DIMACS Technical Report 2009-03
February 2009

The cycle discrepancy of three-regular graphs

by

Sarmad Abbasi[1]
117-BB, Phase IV, DHA
Lahore 08903
Pakistan

Laeeq Aslam
Punjab University College of Information Technology
University of the Punjab
Lahore, Pakistan

[1]The author was supported by DIMACS for a summer internship during his graduate studies.

ABSTRACT

Let $G = (V, E)$ be an undirected graph and $\mathcal{C}(G)$ denote the set of all cycles in G. We introduce a graph invariant cycle discrepancy, which we define as

$$\mathrm{cycdisc}(G) = \min_{\chi: V \mapsto \{+1,-1\}} \max_{C \in \mathcal{C}(G)} \left| \sum_{v \in C} \chi(v) \right|.$$

We show that, if G is a three-regular graph with n vertices, then

$$\mathrm{cycdisc}(G) \leq \frac{n+2}{6}.$$

This bound is best possible and is achieved by very simple graphs. Our proof is algorithmic and allows us to compute in $O(n^2)$ time a labeling χ, such that

$$\max_{C \in \mathcal{C}(G)} \left| \sum_{v \in C} \chi(v) \right| \leq \frac{n+2}{6}.$$

Some interesting open problems regarding cycle discrepancy are also suggested.

1 Introduction

Discrepancy theory is an important and vast subject in combinatorics. One of the major goals of this subject is to study the following problem: given a set V and a collection, $\mathcal{S} = \{S_0, \ldots, S_m\}$, of subsets of V, how can we partition the ground set V into two sets such that each $S_i \in \mathcal{S}$ is partitioned as equally as possible? More formally, the quantity of interest in such a problem is called the discrepancy of the set system $\mathcal{S}$ and is given by:

$$\operatorname{disc}(\mathcal{S}) = \min_{\chi: V \mapsto \{+1,-1\}} \max_{S \in \mathcal{S}} \left| \sum_{v \in S} \chi(v) \right| .$$

This fascinating quantity is the subject of many brilliant and important investigations in combinatorics[9, 2]. We refer the interested reader to [1, 5, 7] for a more complete and thorough introduction to this subject.

In this paper we explore discrepancy when the set system in question is obtained from an undirected graph. More precisely, we study this quantity when the set system consists of the set of all cycles of a graph. This defines an interesting graph invariant, which we call the cycle discrepancy of the graph G. The set of all cycles of a graph typically has exponential size and does not have bounded VC-dimension(See [7] for precise definition of this concept). As we will show, the cycle discrepancy of a maximum degree three graphs can be linear in the number of vertices. This is not the case for typical discrepancy problems which are geometric in nature and the discrepancy is usually around $O(\sqrt{n})$. The problem we pose does not seem to have any geometric roots. It is purely a graph theoretic question and our main result is obtained by using graph-theoretic tools only.

Another invariant closely related to the discrepancy of a set system is its VC-dimension. It is also interested to study the VC-dimension of set systems defined by graphs. This study is conducted by Kraakis *et. al.* in [6]. The authors study the VC-dimension of set systems that are defined by cycles, paths, connected subgraphs and stars. These problems are studied from a complexity-theoretic and extremal point of view. Schaefer[8] also studies set systems that arise from graphs from a complexity-theoretic point of view. He shows that one can obtain rare examples of natural problems that are $\mathbf{\Sigma}_p^3$-complete.

This paper is mainly concerned about the cycle discrepancy of three-regular graphs. We construct very simple three-regular graphs with high cycle discrepancy. The main contribution of this paper is to show that the cycle discrepancy achieved by these graphs is optimal. It is interesting to note that our main result follows a progression. In Theorem 3 we observe that the cycle discrepancy of a three-regular graph cannot be more that $\frac{1}{3}n$. This result is then improved to $\frac{3}{10}n$ in Theorem 4 and to $\frac{1}{5}n$ in Theorem 5. Finally, Theorem 6 gives an optimal bound.

The rest of this paper is organized as follows: in the next section we give some definitions and preliminary results. In Section 3, we construct maximum degree three graphs with high cycle discrepancy. Section 4 is devoted to proving an upper bound on the cycle discrepancy of three-regular graphs. In Section 5, we discuss the algorithmic aspects of Theorem 6.

Section 6 discusses the cycle discrepancy of bounded degree graphs. In the last section we propose some concrete open problems and give some concluding remarks.

2 Definitions and some preliminary results

All graphs we deal with are loop less, without multiple edges, and undirected. We use standard graph theoretic notation which is mostly borrowed from Bollabás' Monograph[3]. For a graph $G = (V, E)$ we say that x is *connected to* y and denote it by $x \sim_G y$ if $\{x, y\} \in E$. For any vertex $v \in V$ and $A \subseteq V$ we let

$$\begin{aligned} N_G(v, A) &= \{w \in A : v \sim_G w\} \text{ and} \\ d_G(v, A) &= |N(v, A)| \end{aligned}$$

$N_G(v, A)$ and $d_G(v, A)$ are called the *neighborhood of* v *in* A and the *degree of* v *in* A, respectively. $N_G(v, V)$ and $d_G(v, V)$ are denoted by $N_G(v)$ and $d_G(v)$, respectively. The *maximum degree*, $\Delta(G)$, of a graph is given by $\Delta(G) = \max_{v \in V} d_G(v)$. The subscript G is usually ignored from this notation when the graph in question is clear from the context.

A graph $G = (V, E)$ is three-regular if the degree of every vertex in G is exactly three. G is called *three colorable* if there exists a $c : V \to \{\text{red, blue, yellow}\}$ such that if $x \sim y$ then $c(x) \neq c(y)$. The map c is called a *tricoloring* of G. A tricoloring naturally partitions the vertex set into three color classes given by: $X = c^{-1}(\text{red})$, $Y = c^{-1}(\text{blue})$ and $Z = c^{-1}(\text{yellow})$. A tricoloring is characterized up to permutation of colors by the partition of the vertex set V into its color classes. Therefore, we sometimes say that $c = (X, Y, Z)$ is a tricoloring of G. Furthermore, when we use this notation, we will tacitly assume that $|Z| \leq |Y| \leq |X|$.

