \sum_{\substack{i:i \notin \mathcal{S} \\ i+1 \in \mathcal{S}, i+2 \notin \mathcal{S}}} \ell_i - \sum_{\substack{i:i \notin \mathcal{S} \\ i-1 \in \mathcal{S}, i-2 \notin \mathcal{S}}} \ell_i \\ & - \sum_{\substack{i \notin \mathcal{S} \\ i-1, i-2 \in \mathcal{S}}} \ell_i - \sum_{\substack{i \notin \mathcal{S} \\ i+1, i+2 \in \mathcal{S}}} \ell_i - 2 \sum_{\substack{i:i \notin \mathcal{S} \\ i-1 \in \mathcal{S}, i+1 \in \mathcal{S}}} \ell_i - \sum_{\substack{i:i+1 \notin \mathcal{S} \\ i \in \mathcal{S}, i+2 \in \mathcal{S}}} \ell_i - \sum_{\substack{i:i-1 \notin \mathcal{S} \\ i \in \mathcal{S}, i-2 \in \mathcal{S}}} \ell_i \end{aligned}$$

Now we use that

$$\sum_{\substack{i:i+1 \notin \mathcal{S} \\ i \in \mathcal{S}, i+2 \in \mathcal{S}}} \ell_i + \sum_{\substack{i:i-1 \notin \mathcal{S} \\ i \in \mathcal{S}, i-2 \in \mathcal{S}}} \ell_i \leq \sum_{\substack{i:i \notin \mathcal{S} \\ i+1 \notin \mathcal{S}, i+2 \in \mathcal{S}}} \ell_i + \sum_{\substack{i:i \notin \mathcal{S} \\ i-1 \notin \mathcal{S}, i-2 \in \mathcal{S}}} \ell_i + 2 \sum_{i:i \in \mathcal{S}} \ell_i$$

and

$$2 \sum_{i \notin S} l_i = \left( \sum_{\substack{i: i \notin S \\ i+1, i+2 \notin S}} l_i + \sum_{\substack{i: i \notin S \\ i+1 \in S, i+2 \notin S}} l_i + \sum_{\substack{i: i \notin S \\ i+1 \notin S, i+2 \in S}} l_i + \sum_{\substack{i: i \notin S \\ i+1, i+2 \in S}} l_i \right) \\ + \left( \sum_{\substack{i: i \notin S \\ i-1, i-2 \notin S}} l_i + \sum_{\substack{i: i \notin S \\ i-1 \in S, i-2 \notin S}} l_i + \sum_{\substack{i: i \notin S \\ i-1 \notin S, i-2 \in S}} l_i + \sum_{\substack{i: i \notin S \\ i-1, i-2 \in S}} l_i \right)$$

to obtain

$$12n \geq 6\delta D + 2s\delta + O(\delta) - 2 \sum_{i \notin S} l_i - 2 \sum_{\substack{i: i \notin S \\ i-1 \in S, i+1 \in S}} l_i - 2 \sum_{i \in S} l_i \\ = 6\delta D + 2s\delta - 2n + O(\delta) - 2 \sum_{\substack{i: i \notin S \\ i-1 \in S, i+1 \in S}} l_i.$$

This gives

$$7n \geq 3\delta D + s\delta + O(\delta) - \sum_{\substack{i: i \notin S \\ i-1 \in S, i+1 \in S}} l_i. \quad (20)$$

## 5 Optimization

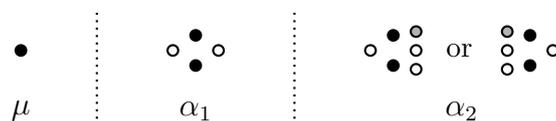


Figure 5: Visual representation for some variables denoted with Greek letters. Layers with black filled circles represent the layers whose vertices we count, the empty circles show how many colors are present in the nearby layers. Gray filled circles represent a third color that may or may not be present in the layer.

