Average Distance in Coloured Graphs

Peter Dankelmann[1]
School of Mathematical and Statistical Sciences
University of Natal, Durban, 4041, South Africa

Wayne Goddard[1]
School of Geological and Computer Sciences
University of Natal, Durban, 4041, South Africa

Peter Slater[1]
Department of Mathematics and Statistics
University of Alabama in Huntsville, Huntsville AL 35899, USA

Abstract

For a graph G where the vertices are coloured, the *coloured distance* of G is defined as the sum of the distances between all unordered pairs of vertices having different colours. Then for a fixed supply s of colours, $d_s(G)$ is defined as the minimum coloured distance over all colourings with s. This generalizes the concepts of median and average distance. In this paper we explore bounds on this parameter especially a natural lower bound and the particular case of balanced 2-colourings (equal numbers of red and blue). We show that the general problem is NP-hard but there is a polynomial-time algorithm for trees.

1 Introduction

Hulme and Slater [6] introduced the following facilities location problem. Given a connected graph $G = (V(G), E(G))$ on n vertices, some number k of facilities is specified. The facilities are to be placed on the graph. At each vertex there is to be either a person or a facility. Each person must visit each facility (perhaps because each facility is different). The question is how to place these facilities. That is, one must place the facilities to minimize the sum over all pairs (u, v), where u is a facility and v is not, of the distance $d(u, v)$ (measured as the number of edges in the shortest path connecting them).

The placement of the facilities may be thought of as a colouring of the vertices with two colours. This generalizes to "*coloured distance*". A colouring of a graph

[1]Research supported in part by South African Foundation for Research Development

is a map c with domain $V(G)$; it may also be thought of as a partition of the vertex set. Then the *coloured distance* of a graph G with a given colouring c of the vertices is defined as

$$d(G;c) = \sum_{\substack{\{u,v\}\subseteq V(G)\\ c(u)\neq c(v)}} d(u,v).$$

If every vertex is a different colour, one obtains the total distance or simply the *distance* of the graph. We will denote this by $d(G)$. The distance is also called the *transmission* of the graph or the *Wiener index*.

In general, there is a fixed *supply* s of colours; this is a partition of the natural number n. We are interested in the minimum coloured distance, given that the cardinalities of the colour classes of c are determined by s. We define

$$d_s(G) = \min_{c \text{ is } s\text{-colouring of } G} d(G;c).$$

For $s = (1, 1, \ldots, 1)$ one obtains the distance of the graph again. Another special case is $s = (1, n-1)$: the *median* problem. These two versions have been well studied; see, for example, the book [1]. In particular, there is a polynomial-time algorithm to calculate the "distance" of the median vertex.

The problem we study is to minimize the coloured distance. Since the supply of colours is fixed, this is equivalent to minimising the average distance. Thus the problem is called the minimum average distance problem (MAD) in [9].

In particular, for a given colouring c we define $\mu(G;c)$ to be the average distance between differently coloured vertices; that is, $d(G;c)$ divided by the number of unordered pairs of differently coloured vertices. We write $\mu(G)$ to be the average distance between two distinct vertices (that is, the value $\mu(G;c)$ where c assigns every vertex a different colour).

This problem has been studied in [5, 6, 8, 9]. For example, in [6] the coloured distance of a path with two colours was given.

Theorem 1 *Consider the path P_n with $s = (a, b)$ where $n = a + b$ and $a \leq b$. Say there are a reds. If the vertices are numbered 1 up to n, then one optimal arrangement is to place the reds at positions $m, m \pm 2, m \pm 4, \ldots, m \pm (a-1)$ if a is odd, and $m \pm 1, m \pm 3, m \pm 5, \ldots, m \pm (a-1)$ if a is even, where $m = \lceil n/2 \rceil$.*

The optimal placement of facilities on grids was studied by Meiers and Slater [8]. Also, the case of $s = (2, 2, \ldots, 2)$ has been studied by Gerstel and Zaks [5] in connection with routing in trees. In this case, one must pair up the vertices such that the sum of the distances between each pair is as large as possible. Their result was that in a tree, a placement is optimal iff for each pair the members are in different branches from the median.

In this paper we look at a natural lower bound on the coloured distance and use it to determine the coloured distance for cycles and some products of cycles. We focus next on the problem for balanced 2-colouring and provide tight bounds for both general graphs and trees. These ideas are then used to show that for any given supply of colours the path is the worst case. We conclude by showing that, while the problem is NP-hard in general, there is a polynomial-time algorithm for two colours on trees based on repeated swopping.

2 Ideal Colourings

Consider a graph G with colouring c. For a subset $R \subseteq E(G)$, we define the *transition value* of R to be the number of unordered pairs of vertices which have different colours and occur in different components of $G-R$. This will be denoted $\partial_c(R)$. (If R is not a cut-set, then the value of the parameter is 0.)