If $W \subseteq V$ then $G[W]$ will denote the graph induced by the set W. K_r denotes the complete graph on r vertices. A *path*, $P = v_1, \ldots, v_t$, is an ordered sequence of distinct vertices in V such that $v_i \sim v_{i+1}$ for $i = 1, \ldots, t-1$. A *cycle* $C = v_1, \ldots, v_t$ is an ordered sequence of distinct vertices with $v_i \sim v_{i+1}$ for $i = 0, \ldots, t-1$ and $v_t \sim v_1$ with $t \geq 3$. A cycle of length t will sometimes be denoted by C_t. We let $\mathcal{P}(G)$ and $\mathcal{C}(G)$ denote the set of all paths and the set of all cycles in G. A subgraph H, such as a path or a cycle, defines a subset of vertices of G that appear in it. When we apply the usual set theoretic operations to H the operations are implicitly being applied to this underlying set of vertices.

Disjoint union is denoted by $\uplus$; that is, $A = B \uplus C$ is a short form for $A = B \cup C$ and $B \cap C = \emptyset$. A *labeling*, χ, of a graph G is a mapping of V to $\{+1, -1\}$. A labeling can also be thought of as coloring of the vertices with two colors without having the restriction that adjacent vertices get different colors. However, we will never use this terminology as it conflicts with the notion of tricoloring. We will use $+$ and $-$ to denote $+1$ and -1. For any $S \subseteq V$ we define,

$$\chi(S) = \sum_{v \in S} \chi(v).$$

We define the *cycle discrepancy of the labeling*, χ by

$$\mathrm{cycdisc}(\chi) = \max_{C \in \mathcal{C}(G)} |\chi(C)|.$$

Finally, cycdisc(G), the *cycle discrepancy of G* is defined as

$$\mathrm{cycdisc}(G) = \min_{\chi : V \mapsto \{+,-\}} \mathrm{cycdisc}(\chi).$$

2.1 Some simple facts about cycle discrepancy

Cycle discrepancy is a graph invariant; that is, isomorphic graphs have the same cycle discrepancy. The following facts are easily established about this invariant.

Fact 1 $\mathrm{cycdisc}(G) = 0$ *if and only if G is bipartite.* ∎

Fact 2 *If G has k connected components $G_1, \ldots, G_k$ then*

$$\mathrm{cycdisc}(G) = \max_{i=1}^{k} \mathrm{cycdisc}(G_i).$$

∎

Fact 3 *If $G_1, \ldots, G_k$ are two-connected components of G then*

$$\mathrm{cycdisc}(G) = \max_{i=1}^{k} \mathrm{cycdisc}(G_i).$$

∎

Fact 4 *Cycle discrepancy is monotone; that is, if H is a subgraph of G then*

$$\mathrm{cycdisc}(H) \leq \mathrm{cycdisc}(G).$$

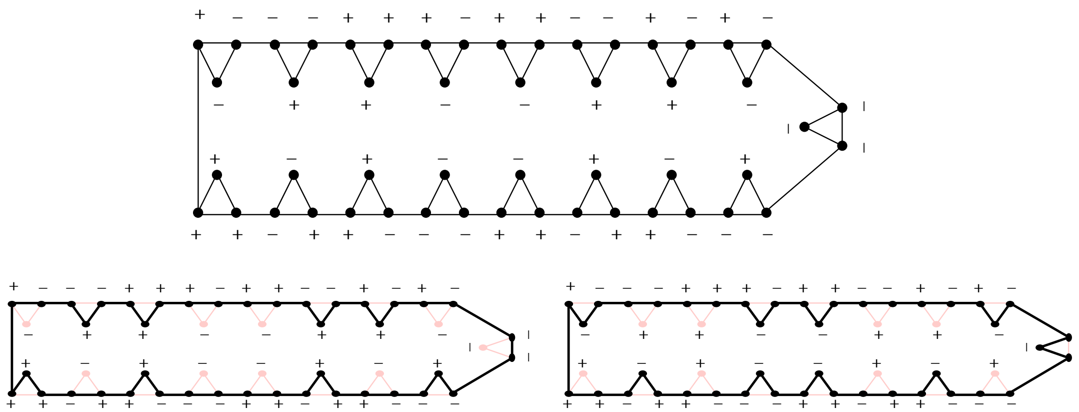
∎

When we are trying to prove an upper bound B on the cycle discrepancy of a graph G, we only have to exhibit one labeling, χ, such that $-B \leq \chi(C) \leq B$ for all cycles C in G. In most of our proofs the above inequalities are treated separately. Therefore, it is convenient for us to define:

$$\chi^+(C) = \chi(C) \text{ and } \chi^-(C) = -\chi(C).$$

We call a cycle C positive-heavy or negative-heavy if $\chi^+(C) > 0$ or $\chi^-(C) > 0$, respectively. Thus to prove that $\mathrm{cycdisc}(G) \leq B$ it suffices to exhibit a χ such that:

1. $\chi^+(C) \leq B$ for all positive-heavy cycles.
2. $\chi^-(C) \leq B$ for all negative-heavy cycles.



C^+ C^-

Figure 1: The graph G_t with high cycle discrepancy: The cycles C^+ and C^- pick up different z-vertices.

3 Maximum degree three graphs with high cycle discrepancy

Since cycle discrepancy is a monotone graph property. Therefore, it is interesting to study how large can the discrepancy of a graph be, if we bound its maximum degree. All maximum degree one graphs have zero cycle discrepancy. A maximum degree two graph is a disjoint union of paths and cycles. Thus, its cycle discrepancy is one if and only if it contains an odd cycle. The first interesting extremal question is to determine the maximum possible cycle discrepancy that an n vertex maximum degree three graphs. We start by constructing such graphs with high cycle discrepancy.

Theorem 1 *For every $n = 3t$ there exists a graph such that $\Delta(G) \leq 3$ and*

$$\mathrm{cycdisc}(G) \geq t/2.$$

Proof. Let $G_t = (V_t, E_t)$ be a graph on $V_t = \{x_0, y_0, z_0, \ldots, x_{t-1}, y_{t-1}, z_{t-1}\}$ consisting of t triangles connected in a cycle (See Figure 1.) Let $\chi : V_t \mapsto \{+, -\}$ be any labeling of G_t. Consider two cycles C^+ and C^-. The cycle C^+ goes through all the vertices x_i, z_i and also includes all y_i such that $\chi(y_i) = +$. Similarly, C^- goes through all the vertices x_i, z_i and includes all the y_i with $\chi(y_i) = -$. Thus

$$\chi^+(C^+) + \chi^-(C^-) = t.$$

This implies that $\chi(C^+) \geq \lceil t/2 \rceil$ or $\chi(C^-) \geq \lceil t/2 \rceil$. ∎

Based on the above graphs, it is easy to construct three-regular graphs with high cycle discrepancy.

