DIMACS Series in Discrete Mathematics
and Theoretical Computer Science

Consensus List Colorings of Graphs and Physical Mapping of DNA

N. V. R. Mahadev and Fred S. Roberts

Abstract. In graph coloring, one assigns a color to each vertex of a graph so that neighboring vertices get different colors. We shall talk about a bioconsensus problem relating to graph coloring and discuss the applicability of the ideas to the DNA physical mapping problem. In many applications of graph coloring, one gathers data about the acceptable colors at each vertex. A list coloring is a graph coloring so that the color assigned to each vertex belongs to the list of acceptable colors associated with that vertex. We consider the situation where a list coloring cannot be found. If the data contained in the lists associated with each vertex are made available to individuals associated with the vertices, it is possible that the individuals can modify their lists through trades or exchanges until the group of individuals reaches a set of lists for which a list coloring exists. We describe several models under which such a consensus set of lists might be attained. In the physical mapping application, the lists consist of the sets of possible copies of a target DNA molecule from which a given clone was obtained and trades or exchanges correspond to correcting errors in data.

1. Introduction

An old problem arising in decision theory, voting theory, and the social sciences is the *consensus problem*: How do we combine individual opinions into a decision by a group? This widely studied problem has a large associated literature. In recent years, methods developed in the social sciences have started to get wide use in the biological sciences. Use of consensus methods in biology is termed in this volume *bioconsensus*. In this paper, we introduce and develop some new models of consensus in a graph theory setting and discuss the application of these concepts in the context of the problem of physical mapping of DNA.

The new consensus models are concerned with graph coloring. Here, we seek to assign a color to each vertex of a graph so that vertices joined by an edge get

2000 *Mathematics Subject Classification.* Primary 05C90; Secondary 05C15, 91B14, 92B99.
Key words and phrases. list coloring, physical mapping, interval graph, fragment overlap graph.
The second author thanks the National Science Foundation for its support under grants NSF-SBR-9709134 and NSF-DBI-9982983 to Rutgers University and the Alfred P. Sloan Foundation for a grant to Rutgers University.

different colors. Graph coloring is an old subject with many important applications. While graph theory has widespread uses in molecular biology in particular and in other areas of biology in general, graph coloring is not widely used. However, it plays a role in the study of DNA physical mapping, as we shall observe. The interesting and important variant called list coloring, closely connected to applications, arises when the color assigned to a vertex in a graph must be chosen from a list of acceptable colors associated with that vertex. In this paper, we consider the situation when such a coloring cannot be found. We introduce and study various consensus-achieving models by which the individuals associated with vertices might change the lists associated with their vertices so that a coloring using acceptable colors might be achieved. These models are applied to the physical mapping problem in the context of handling certain kinds of errors in data.

We begin with some basic definitions and motivations. Suppose $G = (V, E)$ is a graph. A *coloring* is an assignment of a color (herein a positive integer) $f(x)$ to each vertex x in V so that

$$\{x, y\} \in E \rightarrow f(x) \neq f(y).$$

The smallest number of colors in a graph coloring of G is called the *chromatic number* and is denoted $\chi(G)$. Graph colorings arise in a variety of applications. For example, in channel assignment, V is a set of transmitters, an edge means interference, and the color is the channel assigned to the transmitter. In traffic phasing, V is a set of individuals or cars or groups that request use of a facility such as a computer, traffic intersection, or room; an edge means interference; and the color is the time the facility is assigned to the individual, car, or group. For a variety of other applications of graph coloring, see [**23**] [**24**].

In the next section, we review the study of physical mapping and describe the role of graph coloring in that study. We introduce list coloring in Section 3. The following three sections describe the three consensus models for list coloring and the final section gives concluding remarks and describes future directions for research.

2. How Graph Coloring Enters into Physical Mapping

This section will begin with an elementary introduction to physical mapping. It closely follows Setubal and Meidanis [**25**], which is a good source of additional details. A human chromosome is a DNA molecular consisting of about 10^8 base pairs. A *physical map* is defined to be a piece of DNA that tells us the location of certain markers along the molecule. The *markers* are typically small, but precisely defined, sequences. To build physical maps, we begin by making several copies of the DNA molecule to be mapped, the so-called *target DNA*. Then we break up each copy into *fragments* through the use of *restriction enzymes*, with different enzymes used for different copies. We obtain overlap information about the fragments and use this information to obtain the mapping.