The transition value of R is a lower bound on the number of times edges of R are used if one looks at the shortest paths connecting coloured vertices. So if $\tau = (R_1, R_2, \ldots, R_m)$ is a partition of $E(G)$, then

$$d(G; c) \geq \sum_{i=1}^{m} \partial_c(R_i).$$

Furthermore, if we are told how many vertices there are of each colour, that is, that $\|c\| = s$, we can establish a lower bound for $\partial_c(R)$. This lower bound can be obtained by a purely arithmetical calculation and depends only on s and the orders of components in $G - R$. We denote this lower bound by $\partial_s(R)$.

For example, if we have a colouring with $2k$ reds and $2k$ blues (so that $n = 4k$ and $s = (2k, 2k)$), and $G - R$ consists of two components each with $2k$ vertices, then $\partial_c(R) \geq 2k^2 = \partial_s(R)$, attained (only) by placing k reds on each side of the cut.

Hence we have established:

Lemma 2 *Let G be a graph and s a supply of colours. Consider a partition $(R_1, \ldots, R_m)$ of the edge set of G. Then:*

$$d_s(G) \geq \sum_{i=1}^{m} \partial_s(R_i).$$

We say that a colouring c with supply s is *ideal* with respect to a particular edge-partition if equality is attained in the above bound by this colouring. That is, $d(G; c) = \sum_{i=1}^{m} \partial_s(R_i)$. A trivial example is the complete graph with the entire edge-set as the only cut in the partition: any colouring is ideal with respect to the edge-partition.

We look now at some examples. We will mostly consider the case where there are two colours; in such a case the two colours will be red and blue.

2.1 Paths

For paths consider the edge partition formed by each edge in a singleton set. Then, it is not too hard to argue that the path has an ideal colouring for each choice of colour supply. See [9].

2.2 Even Cycles

Consider first an *even cycle* C_n with an even number of reds and an even number of blues; say $n = 2m$ and $s = (2j, 2k)$. Then there is a natural partition $\tau = (R_1, R_2, \ldots, R_m)$ of the edges where each set R_i is a pair of opposite edges.

Each $G - R_i$ consists of two paths, each with m vertices. If there are $j - x$ reds in the first path, then there are $k + x$ blues in the first path, $j + x$ reds in the second and $k - x$ blues in the second, so that the transition value is $(j-x)(k-x)+(j+x)(k+x) = 2jk+2x^2$; this is minimised for $x = 0$. It follows that $\partial_s(R_i) = 2jk$, which is achieved only when each path has j red vertices and k blue vertices. Thus the optimality requirement is that, given any split of the cycle into two equal paths, the distribution of reds must be balanced.

We call a colouring an *antipodal colouring* if diametrically opposite vertices receive the same colour. An antipodal colouring always produces the desired balance. Hence an antipodal colouring is ideal and is therefore optimal.

On the other hand, it is easily shown that only the antipodal colorings are optimal. (Consider rotating the cut one edge around the cycle. This transfers one vertex from left to right and one from right to left. To preserve the balance, these vertices must have the same colour.)

One also obtains ideal colourings if the numbers of reds and blues are both odd. The requirement for optimality is that: given any split of the cycle into two equal paths, the distribution of reds between the two paths is as equal as possible. This is always attainable: for example, do an antipodal colouring except for one pair. However, this is not the only ideal colouring; for example, if the reds and blues are equinumerous, then the bipartite colouring is also ideal.

This result for two colours generalises immediately to any colour supply for an even cycle as follows. By a calculation similar to that above and elementary calculus, the optimality requirement is that given any split of the cycle into two equal paths, the distribution of each colour must be as balanced as possible. An ideal colouring can be constructed as follows. Say that a pair of antipodal vertices is good if they have the same colour. Then take any colouring with a maximum number of good pairs—the number of bad pairs is exactly half the number of colours with an odd supply.

2.3 Cartesian Products

For graphs G and H we let $G \times H$ denote the cartesian product of G and H. Given partitions τ_G of $E(G)$ and τ_H of $E(H)$, there is a natural partition τ of $E(G \times H)$. Namely, for each set R of τ_G define R' as the union of all R in each copy of G, and for each set S of τ_H define S'' as the union of all S in each copy of H. Then τ consists of all R' for $R \in \tau_G$ and all S'' for $S \in \tau_H$.

Furthermore, we define a *splitter* of a graph F as any set $T \subseteq E(F)$ such that $F - T$ consists of two components and these components have the same order. If τ_G and τ_H both consist entirely of splitters, then so does τ.

For example, consider the n-dimensional hypercube with $s = (2^{n-1}, 2^{n-1})$. Then the edge partition τ is the one where the 2^{n-1} edges in each direction form a subset. This partition consists of splitters. Then a colouring is ideal iff each $(n-1)$-dimensional subcube receives equal numbers of red and blue. For example, the bipartite colouring is ideal and hence optimal. But so too is any colouring in which antipodal vertices receive the same colour. (This follows because the requirement for ideal colouring is that in each half of the split there are equal numbers of each colour, and antipodal vertices are always in opposite halves of the split.)