Theorem 2 *For every $n \geq 6$ such that n is even, there exists an three-regular graph, G, on n vertices such that*

$$\text{cycdisc}(G) \geq \left\lfloor \frac{n+2}{6} \right\rfloor.$$

It is curious that the graphs constructed in Theorem 1 are not degree three maximal: there are approximately $\frac{n}{3}$ vertices of degree less than three. It is tempting to try to improve the above theorem by adding additional edges to these graphs. However, as we will prove, no significant improvement in the above construction is possible.

4 The cycle discrepancy of three-regular graphs

In this section we bound the discrepancy of three-regular graphs. If a three-regular graph contains a K_4, then the K_4 forms a connected component. By Fact 2 we can simply remove the K_4 from consideration and bound the discrepancy of the remaining graph.

The crux of our argument is a bound on the cycle discrepancy of K_4-free, three-regular graphs. Throughout this section we let $G = (V, E)$ be a three-regular graph. Furthermore, we assume that G does not contain a K_4. According to a classical theorem of Brooks'[4], G is three colorable. Let $c = (X, Y, Z)$ be a tricoloring of G that minimizes the size of the smallest color class; namely Z. We will now use this tricoloring to obtain our labeling.

4.1 The XYZ-labeling

Note that the minimality of Z implies the following facts:

Fact 5 *For every $z \in Z$, $d(z, X) \geq 1$ and $d(z, Y) \geq 1$.* ∎

Fact 6 *If z, z' are two distinct vertices in Z with $d(z, X) = d(z', X) = 2$, then z and z' do not have a common neighbor in Y.*

Proof. If z and z' have a common neighbor $y \in Y$ then note that $(X, Y \cup \{z, z'\} \setminus \{y\}, Z \cup \{y\} \setminus \{z, z'\})$ is a tricoloring that contradicts the minimality of Z. ∎

Fact 7 *If z, z' are two distinct vertices in Z with $d(z, Y) = d(z', Y) = 2$, then z and z' do not have a common neighbor in X.* ∎

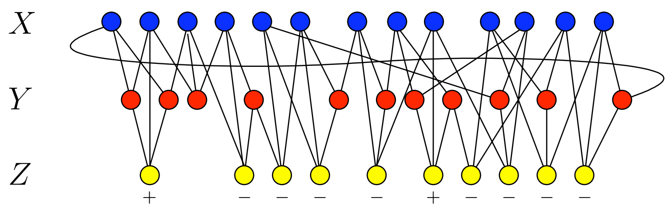


Figure 2: The labeling based on a tricoloring of G: red vertices are labeled $+$, blue vertices are labeled $-$ and each yellow vertices get the label that is opposite of the majority label of its neighbors.

We now define a labeling, χ_c, in which all the vertices in X are labeled $+$ and all the vertices in Y are labeled $-$. Each vertex in Z takes on the label which is the opposite of the majority label of its neighbors. Formally,

$$\chi_c = \begin{cases} + & \text{if } v \in X, \\ - & \text{if } v \in Y, \\ + & \text{if } v \in Z \text{ and } d(v, Y) = 2, \\ - & \text{if } v \in Z \text{ and } d(v, X) = 2. \end{cases}$$

This labeling is illustrated in Figure 2. Fact 5 shows that this labeling is well defined. Fact 6 and 7 together imply the following:

Fact 8 *Each vertex $v \in V$ (and not only the vertices of Z) has the color that is opposite of the majority color of its neighbors.* ∎

This property alone is sufficient to prove the following weak bound.

Theorem 3

$$\mathrm{cycdisc}(G) \leq \left\lfloor \frac{n}{3} \right\rfloor + 2.$$

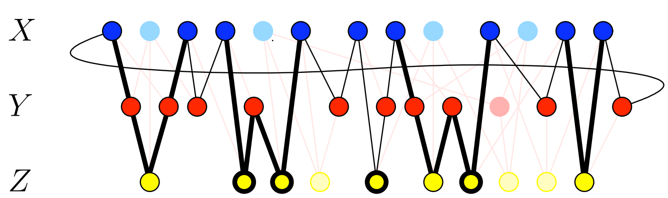


Figure 3: An illustration of Lemma 2 and Lemma 1. The bold paths shows the four cases discussed in Lemma 1. The bold vertices belong to L_C.

Proof. For any path, P, of length three we have $\chi_c(P) = +1$ or $\chi_c(P) = -1$ This is because no three vertices in a path can have the same label; otherwise, the middle vertex violates Fact 8. The theorem follows. ∎

χ_c is actually much better and we can improve the above theorem. Let

$$\begin{aligned} Z^+ &= \{v \in Z : \chi_c(v) = +\} \text{ and} \\ Z^- &= \{v \in Z : \chi_c(v) = -\}. \end{aligned}$$

For a path P, define L_P to be the vertices having exactly one neighbor in X and Y (on the path P). Formally,

$$L_P = \{z \in Z : d_P(z, X) = d_P(z, Y) = 1\}.$$

Furthermore, we define $L_P^+ = L_P \cap Z^+$ and $L_P^- = L_P \cap Z^-$. For a cycle C the sets L_C, L_C^+ and L_C^- are defined analogously.

Lemma 1 *Let P be a path starting from a vertex in X and ending at a vertex in X that does not contain any other vertices of X. Then*

$$\chi_c(P) = \chi_c(L_P) + 1 = |L_P^+| - |L_P^-| + 1.$$

Proof. Let P be a path that starts from x and ends at x', where $x, x' \in X$. Furthermore, assume that P does not contain any other vertices of X. It is easily seen that after possibly reversing P we can classify P in one the four cases discussed below (these cases are illustrated in Figure 3).

Case 1: $P = x, y_1, z_1, \ldots, y_t, z_t, y_{t+1}, x'$. Where $y_1, \ldots, y_{t+1} \in Y$, $z_1, \ldots, z_t \in Z$ and $t \geq 1$. In this case $\chi_c(z_i) = +$ for all $i \leq t$ since all z_i are connected to two nodes in Y; namely y_i and y_{i+1}. Therefore, $\chi_c(P) = 2 - (t+1) + t = 1$. Furthermore, $L_P = \emptyset$ as no vertex in Z has any neighbors in X on P.

Case 2: $P = x, y_1, z_1, \ldots, y_t, z_t, x'$. Where $y_1, \ldots, y_{t+1} \in Y$, $z_1, \ldots, z_t \in Z$ and $t \geq 1$. In this case, we note that $\chi_c(z_i) = +$ for all $i < t$ since all such z_i are connected to two nodes in Y. Thus $\chi_c(P) = 2 - t + (t-1) + \chi_c(z_t) = \chi_c(z_t) + 1$. On the other hand $L_P = \{z_t\}$ as all other edges have no neighbors in X on P.