How do we obtain the overlap information? In general, each fragment is still too long to be sequenced, so we cannot obtain overlap information as is done in the process called *fragment assembly*, i.e., by sequencing and comparing fragments. The method used instead is called *fingerprinting*. We shall describe the fingerprinting method known as hybridization. Another commonly used method is known as *restriction site analysis*.

In *hybridization*, we start with the set of fragments obtained from each copy of the DNA molecule. The fragments are then replicated using the method known as *cloning*. We end up with a collection of thousands of clones, many corresponding to the same fragment. (We shall be sloppy about the distinction between clones and fragments.) Now the idea is to associate with each clone (fragment) information that helps us uniquely describe it (as fingerprints do). This is accomplished by checking whether certain small sequences called *probes* bind (hybridize) to clones (fragments). The subset of probes that do bind is called the clone's *fingerprint*. By comparing fingerprints, we try to determine whether two clones overlap. The principle used is that two clones that share part of their fingerprints are likely to have come from overlapping regions of the target DNA. It should be noted that overlap information does not tell us the location of the corresponding fragment on the target DNA, only something about the relative order of the clones (fragments), information that was lost during the breakup.

It is important to emphasize that there are many errors in hybridization data and this observation is an important motivation for the application of list coloring consensus methods that we shall be investigating. Still following Setubal and Meidanis, we can identify various sources and kinds of errors. A probe might fail to bind where it should or it may bind where it shouldn't. Humans can mis-read or mis-record experimental results. During the cloning process, it is possible for two pieces of the target DNA to join and then be replicated as if they were a single clone. (Such clones are called *chimeric clones* and as many of 60% of the clones might be chimeric.) Probes can also bind along more than one site along the target DNA. (The method of *Sequence Tagged Site* or STS was designed to avoid this.) Finally, the data obtained is often incomplete. (Here, pooling methods can help.)

A key tool in physical mapping is the interval graph model. Let F be a family of sets. The *intersection graph* corresponding to F has the sets in F as its vertices and an edge between two sets if and only if they overlap. A graph is called an *interval graph* if it arises this way when the family F consists of intervals on the real line. Interval graphs have a long history and many applications in both the social and biological sciences, as well as in many other areas. For a discussion of interval graphs and their applications, see for example [**9**],[**11**],[**21**],[**22**]. There are good algorithms for recognizing an interval graph and for constructing a "map" of intervals on the line that have the corresponding intersection pattern.

From the overlap information, we create a *fragment overlap graph* as follows. The vertices are the fragments (clones) and the edges mean that the fragments overlap. If we had complete and error-free information about fragment overlapping, then the resulting graph would be an interval graph. If it is an interval graph, we can use the corresponding "map" of intervals to get information about the relative order of the fragments on the target DNA molecule and begin to obtain a physical map. However, the information we have is not complete and error-free and the fragment overlap graph might not be an interval graph. There are several approaches to handling this problem. We describe three approaches that start by labeling each clone (fragment) with identification of the copy of the target molecule from which it came. This label can be thought of as a color. If two clones come from the same copy of the target molecule, they cannot overlap, and so the labeling gives us a graph coloring of the fragment overlap graph. Later we shall discuss what to do if there are errors in the labeling and we don't have a graph coloring. The possibility that errors in the data lead to the fragment overlap graph not being an interval graph is given a lot of attention in the literature. The possibility that errors in the labeling might lead to our not having a graph coloring is not given this kind of attention. It will be our main emphasis in this paper. However, we first consider what to do if the fragment overlap graph is not an interval graph.

Let us consider first the case where the only type of error we have made is to omit overlaps – the *Case of False Negatives*. Then we might seek to add edges to the fragment overlap graph, enough to obtain an interval graph. We have to do this in such a way that the labeling we have remains a graph coloring. This may not be possible. If it is, we can use the resulting graph to obtain a candidate physical mapping. If there are several ways to add edges, we can look for the smallest number of added edges. It turns out that the problem of whether there is a way to add edges so as to obtain an interval graph for which the coloring remains a coloring is NP-hard. (See [**8**],[**10**],[**12**] for a discussion of the NP-hardness of various physical mapping problems and [**18**] for linear time algorithms for the simpler problem of transforming a given graph into a proper interval graph with the smallest number of edges. A *proper interval graph* is one where no interval properly contains another.)

A second case is the *Case of False Positives* where the only type of error we have made is to include overlaps when they are not actually present. Then we might seek to delete edges from the fragment overlap graph to obtain an interval graph with the labeling remaining a graph coloring. The latter requirement is trivially satisfied since removing edges always maintains the coloring property. It is natural to seek the smallest number of edges to remove. This is also an NP-hard problem.