A similar argument holds for any product of even cycles and K_2's with an even number of each colour: any colouring where antipodal vertices receive the same colour is ideal.

Another example is the grid formed by the product of two even paths using equal numbers of red and blue. Then any colouring in which an equal number of reds and blues are placed in each row and in each column is good since it is ideal. See [8].

There need not be an ideal colouring, however. For example, consider the grid $P_3 \times P_3$ with $s = (2, 7)$. If we consider the four natural cuts, then the two reds should always be in the larger component, but this cannot be simultaneously realised.

2.4 Odd Cycles

To handle odd cycles, one needs a variant of ideal colourings. Consider the cycle C_n, where n is odd, and $s = (2j + 1, 2k)$. We use the word element to refer to both vertices and edges. (Note that a path of length ℓ has $2\ell + 1$ elements.) Let $\tau = (R_1, R_2, \ldots, R_n)$ be a partition of the elements of C_n where each R_i consists of a vertex and its opposite edge.

Then we may bound from below the total number of times that an element of the pair R_i is used in the collection of shortest paths connecting vertices in $C_n - R_i$ with different colours. If one sums up this lower bound over all n pairs R_i one has a lower bound on what one might call the "coloured element-distance". To convert to a lower bound on the coloured distance, we subtract 1 for every path (every pair of differently coloured vertices), and then divide by 2. We omit the details.

We define a colouring of C_n as *near-antipodal* if whenever one contracts out one vertex the result is an antipodal colouring of C_{n-1}. It is not hard to show that a near-antipodal colouring achieves the lower bound, and is therefore optimal.

3 Balanced Colourings

We study next *balanced colourings*; that is, those colourings where there are two colours and an equal supply of them. One can ask for bounds in terms of the total distance of the graph. These are better stated by considering the average distance.

Theorem 3 *Let G be a connected graph of even order n and let c be an optimal balanced 2-colouring of $V(G)$. Then*

$$\frac{2n-2}{3n-4}\,\mu(G) \leq \mu(G;c) \leq \mu(G),$$

and this bound is sharp.

PROOF.To prove the upper bound on $\min \mu(G;c)$, take a balanced colouring c of G at random. By linearity of expectation, the average value of $\mu(G;c)$ taken over all c is precisely $\mu(G)$. Hence, there exists a c with average coloured distance at most $\mu(G)$. Equality in the upper bound holds if and only if every balanced colouring c of G gives the same value of $d(G;c)$; for example G the complete graph or a star.

The key to the lower bound on $\mu(G;c)$ is an estimate of the distance between two vertices of the same colour. Suppose that the two colour classes are B and R. Let $x, y \in B$. Then for any vertex $z \in R$ we have $d(x,y) \leq d(x,z) + d(y,z)$ by the triangle inequality and since the graph is undirected. It follows that

$$d(x,y) \leq \frac{2}{n} \sum_{z \in R} \left(d(x,z) + d(y,z)\right).$$

If we now sum over all unordered pairs of vertices in B, we get that

$$\sum_{\{x,y\} \subseteq B} d(x,y) \leq \frac{2}{n} \sum_{\{x,y\} \subseteq B} \sum_{z \in R} \left(d(x,z) + d(z,y)\right) = \frac{2}{n}\left(\frac{n}{2} - 1\right) \sum_{w \in B,\, z \in R} d(w,z),$$

as for each vertex $w \in B$ the value $d(w,z)$ appears in the middle summation once for each choice of the other vertex in B. The sum in the last expression is by definition $d(G;c)$, so it follows that

$$\sum_{\{x,y\} \subseteq B} d(x,y) \leq \left(1 - \frac{2}{n}\right) d(G;c).$$

An analogous estimate holds for the distance between vertices in R. So

$$\begin{aligned} d(G) &= \sum_{\{x,y\} \subseteq V(G)} d(x,y) \\ &= \sum_{x \in R,\, y \in B} d(x,y) + \sum_{\{x,y\} \subseteq R} d(x,y) + \sum_{\{x,y\} \subseteq B} d(x,y) \\ &\leq \left(3 - \frac{4}{n}\right) d(G;c). \end{aligned}$$

Since $d(G) = \mu(G)n(n-1)/2$ and $d(G;c) = (n^2/4)\mu(G;c)$, simple arithmetic shows that we obtain the desired inequality. Equality holds for G a balanced complete bipartite graph with a proper colouring. QED

3.1 Trees

For a tree the bound is much sharper. In fact, every tree of even order has an ideal balanced 2-colouring, and the average distance of this 2-colouring is intimately linked to the average distance of the uncoloured tree.