Case 3: $P = x, z_1, y_1 \ldots, z_t, y_t, z_{t+1}, x'$. Where $y_1, \ldots, y_{t+1} \in Y$, $z_1, \ldots, z_{t+1} \in Z$ and $t \geq 1$.

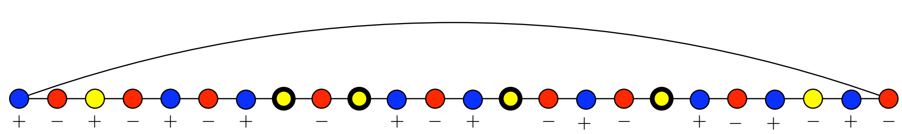


Figure 4: Another way to view Lemma 2. If we ignore the nodes in L_C (shown with bold circles) the labels of all the other nodes on a cycle cancel alternatively.

In this case $\chi_c(z_i) = +$ for all $2 \leq i \leq t$ and therefore, $\chi_c(P) = 2 - t + (t-1) + \chi_c(z_1) + \chi_c(z_{t+1}) = \chi_c(z_1) + \chi_c(z_{t+1}) + 1$ and it is easy to see that $L_P = \{z_1, z_{t+1}\}$.

Case 4: $P = x, z_1, x'$. In this case, $\chi_c(z_1) = -$ as z_1 is connected to two vertices in X. Therefore, $\chi_c(P) = +1$. $L_P = \emptyset$ as both neighbors of z_1 neighbors are in X. ∎

The next Lemma goes a long way in bounding the cycle discrepancy of the labeling χ_c.

Lemma 2 *For any cycle C in G,*

$$\chi_c(C) = \chi_c(L_C) = |L_C^+| - |L_C^-|.$$

Proof. If C does not contain any vertex of X then it must alternate between the vertices of Y and Z. All the Y vertices are labeled $+$ and all the Z vertices must have label $-$ as they have at least two neighbors in Y. Thus $\chi_c(C) = 0$ and in this case $L_C = \emptyset$.

The proof of Lemma 1 applies to a cycle with exactly one vertex in X almost verbatim. The only difference is that $x = x'$ and therefore, all quantities are adjusted by 1 and Case 4 cannot occur.

Now, we can assume $|C \cap X| = t > 1$. Let $x_0, \ldots, x_{t-1}$ be the vertices of X that appear on C and let P_i denote the path from x_i to x_{i+1} (indices here are modulo t). We have

$$\begin{aligned}
\chi_c(C) &= \left(\sum_{i=0}^{t} \chi_c(P_i)\right) - t \\
&= \sum_{i=0}^{t} \chi_c(L_{P_i}) \\
&= \chi_c(L_C).
\end{aligned}$$

The first equality is true because each x_i appears on two paths; namely, P_i and P_{i-1}. Whereas, all other vertices of C appear on exactly one path. The second equality holds by applying Lemma 1 to each P_i. The last equality holds since $L_C = \biguplus L_{P_i}$. ∎

Figure 4 shows the cycle highlighted in Figure 3: when the vertices of L_C are taken out of consideration the labels of the remaining vertices cancel alternatively.

By Lemma 2 for every cycle C, we have

$$|\chi_c(C)| \leq \max(|Z^-|, |Z^+|) \leq |Z| \leq \frac{n}{3}.$$

This gives us an alternate proof of Theorem 3. To improve this bound we set the following parameters:

$$|Z^+| = r, |Y| - |Z| = q \text{ and } |X| - |Y| = k.$$

Our first observation is that the difference in the sizes of Z^- and Z^+ is actually controlled by the difference in sizes of the two sets X and Y.

Lemma 3 *If every vertex of Z has at least one neighbor in X and one neighbor in Y then*

$$|Z^-| - |Z^+| = 3(|X| - |Y|) = 3k.$$

Proof. Counting the number of edges between Z and X in two ways gives us

$$2|Z^-| + |Z^+| = 3|X| - |E(X,Y)|.$$

Similarly, counting the number of edges between Z and Y in two ways yields

$$2|Z^+| + |Z^-| = 3|Y| - |E(X,Y)|.$$

Subtracting these two equations yields the desired result. ∎

The above equation allows us to determine the cardinalities of X, Y, Z, Z^+ and Z^- in terms of r, k and q. We have

$$\begin{array}{rcl} |X| &=& 2r + 4k + q. \\ |Y| &=& 2r + 3k + q, \\ |Z| &=& 2r + 3k, \qquad |Z^+| \;=\; r, \quad |Z^-| \;=\; r + 3k, \end{array}$$

Critically, we obtain that

$$n = 6r + 10k + 2q$$

and we can improve our previous bound.

Theorem 4

$$\text{cycdisc}(G) \le r + 3k \le \frac{3n}{10}.$$

∎

To further improve this bound we tinker with our labeling χ_c. Note that

$$\chi_c^+(C) \le |Z^+| \text{ and } \chi_c^-(C) \le |Z^-|.$$

Since $|Z^-| = |Z^+| + 3k$, it seems wise to change the labels of some vertices in Z^- from $-$ to $+$. This change will tend to reduce the discrepancy in the negative-heavy cycles at the cost of increase in the discrepancy in the positive-heavy cycles. Luckily, we can quantify these gains and losses. We start with a simple fact about two labelings:

Fact 9 *Let $\chi : V \mapsto \{+, -\}$ be a labeling and $W \subseteq V$ such that $\chi(v) = -$ for all $v \in W$. Suppose ρ is defined by switching all the labels of the vertices in W from $-$ to $+$; Formally,*

$$\rho(v) = \begin{cases} \chi(v), & \text{if } v \in V \setminus W \text{ and} \\ -\chi(v) \;=\; +, & \text{if } v \in W. \end{cases}$$

For any $S \subseteq V$ we have,

$$\rho(S) = \chi(S) + 2|W \cap S|.$$

Consequently,

$$\rho^+(S) \le \chi^+(S) + 2|W|.$$

∎

Let us define ρ_c by changing the labels of exactly k vertices in Z^- to a $+$. We now estimate the discrepancy of ρ_c.