The inequalities (16) and (20) are key constraints for our linear program. The linear program is in global variables, which are mostly the fraction of vertices of  $G$  in certain type of layers, which live in a neighborhood of certain type of layers. The global variables,

denoted by Greek letters, will be:

$$\begin{aligned}\mu &= \frac{1}{n} \sum_{i:c(i)=1} \ell_i \\ \alpha_1 &= \frac{1}{n} \sum_{\substack{i:0 < i < D, c(i)=2 \\ i-1, i+1 \in \mathcal{S}}} \ell_i \\ \alpha_2 &= \frac{1}{n} \sum_{\substack{i:0 < i < D, c(i)=2, \\ i-1 \in \mathcal{S}, i+1 \notin \mathcal{S}}} \ell_i + \frac{1}{n} \sum_{\substack{i:0 < i < D, c(i)=2, \\ i+1 \in \mathcal{S}, i-1 \notin \mathcal{S}}} \ell_i \\ \phi &= \frac{D\delta}{n} \\ \psi &= \frac{\delta s}{n}\end{aligned}$$

Figure 5 illustrates the variables whose definition involves sums. Clearly, all variables are non-negative. We use Corollary 9 on what kind of layers can be consecutive. From the definitions, it easily follows that

$$\mu + \alpha_1 + \alpha_2 \leq 1. \tag{21}$$

We have

$$\sum_{\substack{(i,j): i \notin \mathcal{S} \\ j \notin \mathcal{S}, |i-j|=1}} \ell_i = 2n(1 - \alpha_1 - \alpha_2 - \mu) + n\alpha_2 + O(\delta) = n \left( 2 - 2\mu - 2\alpha_1 - \alpha_2 + O\left(\frac{\delta}{n}\right) \right),$$

since (except possibly for  $i = D$ )  $\ell_i$ 's accounted for in the definition of  $\mu$  and  $\alpha_i$  do not contribute to the sum on the left side,  $\ell_i$ -s accounted for in  $\alpha_2$  appear once, and all other  $\ell_i$ 's appear twice. In addition,

$$\sum_{i:i+1 \in \mathcal{S}, i \notin \mathcal{S}} \ell_i + \sum_{i:i-1 \in \mathcal{S}, i \notin \mathcal{S}} \ell_i = n \left( 2\alpha_1 + \alpha_2 + O\left(\frac{\delta}{n}\right) \right).$$

Using these observations, simple algebra derives from (16)

$$12\phi + 4\mu - 2\alpha_1 - \alpha_2 \leq 28 + O\left(\frac{\delta}{n}\right). \tag{22}$$

From (20), using

$$\sum_{\substack{i:i \notin \mathcal{S} \\ i-1 \in \mathcal{S} \vee i+1 \in \mathcal{S}}} \ell_i = n(\alpha_1 + \alpha_2)$$

we get

$$7 \geq 3\phi + \psi - \alpha_1 - \alpha_2 + O\left(\frac{n}{\delta}\right). \tag{23}$$

Let  $\mathcal{D}$  denote the set of layers with 2 colors, with singles on both side. (Their cardinalities added up to  $\alpha_1$ .) Let  $\mathcal{E}$  denote the set of layers that are adjacent to at least one layer from  $\mathcal{D}$ . Hence all layers in  $\mathcal{E}$  are singles. Let  $\mathcal{F}$  denote the set of remaining layers, i.e. not in  $\mathcal{D} \cup \mathcal{E}$ . First note that

$$|\mathcal{D}| + |\mathcal{E}| + |\mathcal{F}| = D + 1. \quad (24)$$

By the minimum degree condition, for all  $i : 0 < i < D$  we have  $\delta \leq \ell_{i-1} + \ell_i + \ell_{i+1}$ . Hence,  $\delta|\mathcal{F}| \leq \sum_{i \in \mathcal{F}} (\ell_{i-1} + \ell_i + \ell_{i+1}) \leq 3n(1 - \alpha_1) + O(\delta)$  and

$$|\mathcal{F}| \leq 3(1 - \alpha_1) \frac{n}{\delta} + O(1). \quad (25)$$

It is not difficult to see that  $|\mathcal{D}| + |\mathcal{E}| \leq 3s$ . Using this observation with (24) and (25), we obtain

$$3\phi = \frac{3\delta}{n} (|\mathcal{D}| + |\mathcal{E}| + |\mathcal{F}| - 1) \leq \frac{3\delta}{n} (3s) + \frac{3\delta}{n} |\mathcal{F}|$$

and hence

$$\phi \leq 3\psi + 3(1 - \alpha_1) + O\left(\frac{\delta}{n}\right). \quad (26)$$