Consider the following algorithm:

Algorithm 4 *For a tree of even order: Repeatedly choose vertices u, v such that either they are end-vertices at distance 2 or they are adjacent with degree sum at most 3. Colour u and v with different colours and delete.*

Theorem 5 *The above algorithm produces a balanced ideal 2-colouring (with respect to the edge partition where each edge is by itself).*

PROOF.We need to show that, for each edge e of the tree, the numbers of red and blue on the one side of the edge differ by at most one, and similarly on the other side.

We do this by induction on the size of the tree. Consider the pruning step in the above algorithm. If e is a leaf edge, then the transition value of e is the same for any balanced colouring. So suppose e is not a leaf-edge. Then the vertices u and v being deleted are on the same side of e; thus, if the colouring without u and v is ideal, the colouring with u and v is ideal. QED

For a tree T of even order we define an edge to be an *even (odd) split* if there is an even (odd) number of vertices on both sides of the edge. Let $z(T)$ denote the number of odd splits. We will need the following simple lemma.

Lemma 6 *For any tree of even order n, $n/2 \leq z(T) \leq n-1$.*

PROOF.For the lower bound we induct on n. The result is true if $n = 2$; so assume $n \geq 4$. Note that an edge incident with an end-vertex is an odd split.

If there are two end-vertices with the same neighbour, then removing both vertices preserves the parity of the remaining edge splits; so the result follows by the inductive hypothesis. Otherwise, there is an end-vertex v whose neighbour w has degree 2. The two edges incident with w have opposite parities, and removing both v and w preserves the parity of the remaining edge splits; so the result follows by the inductive hypothesis.

The two extreme cases include the paths and the stars. QED

Theorem 7 *Let T be a tree of even order n. If c is an optimal balanced 2-colouring of T, then*

$$\frac{n-1}{n}\mu(T) + \frac{1}{n-1} \leq \mu(T;c) \leq \frac{n-1}{n}\mu(T) + \frac{2}{n}$$

and this bound is sharp.

Proof.By Theorem 5 we know that an optimal balanced 2-colouring is ideal. Consider any edge e where the components of $T-e$ have orders x and $n-x$. If x is even, then the contribution of e to $d(T)$ is $x(n-x)$ and the contribution to $d(T;c)$ is $x(n-x)/2$. If x is odd, then the contribution of e to $d(T)$ is $x(n-x)$ and the contribution to $d(T;c)$ is $(x-1)(n-x-1)/4+(x+1)(n-x+1)/4 = x(n-x)/2+1/2$.

So

$$d(T;c) = \frac{d(T)}{2} + \frac{z(T)}{2}.$$

Since $d(T) = \binom{n}{2}\mu(T)$ and $d(T;c) = (n^2/4)\mu(T;c)$, the bounds follow from the above lemma.

Paths achieve the lower and stars the upper bound. QED

4 Paths are the Worst

Another consequence of Algorithm 4 is that the path is the worst for balanced 2-colourings:

Theorem 8 *If c is an optimal balanced 2-colouring of a tree T, then $d(T;c) \leq n(n^2+2)/12 = d(P_n;c)$.*

Proof.The proof is by induction on n. The case $n=2$ is trivial. Let $u,v \in V(T)$ be as in Algorithm 4 and let $S = \{u,v\}$. Then since u and v have different colours, their combined coloured distance is maximised if $T-S$ is a path (on $n-2$ vertices). So their total distance is at most $2 + \sum_{i=1}^{n-2} i = (n^2-3n+6)/2$ if u and v are both end-vertices, and at most $n/2 + \sum_{i=1}^{n-2} i = (n^2-2n+2)/2$ otherwise. The latter is larger since $n \geq 4$. Thus,

$$d(T;c) \leq (n-2)((n-2)^2+2)/12 + (n^2-2n+2)/2 = n(n^2+2)/12,$$

as required.

One can verify that for the path P_n on n vertices $d(P_n; c) = n(n^2 + 2)/12$. QED

We show next that for any distribution of 2 colours the path is the worst. While the result is not surprising, there does not seem to be a short proof.

Suppose the colour supply is k reds and $n - k$ blues, with $k \leq n/2$. Then our strategy is as follows. We choose a "central" subtree H with $2k$ vertices; this we colour with k reds and k blue, while the remaining vertices are coloured blue. To get this to work, we choose for H an ideal colouring with a few extra properties.

We discuss these properties next. We define the *red distance* of a vertex v in a graph G, $dr(v, G)$, as the sum of the distances from v to all red vertices, regardless of whether v is red or blue.

Lemma 9 *Let T be a tree of even order n.*
a) For any one specified vertex w there exists an ideal balanced 2-colouring such that the red distance is at most $(n^2 - 2n)/4$ for w and at most $n^2/4$ for all other vertices.
b) If n is a multiple of 4 then there exists an ideal balanced 2-colouring such that the red distance is at most $(n^2 - n)/4$ for all vertices. If n is not a multiple of 4, then there exists an ideal balanced 2-colouring such that the red distance is at most $(n^2 - n + 2)/4$ for all vertices with equality for at most one vertex.