Lemma 4

$$\text{cycdisc}(\rho_c) \leq r + 2k.$$

Proof. $\rho_c^+(C) \leq r + 2k$ follows from Fact 9. To show that $\rho_c^-(C) \leq r + 2k$ we have to do a little more work. Intuitively, the negative-heavy cycles are better off if they avoid all the vertices whose labels have been switched to $+$. However, avoiding all these vertices means that a negative-heavy cycle can now only pick up $r + 3k - k = r + 2k$ vertices from Z^-. Formally, let $z_1, \ldots, z_k$ be the vertices whose label has been changed from a $-$ to a $+$. Let C be a cycle and let us assume that after appropriate relabeling $z_1, \ldots, z_t$ do not appear on the cycle C and $z_{t+1}, \ldots, z_j$ appear on the cycle. By Lemma 2 we have

$$\chi_c(C) = |L_C^+| - |L_C^-|,$$

and therefore,

$$\chi_c^-(C) \leq |L_C^-|.$$

Since $L_C^- \subseteq Z^- \setminus \{z_1, \ldots, z_t\}$. Hence

$$\chi_c^-(C) \leq |Z^- \setminus \{z_1, \ldots, z_t\}| \leq r + 3k - t.$$

Furthermore,

$$\rho_c(C) = \chi_c(C) + 2(k - t).$$

Hence we get

$$\rho_c^-(C) \leq r + 3k - 2k + t \leq r + 2k.$$

∎

We have proved the following useful Lemma that we record for future use.

Lemma 5 *Let $G = (V, E)$ be a K_4-free, three-regular graph. Let (X, Y, Z) be a tricoloring of G such that each vertex in $z \in Z$ has at least one neighbor in X and one neighbor in Y. Define,*

$$k = |X| - |Y| \text{ and } q = |Y| - |Z|.$$

Then for the coloring ρ_c we have,

$$\text{cycdisc}(\rho_c) \leq \frac{n + 2k - 2q}{6}.$$

∎

A straightforward application of this Lemma gives us the following bound:

Theorem 5

$$\text{cycdisc}(G) \leq \frac{n}{5}.$$

∎

In order to improve this bound we start with a simple observation about bipartite graphs.

Lemma 6 *Let $H = (A, B, E)$ be a connected bipartite graph with maximum degree three. If α is the number of vertices of degree three in B then*

$$|A| \leq |B| + \alpha + 1.$$

Proof. The number of vertices in H is $|A| + |B|$. Furthermore, we have

$$|E| \leq 3\alpha + 2(|B| - \alpha) = 2|B| + \alpha.$$

As, the graph is connected $|E|$ must be at least $|A| + |B| - 1$. The Lemma follows. ∎

We are ready to give the main argument that will help us prove an optimal upper bound.

Lemma 7 *Let $G = (V, E)$ be a K_4-free, three-regular graph. There is a tricoloring $c = (X, Y, Z)$ of G such that each vertex $z \in Z$ has at least one neighbor in X and one neighbor in Y. Furthermore,*

$$k \leq q + 1$$

where $k = |X| - |Y|$ and $q = |Y| - |Z|$.

Proof. Let (X, Y, Z) be a tricoloring that minimizes $|Z|$. Furthermore, out of all such tricolorings we pick the one that minimizes $k = |X| - |Y|$. We set $q = |Y| - |Z|$.

We may assume that $k > 0$; otherwise, there is nothing to prove. Let

$$\hat{Y} = \{y \in Y : d(y, X) = 3\}.$$

We note that $|\hat{Y}| \leq q$ otherwise; $(X, Z \cup \hat{Y}, Y \setminus \hat{Y})$ is a proper three coloring of G that contradicts the minimality of $|Z|$.

We now examine $G[X \cup Y]$. The fact that $|X| > |Y|$ implies that there must be a component $\mathcal{C}$ of $G[X \cup Y]$, such that $|\mathcal{C} \cap X| > |\mathcal{C} \cap Y|$. Define,

$$\begin{aligned} \mathcal{C}_X &= \mathcal{C} \cap X \text{ and} \\ \mathcal{C}_Y &= \mathcal{C} \cap Y. \end{aligned}$$

We argue that this component must satisfy

$$|\mathcal{C}_X| \geq |\mathcal{C}_Y| + k. \tag{1}$$

Indeed, if $|\mathcal{C}_X| < |\mathcal{C}_Y| + k$, then, we can invert this component to obtain a contradiction; that is, we define

$$\begin{aligned} A &= (X \setminus \mathcal{C}_X) \cup \mathcal{C}_Y, \\ B &= (Y \setminus \mathcal{C}_Y) \cup \mathcal{C}_X. \end{aligned}$$

This yields two sets A and B with $|B| - k < |A| < |B| + k$. Thus (A, B, Z) or (B, A, Z) is a tricoloring that contradicts the minimality of k.

Let $\hat{\mathcal{C}}_Y$ be the vertices in $\mathcal{C}$ that have degree three. By Lemma 6

$$|\mathcal{C}_X| \le |\mathcal{C}_Y| + |\hat{\mathcal{C}}_Y| + 1.$$

Comparing this with the inequality 1, we get $k \le |\hat{\mathcal{C}}_Y| + 1$. Our last observation is $\hat{\mathcal{C}}_Y \subseteq \hat{Y}$, and therefore, we have $|\hat{\mathcal{C}}_Y| \le q$. ∎

Combining Lemma 5 and Lemma 7 we get our main result. It is also easy to see that we can remove the restriction on G being K_4 free.

Theorem 6 *For every three-regular graph G with n vertices we have*

$$\operatorname{cycdisc}(G) \le \frac{n+2}{6}.$$

∎

5 An $O(n^2)$ algorithm for finding a good labeling

The proof of Theorem 3 is algorithmic and can be easily converted into an $O(n^3)$ algorithm. The algorithm takes a three-regular graph, $G = (V, E)$, and outputs labeling ρ_c such that

$$\operatorname{cycdisc}(\rho_c) \le \frac{n+2}{6}.$$

In this section we show that this can also be done in $O(n^2)$ time. We show how to obtain a coloring that satisfies the conditions of Lemma 7. Given such a coloring, an appropriate labeling can be easily obtained in linear time. The algorithm has two phases, which we discuss below.

Phase I: The algorithm starts by finding a tricoloring $c = (X, Y, Z)$ of G promised by Brooks' Theorem. Such a coloring can be found in linear time. Call a $z \in Z$ *blocked* if it is connected to an $x \in X$ and a $y \in Y$ such that x and y belong to the same connected component of $G[X \cup Y]$. Furthermore, define

$$\hat{Y} = \{y \in Y : d(y, X) = 3\}.$$

Our algorithm maintains a coloring c and repeatedly checks if the following conditions are satisfied.