We tried to use more inequalities and more variables, splitting  $\alpha_1$  further, based on the number of colors in the layers before and after. Removing redundant variables and conditions, we finalized our linear program based on constraints (21), (22), (23) and (26) as follows:

**Maximize**  $\phi = \frac{D\delta}{n}$   
**subject to**

$$\begin{aligned} \mu + \alpha_1 + \alpha_2 &\leq 1 \\ 12\phi + 4\mu - 2\alpha_1 - \alpha_2 &\leq 28 + O\left(\frac{\delta}{n}\right) \\ 3\phi + \psi - \alpha_1 - \alpha_2 &\leq 7 + O\left(\frac{\delta}{n}\right) \\ \phi - 3\psi + 3\alpha_1 &\leq 3 + O\left(\frac{\delta}{n}\right) \\ \phi, \mu, \psi, \alpha_1, \alpha_2 &\geq 0 \end{aligned}$$

Let  $\mathbf{x} = (x_1, x_2, x_3, x_4, x_5)^T = (\phi, \mu, \psi, \alpha_1, \alpha_2)^T$ , let  $A$  be the  $4 \times 5$  coefficient matrix above,  $\mathbf{b} = (1, 28, 7, 3)^T$ , and  $\mathbf{h}$  be any concrete error term in the constraint column within the  $O\left(\frac{\delta}{n}\right)$  bounds. Let  $\mathbf{y} = (y_1, y_2, y_3, y_4, y_5)$ . Consider now four closely related linear programs:

$$A\mathbf{x} \leq \mathbf{b} + \mathbf{h}; \quad \mathbf{x} \geq 0; \text{ maximize } x_1; \quad (27)$$

$$A\mathbf{x} \leq \mathbf{b}; \quad \mathbf{x} \geq 0; \text{ maximize } x_1; \quad (28)$$

$$\mathbf{y}A \geq (1, 0, 0, 0, 0); \quad \mathbf{y} \geq 0; \text{ minimize } \mathbf{y}(\mathbf{b} + \mathbf{h})^T; \quad (29)$$

$$\mathbf{y}A \geq (1, 0, 0, 0, 0); \quad \mathbf{y} \geq 0; \text{ minimize } \mathbf{y}\mathbf{b}^T. \quad (30)$$

Our standard reference to linear programming is [5]. Note that (27) is identical to the displayed linear program, and that (27) and (29), and (28) and (30) are dual linear programs, respectively, and the Duality Theorem of Linear Programming applies to them. Utilizing the open source online tool [10], we solved (28) with optimum  $\phi = \frac{57}{23}$  attained at  $(\frac{57}{23}, 0, \frac{13}{22}, \frac{17}{23}, \frac{6}{23})^T$ . By duality,  $\frac{57}{23}$  is the optimum of (30) as well. The polytope defined by the constraints of (28) has a feasible solution  $\mathbf{x}^*$ , for which inequalities in the 3<sup>rd</sup>, 4<sup>th</sup> and 5<sup>th</sup> constraints hold strictly—just modify the optimal solution by reducing  $\phi$  a bit. We want to show that (27) has a finite optimum, if  $n$  is sufficiently large. Using the first and third constraints in (27), we get that  $12\phi \leq 30 + O(\frac{\delta}{n})$ , which implies that  $\phi \leq 3$  for  $n$  sufficiently large. Our only concern is whether (27) has a feasible solution at all, as negative error terms might eliminate it. Clearly  $\mathbf{x}^*$  is a feasible solution, if  $n$  is sufficiently large. By the Duality Theorem, (29) has a finite minimum value, which is equal to the maximum value for (27). As the polytopes of (29) and (30) are the same, this finite minimum is achieved in one of the finitely many vertices of this polytope, say  $\mathbf{y}^{(1)}, \dots, \mathbf{y}^{(m)}$ , as these linear programs only differ in their objective functions. Now we have

$$\begin{aligned} \max x_1 \text{ in (27)} &= \min_{\mathbf{y} \geq 0} \mathbf{y}(\mathbf{b} + \mathbf{h})^T = \min_{i=1}^m \mathbf{y}^{(i)}(\mathbf{b} + \mathbf{h})^T \\ &\geq \min_{i=1}^m \mathbf{y}^{(i)}\mathbf{b}^T + \min_{i=1}^m \mathbf{y}^{(i)}\mathbf{h}^T = \frac{57}{23} + O\left(\frac{\delta}{n}\right). \end{aligned}$$