PROOF.a) Follow the algorithm for an ideal colouring given in Algorithm 4. For each pair u and v chosen by the algorithm, colour the one nearer w with red and the other with blue. We prove the bounds on the red distance by induction on the order of the tree. Consider first the bound for any vertex x.

Look at the pruning step when u and v are removed. If $x \notin \{u, v\}$, then in $G - \{u, v\}$ the vertex x has red distance at most $(n - 2)^2/4$ while the distance between x and whichever of u and v is red is at most $n - 1$. It follows that the red distance of x in G is at most $(n - 2)^2/4 + n - 1 = n^2/4$. On the other hand, if $x \in \{u, v\}$, say $x = u$, then let y be the vertex of $G' = G - \{u, v\}$ nearest x. Then if x and v are nonadjacent, it follows that the red distance of x is at most the red distance of y in G' plus an extra 1 for each red vertex of G' plus 2 to get to v. So

$$\text{red distance of } x \ \leq (n - 2)^2/4 + (n/2 - 1) + 2 = n^2/4 - n/2 + 2,$$

which is at most $n^2/4$ since $n \geq 4$ (forced by u and v being nonadjacent). And if x and v are adjacent, it follows that the red distance of x is at most the red distance of y in G' plus an extra 2 for each red vertex of G' plus 1 to get to v. So

$$\text{red distance of } x \ \leq (n-2)^2/4 + 2(n/2-1) + 1 = n^2/4.$$

The bound for w is proved in a similar way and we omit the details.

b) We use induction on n. The cases $n = 2$ and $n = 4$ are easily checked. We use a slight variation of the algorithm.

Case 1: Assume there are two end-vertices u and v with common neighbour w. Then colour u red and v blue, and colour $T - \{u, v\}$ inductively. Let x be a vertex not in $\{u, v\}$. Since the diameter of T is at most $n-2$, the red distance of x in T is at most

$$((n-2)^2 - (n-2) + 2)/4 + (n-2) = (n^2 - n)/4,$$

if n a multiple of 4. A similar calculation unfolds if n is not a multiple of 4: $dr(x, T) \leq (n^2 - n - 2)/4$. On the other hand, the red distance of u or v is at most $n/2 + 1$ plus the red distance of w in $T - \{u, v\}$. So we are done.

Case 2: Assume no two end-vertices have a common neighbour. Then there are end-vertices a and z whose respective neighbours b and y both have degree 2 (for example, take a and z to be the endpoints of a diametrical path). Colour a and z red, b and y blue, and colour $T - \{a, b, y, z\}$ inductively.

Let v be a vertex not in $\{a, b, y, z\}$. Since the eccentricity of v in T is at most $n-3$, the red distance of v in T is at most

$$((n-4)^2 - (n-4))/4 + 2(n-3) < (n^2 - n)/4,$$

if n a multiple of 4. A similar calculation unfolds if n is not a multiple of 4: $dr(x, T) < (n^2 - n + 2)/4$.

On the other hand, if $v \in \{a, b, y, z\}$ the red distance of v in T is at most $n - 1$ (to the rest of $\{a, b, y, z\}$) plus $2(n/2 - 2) + dr(T', c)$ where c is the closest vertex of T'. Thus

$$dr(v, T) \leq ((n-4)^2 - (n-4))/4 + 2n - 5) = (n^2 - n)/4,$$

if n is a multiple of 4. If n is not a multiple of 4 we obtain that $dr(v, T) \leq (n^2 - n + 2)/4$; equality requires that v be either a or z, and that c have equality in the bound in T'. So equality can only hold for one vertex. QED

Theorem 10 *Let T be a tree of order n and let $1 \le k \le n/2$. Then there exists a $(k, n-k)$-colouring c of $V(T)$ with blue B and red R such that*

$$d(T;c) \le \begin{cases} k(3n^2 - 4k^2 + 4)/12 & \textit{if } n \textit{ is even} \\ k(3n^2 - 4k^2 + 1)/12 & \textit{if } n \textit{ is odd.} \end{cases}$$

PROOF.We already know this for $k = n/2$. (Theorem 8.) So assume $k < n/2$.

Let P be a diametrical path in T. Pick $2k$ vertices of T as follows: take them in increasing eccentricity subject to the constraint that when vertices have equal eccentricity the vertices of P have preference. Let H be the subgraph of T induced by these $2k$ vertices. Note that H is connected.

We define the *weight* of a vertex $w \in V(H)$ as the number of vertices outside H for which w is the closest vertex of H. For a vertex x we denote the closest vertex of H by $\hat{x}$. There are two cases:

Case 1: Assume there is a vertex w of H of weight at least $n/2 - k$. Then colour H with the colouring from Lemma 9(b) (with the vertex w designated). Extend the colouring of H to a colouring of T by colouring every vertex not in H blue.