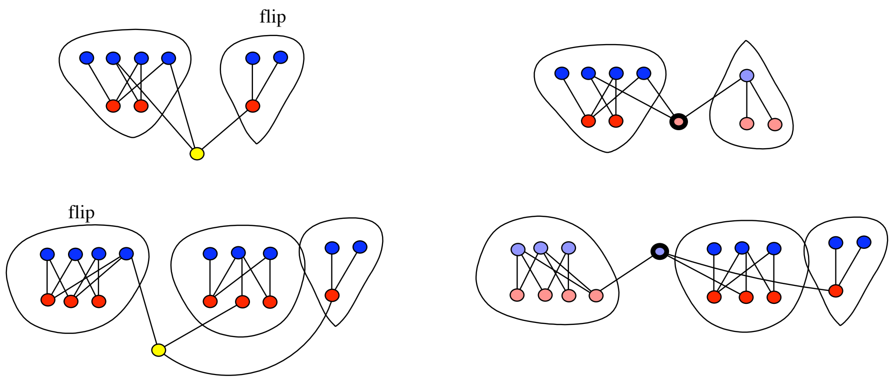


Figure 5: Moving an unblocked vertex z out of the smallest color class: all the possible cases are not shown.

1. Each vertex $z \in Z$ is blocked.
2. $|\hat{Y}| \leq |Y| - |Z|$.

If the first condition is violated by some vertex z, then by inverting some components of $G[X \cup Y]$ the vertex z can be moved into the set X or Y (See Figure 5). In case the second condition fails we exchange Z and $Y \setminus \hat{Y}$ (See Figure 6). In both cases the number of vertices in the smallest class is reduced by at least one. Thus the above steps are applied $O(n)$ times. It is easily seen that both of these steps can be implemented in linear time thus the entire phase takes $O(n^2)$ time.

Phase II: Let $c_0 = (X_0, Y_0, Z_0)$ be the coloring obtained from Phase I. Let

$$k_0 = |X_0| - |Y_0|, q_0 = |Y_0| - |Z_0|, \hat{Y}_0 = \{y \in Y_0 : d(y, Y_0) = 3\} \text{ and } \beta = |\hat{Y}_0|.$$

We note that $\beta \leq q_0$ and all vertices in Z_0 are blocked. We may assume that $k_0 > 1$; otherwise, our algorithm can output (X_0, Y_0, Z_0).

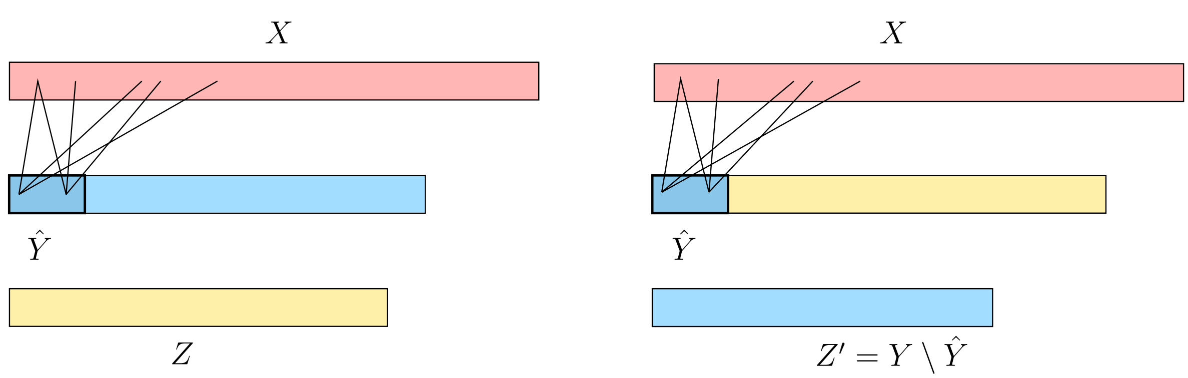


Figure 6: Exchanging Z and $Y \setminus \hat{Y}$ in case $|\hat{Y}| > |Y| - |Z|$: the size of the smallest color class reduces.

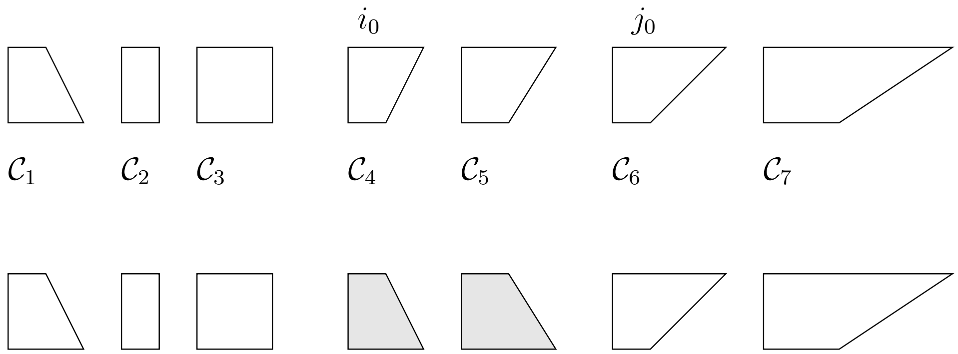


Figure 7: Flipping the connected components in $G[X_0 \cup Y_0]$ as discussed in Case 2

Let $\mathcal{C}_1, \ldots, \mathcal{C}_t$ be the connected components of $G[X_0 \cup Y_0]$ sorted by

$$d_i := |\mathcal{C}_i \cap X_0| - |\mathcal{C}_i \cap Y_0|$$

(See Figure 7). Let i_0 be the smallest index with $d_{i_0} > 0$. For $j \geq i_0$ let X_j and Y_j denote the sets obtained by flipping $\mathcal{C}_{i_0}, \mathcal{C}_{i_0+1}, \ldots, \mathcal{C}_j$. For convenience, we define $X_i = X_0$ and $Y_i = Y_0$ for all $i < i_0$. Let us define, $k_j = |X_j| - |Y_j|$ and $q_j = |Y_j| - |Z_0|$. Note that,

$$k_0 = \cdots = k_{i_0-1} > k_{i_0} > k_{i_0+1} > \cdots > k_t.$$

and

$$q_0 = \cdots = q_{i_0-1} < q_{i_0} < q_{i_0+1} < \cdots < q_t.$$

Let $j_0 \geq i_0$ be the first index such that $|Y_{j_0}| > |X_{j_0}|$ or equivalently k_{j_0} is negative. Note that such an index must exist as $|Y_t| \geq| X|$ and $|X_t| \leq| Y|$. We claim that the tricoloring $(X_{j_0-1}, Y_{j_0-1}, Z_0)$ satisfies constraints of Lemma 5. We discuss two cases:

Case 1: $j_0 = t$. In this case, we have

$$k_{j_0-1} = d_{j_0} + \sum_{i=0}^{i_0-1} d_i - \sum_{i=i_0}^{j_0-1} d_i.$$

As all d_i with $0 \leq i < i_0$ are negative and all d_i with $i_0 \leq i \leq j_0 - 1$ are positive, therefore $k_{j_0-1} \leq d_{j_0}$. We make the critical observation that the vertices of degree three in $\mathcal{C}_{j_0} \cap Y_{j_0-1}$ are a subset of $\hat{Y}_0$. Therefore, there can be at most β vertices of degree three in $\mathcal{C}_{j_0} \cap Y_{j_0-1}$. Thus by Lemma 6 we have $d_{j_0} \leq \beta + 1 \leq q_0 + 1$. Hence $k_{j_0-1} \leq q_0 + 1 \leq q_{j_0-1} + 1$.

Case 2: $j_0 < t$. In this case,

$$0 < k_{j_0} = k_{j_0-1} - 2d_{j_0}.$$

Hence, we have $2d_{j_0} > k_{j_0-1}$. Because d_i's are sorted $d_{j_0+1} \geq d_{j_0}$. Thus,

$$d_{j_0} + d_{j_0+1} \geq 2d_{j_0} > k_{j_0-1}$$

(Here the strictness of the inequality is important). Let β' and β'' be the number of vertices of degree three in $\mathcal{C}_{j_0} \cap Y_{j_0-1}$ and $\mathcal{C}_{j_0+1} \cap Y_{j_0-1}$, respectively. Critically, these vertices are a subset of $\hat{Y}$ and therefore, we have $\beta' + \beta'' \leq \beta$. By Lemma 6, $d_{j_0} \leq \beta' + 1$ and $d_{j_0+1} \leq \beta'' + 1$. Therefore, we have

$$d_{j_0} + d_{j_0+1} \leq \beta + 2 \leq q_0 + 2 \leq q_{j_0-1} + 2.$$

Hence, $k_{j_0-1} \leq q_{j_0-1} + 1$.

Thus in both cases we have $k_{j_0-1} \leq q_{j_0-1} + 1$. Lastly, we recall that all the vertices in Z_0 were blocked in (X_0, Y_0, Z_0). Since $(X_{j_0-1}, Y_{j_0-1}, Z_0)$ is obtained by flipping some connected components of $G[X_0 \cup Y_0]$, therefore, each vertex in $z \in Z_0$ has at least one neighbor in X_{j_0-1} and one neighbor in Y_{j_0-1}. This shows that the tricoloring $(X_{j_0-1}, Y_{j_0-1}, Z_0)$ satisfies the conditions of Lemma 5.

5.1 Cycle discrepancy of maximum degree three graphs

Once we have established an upper bound on the maximum cycle discrepancy of three-regular graphs, it is straightforward to extend a similar bound on graphs with maximum degree three. We can show the following result:

Theorem 7 *If G is a maximum degree three graph on n vertices, then*

$$\text{cycdisc}(G) \leq \frac{n+7}{6}.$$

Proof. We proceed by induction on the number of vertices in G. The base case $n = 5$ being easy to handle. Since cycle discrepancy is monotone, it suffices to study maximum degree three graphs that are edge maximal. By Fact 3 we can further restrict our attention to graphs that are also two connected. Let G be an edge maximal, two connected graph with maximum degree three.

Let S be the vertices of degree at most two in G. It is easily seen that such a graph falls into the following two cases.

Case 1: $S = \{a, b\}$ and $a \sim b$. Furthermore, there are two distinct vertices u and v such that $a \sim u$ and $b \sim v$. Removing a and b from the graph and adding the edge $\{u, v\}$ in the graph (if it is not already present) gives us a graph G' such that

$$\text{cycdisc}(G') = \text{cycdisc}(G).$$

Thus by induction, we have

$$\text{cycdisc}(G) \leq \frac{n+5}{6}.$$

Case 2: $S = \{a\}$ and $d(a) = 2$. Let $N(a) = \{u, v\}$. If $\{u, v\}$ is not an edge in G then we remove a from G and add the edge $\{u, v\}$ to obtain a graph G'. G' is three-regular and has cycle discrepancy at most $\frac{n+1}{6}$. It is easily seen that

$$\text{cycdisc}(G) \leq \text{cycdisc}(G') + 1 \leq \frac{n+7}{2}.$$

On the other hand, if $\{u, v\}$ is already an edge in G, then we can remove $\{a, u, v\}$ from G to obtain G'. It is again easily seen that $\mathrm{cycdisc}(G) \leq \mathrm{cycdisc}(G') + 1$. Now, we observe that in this case G' is a graph that belongs to Case 1 and hence has cycle discrepancy at most $\frac{(n-3)+5}{6}$. The bound on the cycle discrepancy of G follows. ∎

6 Cycle discrepancy of bounded degree graphs

An interesting question is to find a tight upper bound on the cycle discrepancy of bounded degree graphs. In this section, we make some preliminary observations.

We define the concept of st-discrepancy that will be useful in constructing graphs of high cycle discrepancy. Let $H = (U, E)$ be a graph with $|U| \geq 2$ and $s, t \in U$ be any two fixed vertices. Let $\mathcal{P}_{s,t}(H)$ denote the set of all st-paths in H. For any labeling $\chi : U \mapsto \{+, -\}$ we let

$$st\text{-disc}(\chi) = \max_{P \in \mathcal{P}_{s,t}} \chi^+(P) + \max_{Q \in \mathcal{P}_{s,t}} \chi^-(Q).$$

Furthermore, we define

$$st\text{-disc}(H) = \min_{\chi : \mapsto \{+,-\}} st\text{-disc}(\chi).$$

A consequence of the definition of st-discrepancy is that for any labeling χ there exists a pair of paths P, Q such that

$$\chi^+(P) + \chi^-(Q) \geq st\text{-disc}(H).$$

Let $\mathrm{cycle}_k(H, s, t)$ be the graph that consist of k copies of H connected in a cycle as shown in Figure 8. The following Lemma allows us to construct high discrepancy graphs on arbitrarily large number of vertices, provided H has large st-discrepancy.