If both kinds of errors might occur, we can think of having a set of edges E_1 on a vertex set V that are definitely there (overlaps that are definitely known to occur) and a set of edges that are definitely not there. The latter defines a set E_2 of edges on V that might be there, and we have $E_1 \subset E_2$. We have the same coloring on both graphs (V, E_1) and (V, E_2). We then ask if there is a set of edges E so that $E_1 \subset E \subset E_2$ and so that (V, E) is an interval graph and the coloring remains a coloring for (V, E). The latter requirement is automatically satisfied. The problem of determining if we can satisfy the former, known as the *interval sandwich problem*, is also NP-hard [**13**].

3. List Coloring

In many applications of graph coloring, there is an extra complication. There are some acceptable colors for each vertex and the color assigned to a vertex must be chosen from the set of acceptable colors. For instance, in channel assignment, we specify a set of acceptable channels and in traffic phasing a set of acceptable times. In physical mapping, we might lose or inaccurately record information about which copy of the target DNA molecule a given clone came from, and only know that the copy it came from belongs to one of a set of possible copies.

To formalize this idea, we let $S(x)$ denote a nonempty set of integers assigned to vertex x of graph G and call S a *list assignment* for G. We seek a graph coloring f so that for every $x \in V$, we have $f(x) \in S(x)$. Such a coloring is called a *list coloring* for (G, S). If a list coloring exists, we say that (G, S) is *list colorable*. List colorings were introduced independently by Vizing [**27**] and Erdös, Rubin, and Taylor [**6**] and there have been a very large number of papers about this subject in the past decade. The recent flurry of activity concerning list colorings is surveyed in the papers [**26**] and [**19**]. Earlier surveys are in [**1**] and [**16**].

Since it is NP-complete to determine if a graph is colorable in at most k colors if $k \geq 3$, it is easy to see that it is NP-complete to determine if there is a list coloring for (G, S) if $|\cup S(x)| \geq 3$. (It is even NP-complete for special cases such as bipartite graphs – see [**15**], [**20**]. See [**26**] for other special cases that are NP-complete.) If $|\cup S(x)| = 2$, then it is easy to see that it is polynomial to determine if (G, S) has a list coloring (see [**27**] and [**6**]).

In physical mapping, in the Case of False Negatives, we might now ask: Given (G, S), can we add edges to G to obtain an interval graph H so that (H, S) is list colorable? The answer is always "no" if (G, S) is not list colorable. Adding edges makes it harder to obtain a list coloring.

However, the comparable question in the Case of False Positives is interesting: Given (G, S), can we subtract edges from G to obtain an interval graph H so that (H, S) is list colorable? If so, what is the smallest number of edges we can remove to accomplish this?

The Sandwich question is also interesting. Given graphs $G_1 = (V, E_1)$ and $G_2 = (V, E_2)$, with $E_1 \subset E_2$, and given a list assignment S on V, is there an interval graph $H = (V, E)$ with $E_1 \subset E \subset E_2$ so that (H, S) is list colorable?

While these questions are all interesting, we defer answers to them. Instead, we shall ask what is perhaps a more fundamental question, one that doesn't involve changing edges. Namely, we shall simply think of making modifications in the assignment S to obtain an assignment S' with (G, S') list colorable.

To put our problem in a social choice or consensus context, we think of vertices as corresponding to individuals. It is this point of view we shall take in what follows. If (G, S) has no list coloring, some individuals represented by vertices may have to make sacrifices by exchanging or expanding their lists in order for a list coloring to exist. In physical mapping, the motivation is a bit different. It is based on the idea that there could be errors in the data about the possible copies of target DNA a clone came from. In the next three sections, we introduce and discuss three models for how individuals might change their lists. We think of these models as providing procedures for the group of individuals corresponding to the vertices to reach a consensus about a list coloring. In the physical mapping context, we think of these as providing procedures for modifying the sets of possible copies associated with a given clone with as few changes as possible.

In what follows, we use the notation K_n for the complete graph with n vertices and I_n for its complement, the graph with n vertices and no edges. If G and H are graphs with disjoint vertex sets, then $G + H$ denotes the graph whose vertex set is the union of the vertex sets of G and H and whose edge set is the union of the edge sets of G and H plus all edges joining a vertex in G to a vertex in H.