On the other hand,

$$\begin{aligned} \max x_1 \text{ in (27)} &= \min_{\mathbf{y} \geq 0} \mathbf{y}(\mathbf{b} + \mathbf{h})^T = \min_{i=1}^m \left( \mathbf{y}^{(i)}\mathbf{b}^T + \mathbf{y}^{(i)}\mathbf{h}^T \right) \\ &\leq \min_{i=1}^m \left( \mathbf{y}^{(i)}\mathbf{b}^T + \max_{j=1}^m \mathbf{y}^{(j)}\mathbf{h}^T \right) = \min_{i=1}^m \mathbf{y}^{(i)}\mathbf{b}^T + \max_{i=1}^m \mathbf{y}^{(i)}\mathbf{h}^T \\ &= \frac{57}{23} + O\left(\frac{\delta}{n}\right). \end{aligned}$$

We concluded the proof of Theorem 6. The linear programming arguments above should be well-known, but we were unable to find a reference.

The following theorem proves the weaker version of Conjecture 4 for  $k = 3$ , in a restricted case of no single layers:

**Theorem 12.** *For every connected 3-colorable graph  $G$  of order  $n$  and minimum degree at least  $\delta \geq 1$ , such that in the canonical clump graph of  $G$  no layer  $L_i$  is a single for  $0 < i < D$ , we have*

$$\text{diam}(G) \leq \frac{7n}{3\delta} + O(1).$$

*Proof.* If there are no single color layers besides  $L_0$  and  $L_D$ , in (16) the second and third sums are zero, and the first is upper bounded by  $\frac{2}{3}n$ . This yields  $14n/3 \geq 2D\delta + O(\delta)$ . An alternative proof of the theorem is from (20), in which  $s = 0$  and the sum is  $O(\delta)$  in this case.  $\square$

The theorem also holds if the number of single layers is bounded as  $n \rightarrow \infty$ . We are not aware of constructions getting close to this upper bound without single layers.

## Acknowledgements

The first and third authors thank Peter Dankelmann for introducing them to the problem and for suggesting the approach of using  $k$ -colorability instead of forbidden cliques. We also want to thank an anonymous referee for their careful reading.

## References

- [1] D. Amar, I. Fournier, A. Germa. Ordre minimum d'un graphe simple de diamètre, degré minimum et connexité donnés. *Ann Discrete Math.*, **17** (1983), 7–10.
- [2] L. Caccetta, W.F. Smyth. Graphs of maximum diameter. *Discrete Math.*, **102** (1992), 121–141.
- [3] É. Czabarka, P. Dankelmann, L. A. Székely. Diameter of 4-colorable graphs. *Europ. J. Comb.* **30** (2009), 1082–1089.
- [4] É. Czabarka, I. Singgih, L. A. Székely. Counterexamples to a conjecture of Erdős, Pach, Pollack and Tuza. *J. Combin. Theory B* **151** (2021), 38–45.
- [5] G. Dantzig. *Linear Programming and Extensions*. Princeton University Press and the RAND Corporation, 1963.
- [6] P. Erdős, J. Pach, R. Pollack, and Z. Tuza. Radius, diameter, and minimum degree. *J. Combin. Theory B* **47** (1989), 279–285.
- [7] Fan Chung, Ron Graham. *Erdős on Graphs: His Legacy of Unsolved Problems*, A K Peters Ltd., CRC Press, Taylor & Francis, 1998.
- [8] D. Goldsmith, B. Manvel, V. Faber. A lower bound for the order of the graph in terms of the diameter and minimum degree. *J. Combin. Inform. Syst. Sci.* **6** (1981), 315–319.
- [9] J.W. Moon. On the diameter of a graph. *Mich. Math. J.* **12** (1965), 349–351.
- [10] PHPSimplex. Optimizing resources with Linear Programming. <http://www.phpsimplex.com/en/>
- [11] A. A. Razborov. Flag algebras. *J. Symbolic Logic* **72**(4) (2007), 1239–1282.