Lemma 8 *Let* $st\text{-disc}(H) = d$ *then*

$$\mathrm{cycdisc}\left(\mathrm{cycle}_k(H, s, t)\right) \geq \frac{dk}{2} = \frac{d}{2|U|} n.$$

Where $n = k|U|$ *is the number of vertices in* $\mathrm{cycle}_k(H, s, t)$*.*

Proof. Let χ be a labeling of vertices of $\mathrm{cycle}_k(H, s, t)$. This labeling naturally induces a labeling on all the copies of H. We can find k pairs of paths P_i, Q_i in the i-th copy of H such that

$$\chi^+(P_i) + \chi^-(Q_i) \geq d.$$

The paths P_i define a cycle in $\mathrm{cycle}_k(H, s, t)$ that we denote by C_P. Similarly, the paths Q_i define a cycle that we denote by C_Q. We have,

$$\begin{aligned}
\chi^+(C_P) + \chi^-(C_Q) &= \sum_{i=1}^{k} \chi^+(P_i) + \sum_{i=1}^{k} \chi^-(Q_i) \\
&= \sum_{i=1}^{k} \chi^+(P_i) + \chi^-(Q_i) \\
&= dk
\end{aligned}$$

Thus, either $\chi^+(C_P) \geq dk/2$ or $\chi^-(C_Q) \geq dk/2$. ∎

We can find the st-discrepancy of the complete graph very easily:

Lemma 9 *For any two distinct vertices s and t of K_d, we have*

$$st\text{-disc}(K_d) = d - 2.$$

Proof. Let $\chi : V_d \mapsto \{+,-\}$ and P be any path from s to t that contains all the vertices $v \in V_d \setminus \{s,t\}$ such that $\chi(v) = +$. Similarly, we let Q be the path from s to t that contains all the vertices v with $\chi(v) = -$. Now, it is easy to see that

$$\chi^+(P) + \chi^-(Q) = |U \setminus \{s,t\}| = d - 2$$

∎

Our previous theorem shows the following result.

Theorem 8

$$\text{cycdisc}\left(\text{cycle}_k\left(K_d, s, t\right)\right) = (d-2)k = \frac{(d-2)}{2d}n.$$

Where $n = dk$ is the number of vertices in $\text{cycle}_k(K_d, s, t)$.

It is also quite simple to estimate the st discrepancy of an odd cycle.

Lemma 10 *Let s and t be two adjacent vertices on C_{2d+1},*

$$st\text{-disc}(C_{2d+1}) = 1.$$

∎

Theorem 9

$$\text{cycdisc}\left(\text{cycle}_k(C_{2d+1}, s, t)\right) \geq k/2 = \frac{n}{2(2d+1)}.$$

Where $n = (2d+1)k$ is the number of vertices in $\text{cycle}_k(K_d, s, t)$.

An interesting question is to study the discrepancy of a bounded degree graph and bounded degree graph with girth g. More precisely, let $\mathcal{G}_{k,n}$ and $\mathcal{G}_{g,k,n}$ denote the set of all n vertex maximum degree k graphs and the set of all n vertex maximum degree k graphs with girth g, respectively. Let

$$f(k,n) = \max_{G \in \mathcal{G}_{k,n}} \text{cycdisc}(G) \text{ and } h(g,k,n) = \max_{G \in \mathcal{G}_{g,k,n}} \text{cycdisc}(G).$$

Furthermore, let us define:

$$c_k = \overline{\lim}_{n \to \infty} \frac{f(k,n)}{n} \text{ and } c_{g,k} = \overline{\lim}_{n \to \infty} \frac{f(g,n,k)}{n}.$$

By Theorem 8 and Theorem 9 we have

$$c_d \geq \frac{d-2}{2d} \text{ and } c_{2d+1,3} \geq \frac{1}{2(2d+1)}.$$

The main result of this paper states that $c_3 = c_{3,3} = \frac{1}{6}$. This problem is open for all other values of k and g.

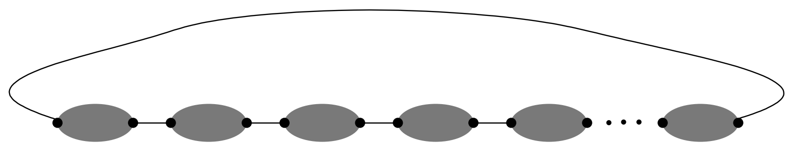


Figure 8: The graph $\text{cycle}_k(H, s, t)$.

7 Conclusion

We have introduced cycle discrepancy as a new graph invariant. We have also found an optimal bound on the cycle discrepancy of three-regular graphs. It would be very interesting to study this invariant from an extremal point of view. A similar invariant, path discrepancy, can also be defined. All our results on cycle discrepancy carry over to path discrepancy with an adjustment of some additive constants. Many interesting problems regrading these two invariants can be posed. We conclude our discussion with the following three concrete problems, whose solution, we believe, will involve developing more interesting methods to study cycle discrepancy.

Problem 1 *Give a tight upper bound on the cycle discrepancy of maximum degree four graphs; that is, determine* c_4*.*

Problem 2 *Give a tight upper bound on the cycle discrepancy of maximum degree three graphs the do not contain any triangles; that is determine* $c_{4,3}$*.*

Problem 3 *Develop an approximation algorithm for estimating the cycle discrepancy of a maximum degree three graph.*

References

[1] J. Beck and W. L. Chen, *Irregularities of distribution*, Cambridge University Press, New York, 1987.

[2] J. Beck and W. Fiala, *Integer making theorems*, Discrete Applied Mathematics **3** (1981), 1–8.

[3] B Bollobás, *Modern graph theory*, Springer-Verlag, 1998.

[4] R. L. Brooks, *On coloring the nodes of a network*, Proc. Cambridge Philos. Soc. **37** (1941), 194–197.

[5] B. Chazelle, *The discrepancy method: Rrandomness and complexity*, Cambridge University Press, New York, 2000.

[6] Evangelos Kranakis, Danny Krizanc, Berthold Ruf, Jorge Urrutia, and Gerhard Woeginger, *The vc-dimension of set systems defined by graphs*, Discrete Applied Mathematics **77** (1997), no. 3, 237–257.

[7] J. Matousek, *Geometric discrepancy: An illustrated guide. algorithms and combinatorics*, Springer, Berlin, 1999.

[8] Marcus Schaefer, *Deciding the Vapnik-Chervonenkis dimension is Σ_p^3-complete*, Journal of Computer and System Sciences **58** (1999), no. 1, 177–182.

[9] J. Spencer, *Six standard deviations suffice*, Transactions of the American Mathematical Society **289** (1985), no. 2, 679–706.