Let x be a vertex outside H. Then its coloured distance (that is, to its red neighbours) is at most k times the distance from x to $\hat{x}$ plus the total red distance in H of $\hat{x}$. Thus

$$d(T;c) \le d(H;c) + k \sum_{x \notin H} d(x, \hat{x}) + \sum_{x \notin H} dr(\hat{x}, H).$$

For at least c vertices x it holds that $\hat{x} = w$ and, by the choice of colouring of H, $dr(\hat{x}, H) \le k^2 - k$. For the remaining $n - 2k - c$ vertices x it holds that $dr(\hat{x}, H) \le k^2$. Thus

$$\sum_{x \notin H} dr(\hat{x}, H) \le k(n/2 - k)(2k - 1).$$

On the other hand, by considering the diametrical path P it follows that if there is a vertex at distance l from $V(H)$, then both sections of $P - V(H)$ contain a vertex of distance $l - 1$ from $V(H)$. So

$$\sum_{x \notin H} d(x, \hat{x}) \le 1+1+2+2+3+\ldots+\lceil (n/2-k) \rceil = \begin{cases} (n-2k+2)(n-2k)/4, & n \text{ even} \\ (n-2k+2)(n-2k+1)/4, & n \text{ odd.} \end{cases}$$

For n even, we thus obtain:

$$d(T;c) \leq k(2k^2+1)/3+k(n/2-k)(2k-1)+k(n-2k+2)(n-2k)/4 = k(3n^2-4k^2+4)/12.$$

The calculations for n odd are similar.

Case 2: Assume every vertex of H has weight less than $n/2-k$. We colour H with the colouring from Lemma 9(a) and extend to a colouring of T by colouring every vertex not in H blue.

The calculations are almost identical to the above case: we again bound $d(T;c)$ as the sum of three quantities. We obtain the exact same upper bounds for each of the quantities, though the calculation of the bound on $\sum_{x \notin H} dr(\hat{x}, H)$ is slightly different as it uses the fact that at most half of the x can have $dr(\hat{x}, H) = k^2 - k + \frac{1}{2}$. QED

So we have shown that for 2 colours the path is the worst graph if the order is specified.

If the order and size are specified, then the worst graph for the (total) distance case is the path-complete graph [10]. This graph is not always the extremum for coloured distance, though; for example, the worst graph for median where the size equals the order is the cycle.

If the order and minimum degree are specified, we know very little. For (total) distance, the problem is open: Graffiti [3] conjectured that for a r-regular graph on n vertices the average distance is at most n/r. Recently, Kovider and Winkler [7] proved that, for a (not necessarily regular) graph G of minimum degree δ, it holds that $\mu(G) \leq n/(\delta+1)+2$. Similar bounds for triangle-free and C_4-free graphs were obtained in [2]. The latter can be extended to bounds on the coloured distance of balanced colourings.

5 An Algorithm for Trees

In this section we give a polynomial-time algorithm for finding the coloured distance of a tree for two colours.

Define a *semi-coloured* graph to be a graph in which some of the vertices are coloured. Then the coloured distance of the graph is the sum of the coloured distances between different-coloured vertices.

We will also need the following notation: for vertices u and v we define $\tilde{d}(u,v)$ to be the distance between u and v if u and v are both coloured but with different colours, and to be 0 otherwise. We will call this the coloured distance of the pair u and v. So the coloured distance of a semi-coloured graph is the sum of $\tilde{d}(u, v)$ over all unordered pairs of vertices. Furthermore, we define the *coloured distance of a subset* $S \subseteq V(G)$ to be the sum of the coloured distances of pairs of vertices in S, and define the coloured distance from set X to set Y to be the sum of the coloured distances of pairs of vertices composed of one member of X and one member of Y.

We will need the concept of a *swop*. If G is a semi-coloured graph, then $G \triangle \{u, v\}$ denotes the semi-coloured graph in which the colours of u and v have been swopped. That is, the colours of the other vertices remain the same, and the colour of u becomes the colour of v and vice versa.

We also need to recall that the coloured distance can be analysed by looking at the edges. If one determines how many times each edge occurs in the path between two coloured vertices, then the sum of usage is the coloured distance of that colouring.

5.1 Balanced Semi-Colourings in Trees

We consider trees with two colours blue and red. The set of blue vertices will be denoted by B and the red vertices by R. The next lemma shows that for balanced 2-semi-colourings, one can pair off the coloured vertices, each pair containing one red and one blue vertex, such that performing the swop of any pair does not increase the coloured distance.

Lemma 11 *Let T be a semi-coloured tree with equal number of blues and reds. Then there exists a bijection π from B to R such that for all $b \in B$*

$$\tilde{d}(T \triangle \{b, \pi(b)\}) \leq \tilde{d}(T).$$

PROOF.The proof is by induction on the number of non-end-vertices.