4. The Adding Model

In the first consensus model, we allow each individual to add one color from the set of colors already used, i.e., from $\cup S(x)$. For example, in the channel assignment application, a person may add one acceptable channel, and in the traffic phasing application, a person may add one acceptable time. In the physical mapping application, this is the case where a possible copy number might have been omitted from the list associated with a clone. Note that allowing the addition of more than one color does not add any flexibility since a person can always add only the color that he/she actually uses.

We say that (G, S) is *p-addable* if we can identify p distinct vertices $x_1, x_2, \ldots,$ x_p in G and (not necessarily distinct) colors $c_1, c_2, \ldots, c_p$ in $\cup S(x)$ so that if $S'(x_i) = S(x_i) \cup \{c_i\}, i = 1, 2, \ldots, p$, and $S'(u) = S(u)$ otherwise, then (G, S') is list-colorable. It is straightforward to observe that (G, S) is p-addable for some p iff

$$|\cup \{S(x) : x \in V\}| \geq \chi(G). \tag{4.1}$$

Necessity follows from the observation that if (G, S) is p-addable, then there is a graph coloring using colors from $\cup S(x)$. Sufficiency follows since Equation 4.1 implies that we can find a coloring f using colors in $\cup S(x)$. We then let c_i be $f(x_i)$ for all vertices x_i in G.

If $k \geq 3$, it is NP-complete to determine if $\chi(G) \leq k$, so it follows that if $|\cup S(x)| \geq 3$, it is NP-complete to determine if (G, S) is p-addable for some p. On the other hand, since there is a polynomial algorithm to determine whether or not $\chi(G) \leq 2$, there is a polynomial algorithm to determine whether or not (G, S) is p-addable for some p in the case where $|\cup S(x)| = 2$.

One way to measure how hard it is to reach consensus in problems such as channel assignment or traffic phasing or physical mapping is to determine the smallest number of "individuals" who have to add an acceptable choice, i.e., the smallest p such that (G, S) is p-addable. If such a p exists, we define $I(G, S)$ to be p and otherwise we say $I(G, S)$ is undefined. We call $I(G, S)$ the *inflexibility*. The higher it is, the more inflexible the group is. Note of course that $I(G, S) = 0$ iff (G, S) is 0-addable iff (G, S) is list colorable.

4.1. Complete Bipartite Graphs. To illustrate the notions of p-addability and inflexibility and give the combinatorial flavor of the arguments, we note that complete bipartite graphs have played an important role in the history of list colorings and so we give a few sample results about these graphs in this subsection. Let $K(n, m)$ denote the graph with two "partite" classes of vertices, n vertices in one class A and m in the other B, and all vertices in class A joined to all vertices in class B by edges.

Consider $K(10, 10)$. Define S as follows. On the 10 vertices of class A, use for the sets $S(x)$ the ten 2-element subsets of $\{1, 2, 3, 4, 5\}$, and similarly for the 10 vertices of class B. Suppose $f(x)$ is a list coloring obtained after additions change list assignment S into list assignment S'. Suppose f uses r colors on A and s colors on B. Then of course $r + s \leq 5$. Now $\binom{5-r}{2}$ sets on A don't use these r colors and so at least $\binom{5-r}{2}$ sets on A need colors added. Similarly, at least $\binom{5-s}{2}$ sets on B need colors added. This number of additions will work since all other sets on A have one of the r colors and similarly for B. It follows that

$$I(K(10, 10), S) \leq \binom{5-r}{2} + \binom{5-s}{2}$$

In fact, it is easy to show that one has equality here for $r = 3$ and $s = 2$, thus giving us $I(K(10, 10), S) = 4$.

A similar argument shows that $I(K(\binom{m}{2}, \binom{m}{2}, S)$ for an appropriate list assignment S is given by $\lfloor m/2 \rfloor \lfloor m/2 - 1 \rfloor$ if m is even and $\lfloor m/2 \rfloor^2$ if m is odd.

To give another sample result, consider $K(7,7)$ and the special case where all sets $S(x)$ have the same size, a common assumption in practice and theory. In particular, consider the case where all $|S(x)| = 3$. If $|\cup S(x)| = 6$, then we shall show that $(K(7,7), S)$ is 1-addable. Consider the seven 3-element sets $S(x)$ for x in partite set A. A simple combinatorial argument shows that there are integers i and j in $\{1,2,\ldots,6\}$ so that at most one of these seven sets does not contain either i or j. Add i to the set missing i and j if there is one, obtaining S'. Let $f(x)$ be i or j for x in A. Then for every y in partite set B, $S(y) = S'(y)$ has a third element different from i and j and we can define $f(y)$ to be that element. Hence, f is a list coloring for (G, S'). We conclude that $(K(7,7), S)$ is 1-addable.