The base case is the set of all stars. Since the blue and red vertices are equinumerous, the contribution to the tree's coloured distance by a leaf-edge is purely determined by whether the end-vertex is coloured or not. Hence any swop of a red and a blue vertex in a semi-coloured star does not alter the distance of the graph. Thus any bijection will work.

Now the inductive step. If there is an end-vertex that is uncoloured, we may discard it as its associated leaf-edge is never used. Therefore, we may assume that every end-vertex is coloured.

Look at a penultimate vertex w in T (not an end-vertex, but all neighbours bar one are end-vertices). Let x be the neighbour of w that is not an end-vertex. (T is not a star.) Let W denote the set containing w and the end-vertices adjacent to w. Proceed as follows: arbitrarily pair off as much as possible vertices in W that have opposite colours. Call the set of paired vertices P. (For example, if W contains 7 reds and 4 blues then P will consist of 8 vertices and $W - P$ will consist of 3 red vertices.) Only one colour is present in $W - P$.

Let T' be the semi-coloured tree resulting from the deletion of P and all edges incident with w, and adding edges making each vertex of $W - P$ adjacent to x. (This reduces the number of non-end-vertices.)

By the inductive hypothesis, in T' there is a bijection π from $B' = B - P$ to $R' = R - P$ such that any swop of paired vertices does not increase the coloured distance. Extend π to a bijection from B to R of T by adding the pairings in P removed earlier.

We claim that this π has the desired properties. There are three types of pairings:

• *b or $\pi(b)$ is in $W - P$.* Clearly the swop of the colours for b and $\pi(b)$ affects the coloured distance between two vertices if and only if exactly one of them is in the set $\{b, \pi(b)\}$. The coloured distance from $\{b, \pi(b)\}$ to P is not affected by the swop in T, since P has an equal number of red and blue vertices.

For each end-vertex u of T with $u \in W - P$, identify the edge ux in T' with the edge uw in T. Now, any path R' between two vertices r and s in T' corresponds to an r–s path in T (with precisely the same edges), unless one of them, say r, is an end-vertex $u \in W - P$. For a corresponding u–s path in T, consider the path with edge set $E(R') - ux + uw + wx$.

Thus the difference in the coloured distance for T before and after the swop for $\{b, \pi(b)\}$ is the same as the difference in T' plus the difference in the number of times the edge wx is used. Recalling that $W - P$ is monochromatic and $|B| = |R|$, the latter difference is $2(|W - P| - 1) \leq 0$. So $\tilde{d}(T \triangle \{b, \pi(b)\}) \leq \tilde{d}(T)$.

• *both b and $\pi(b)$ are outside W.* Again the coloured distance from $\{b, \pi(b)\}$ to P is not affected by the swop. Also, the number of times an edge incident with w is used is not affected by this swop. Therefore if the swop does not increase the coloured distance in T', the swop does not increase it in T.

• *b and $\pi(b)$ are in P.* This is a swop either of the two ends of a leaf-edge, or of two end-vertices with the same neighbour. Such a swop cannot affect the coloured distance.

This completes the inductive step. QED

5.2 Two-Colourings of Trees

Lemma 12 *Let T and T^* be two trees with the same vertex set and edge set which are both totally coloured red and blue using red the same number of times. If $\tilde{d}(T^*) < \tilde{d}(T)$, then there exist vertices b and r, with b blue in T and red in T^* and with r red in T and blue in T^*, such that*

$$\tilde{d}(T \triangle \{b, r\}) < \tilde{d}(T).$$

PROOF.Let R' denote the set of vertices that are red in T and blue in T^*. Let B' denote the set of vertices that are blue in T and red in T^*. By the equinumerosity of B and R in T and T^*, it follows that $|R'| = |B'| = s$. Let $P = B' \cup R'$.

Consider any bijection π from B' to R'. Say π maps b_i onto r_i for $1 \leq i \leq s$. Define $T_i = T \triangle \{b_i, r_i\}$. Denote the coloured distance of P by C, C^* and C_i in trees T, T^* and T_i respectively. Note that $C = C^*$.

Consider

$$X = (s-1)\tilde{d}(T) + \tilde{d}(T^*) \qquad \text{and} \qquad Y = \sum_{i=1}^{s} \tilde{d}(T_i)$$

We first show that

$$Y - X = \left[\sum C_i\right] - [(s-1)C + C^*] = \left[\sum C_i\right] - sC. \tag{1}$$

If u and v are both not in P, then the swop of $\{b_i, r_i\}$ does not affect their colours and hence does not affect the coloured distance $\tilde{d}(u, v)$. Hence the contribution of $\tilde{d}(u, v)$ is the same to X as it is to Y.

Consider some vertex $u \notin P$, and some $i \in [s]$, and consider the contribution of $\tilde{d}(u, r_i) + \tilde{d}(u, b_i)$ to X and to Y. We claim that this is the same. Without loss of generality we may assume that u is red. Then the contribution to X is $(s-1)d(u, b_i) + d(u, r_i)$. The contribution to $\tilde{d}(T_j)$ is $d(u, b_i)$ if $i \neq j$, and $d(u, r_i)$ if $i = j$.

This means that the net contribution of all coloured distances where at least one of the vertices involved is outside P is the same to X as it is to Y, and equation (1) is established.

Aside: the above argument is not dependent on the fact that we are dealing with trees.

So, we must consider the contribution of $\tilde{d}(u,v)$ to X and to Y for $u,v \in P$.

Now, let T' be the graph resulting from taking T and uncolouring all the vertices not in P. The coloured distance of T' is C. By Lemma 11, there exists a bijection π from B' to R' such that

$$\tilde{d}(T' \triangle \{b_i, \pi(b_i)\}) \leq \tilde{d}(T')$$

for all $b_i \in B'$. The left-hand quantity is C_i and the right-hand quantity is C.

It follows that

$$Y = \sum_{i=1}^{s} \tilde{d}(T_i) \leq (s-1)\tilde{d}(T) + \tilde{d}(T^*) = X.$$

Since $\tilde{d}(T^*) < \tilde{d}(T)$, it follows that $X < s\tilde{d}(T)$. Thus there exists an i such that $\tilde{d}(T_i) < \tilde{d}(T)$. But that is what we wanted to show. QED

Thus there is a polynomial-time algorithm for finding an optimal 2-colouring in trees.

Algorithm 13 *To optimally (a,b)-colour a tree, start with any (a,b)-colouring; then repeatedly look for a swop that improves matters and perform the swop until no such swop exists.*

The number of swops is at most quadratic.

6 Diameter 2 and Complexity

In this section we observe that balanced bicolouring is hard. (A related problem is the optimal linear arrangement problem, which is also NP-hard. See [4].)

Define *maxcut*(G) as the maximum number of two-coloured edges in any 2-colouring of G with the supply of colours arbitrary.

Lemma 14 *Let G be a graph of order n and let $H = (G \times K_2) + K_2$. Then for $s = (n+1, n+1)$*

$$d_s(H) = 2(n+1)^2 - 2\,maxcut(G) - 3n - 1.$$

PROOF.H has diameter 2. Consider any colouring of H with equal reds and blues. Note that there are $(n+1)^2$ pairs of oppositely coloured vertices, each at a distance of one or two. For $d_s(H)$ we therefore consider $2(n+1)^2$ minus the number of two-coloured edges.

The edges of H can be partitioned into three classes, and we can upper bound the number of two-coloured edges in each class as follows:
(1) Inside the two copies of G: at most $2maxcut(G)$ two-coloured edges.
(2) Between the two copies of G: at most n two-coloured edges.
(3) Incident with the dominating K_2: at most $2(n+1)$ edges.

But it is impossible to have equality simultaneously in the upper bounds of (2) and (3). For, equality in (2) requires each colour used n times in the two copies of G, and so the two vertices in the dominating K_2 must receive different colours; equality in (3) requires that the two dominators receive the same colour. Hence the number of 2-coloured edges is at most:

$$2maxcut(G) + 3n + 1.$$

Equality is obtained by considering that partition (P, Q) of $V(G)$ which achieves $maxcut(G)$, and colouring red as follows: the vertices of P in the first copy of G, the vertices of Q in the second copy of G, and one dominator. QED

Since the decision problem for cut greater then k is NP-complete (see [4]), it follows that:

Theorem 15 *The decision problem for coloured distance less than k in a balanced colouring is NP-complete.*

References

[1] Buckley, F. & Harary, F., *Distance in Graphs*, Addison-Wesley, 1990.

[2] Dankelmann, P. & Entringer, R., Average distance, minimum degree and spanning trees, *J. Graph Theory*, **33** (2000), 1–13.

[3] Fajtlowicz, S., On conjectures of Graffiti, II, *Congressus Numer.*, **60** (1987), 189–197.

[4] Garey, M. & Johnson, D., *Computers and Intractability: A guide to the theory of NP-completeness,*, W.H. Freeman and Company, New York, 1983.

[5] Gerstel, O. & Zaks, S., A new characterization of tree medians with applications to distributed sorting, *Networks*, **24** (1994), 23–29.

[6] Hulme, B.L. & Slater, P., Minimean location of different facilities on a line network, *SIAM J. Algebraic Discrete Methods*, **2** (1981), 411–415.

[7] Kovider, M. & Winkler, P., Mean distance and minimum degree, *J. Graph Theory*, **25** (1997), 95–99.

[8] Meiers, D. & Slater, P., MAD partitioning for grid graphs, submitted.

[9] Slater, P., Minimum average distance partitions of the vertices on a line network, *Congressus Numer.*, **49** (1985), 215–221.

[10] Šoltés, L., Transmission in graphs: a bound and vertex removing, *Math. Slovaca*, **41** (1991), 11–16.