On the other hand, we shall show that there is a list assignment S for $K(7,7)$ with each $|S(x)| = 3, |\cup S(x)| = 7$, and so that $(K(7,7), S)$ is not 0-addable. On partite set A, use the seven sets $\{i, i+1, i+3\}, i = 1,2,\ldots,7$, where addition is modulo 7. Use the same sets on partite set B. Suppose there is a list coloring f. We shall show that $F = \{f(x) : x \in A\}$ contains one of the sets $\{i, i+1, i+3\}$. Since this set is $S(y)$ for some y in partite set B, there is no way to pick a color $f(y)$ from $S(y)$. Suppose that F has consecutive i and $i+1$ for some i, including $i = 7$. Without loss of generality, assume that 1 and 2 are in F. If 4 is in F, then $\{1,2,4\}$ is in F. Suppose 4 is not in F. Since for some x in A, $S(x) = \{3,4,6\}$, we must have 3 or 6 in F. Similarly, using $\{4,5,7\}$, we see that 5 or 7 is in F. If 3 and 5 are in F, then $\{2,3,5\}$ is contained in F. Other cases are similar. The case where F has no consecutive i and $i+1$ is also similar.

4.2. Upper Bounds on $\mathbf{I(G, S)}$. Since we can add colors to at most each vertex, it is obvious that $I(G,S) \le |V(G)|$ provided that (G,S) is p-addable for some p. (In fact, it is not hard to prove the stronger result that if (G,S) is p-addable for some p, then $I(G,S) \le n - \omega(G)$, where $\omega(G)$ is the size of the largest clique of G.)

THEOREM 4.1. *There are (G,S) such that $I(G,S)/|V(G)|$ is arbitrarily close to 1.*

This theorem has the interpretation that there are situations where almost everyone has to "sacrifice" by taking as acceptable an alternative not on their initial list. In physical mapping, there are situations where essentially every list of copies needs to be expanded.

PROOF. Let $G = G_m = I_{\binom{m}{2}}$. Define S as follows. On $I_{\binom{m}{2}}$, use all pairs from $\{1,2,\ldots,m\}$ and on K_{m-1} use the pairs $\{i, i+1\}, i = 1,2,\ldots,m-1$. We shall show that $I(G,S) = \binom{m-1}{2}$. To show $\ge$, we note that no matter what additions are made to the sets $S(x_i)$ for x_i in K_{m-1}, we still need $m-1$ colors to color K_{m-1}. Then the $\binom{m-1}{2}$ vertices of $I_{\binom{m}{2}}$ whose sets consist of two of these colors cannot be colored from their sets. The inequality $\le$ follows if we take $f(x)$ to be i if $S(x) = \{i, i+1\}$ for x in K_{m-1}, $f(x) = m$ if $m \in S(x)$ for x in $I_{\binom{m}{2}}$, and otherwise add m to set $S(x)$ for x in $I_{\binom{m}{2}}$, and take $f(x) = m$. Since $|V(G)| = \binom{m}{2} + m - 1$, it is easy to show that the theorem holds. □

Note that the result still holds if we add the common requirement that all sets $S(x)$ have the same cardinality k. Take $m > k$ and use $I_{\binom{m}{k}}$ in place of $I_{\binom{m}{2}}$

On $I_{\binom{m}{k}}$, use all k-element subsets from $\{1, 2, \ldots, m\}$ and on K_{m-1}, use the sets $\{i, i+1, \ldots, i+k-1\}, i = 1, 2, \ldots, m-1$, where addition is modulo m. Then, $I(G, S) = \binom{m-1}{k}$.

5. The Trading Model

A second consensus model considers the possibility of side agreements among individuals to lead to increased flexibility. In particular, it allows the trade (or purchase) of colors from one individual's acceptable set to another's. In the first model, we paid no attention to where the added colors in a color set come from. In the new model, we think of trades as taking place in a series of steps. A *trade* from x to y means that we remove color c from $S(x)$ and add it to $S(y)$, i.e., we define a new list assignment S' by taking $S'(x) = S(x) - \{c\}, S'(y) = S(y) \cup \{c\}$, and otherwise $S'(u) = S(u)$. In this new model, in the physical mapping application, we think of the possibility that a label was incorrectly recorded in a set of possible labels of another clone and should be moved.

We shall ask how many trades are required in order to obtain a list assignment that has a list coloring. We say that (G, S) is *p-tradeable* if this can be done in p trades. For instance, if we take the path P_k consisting of k vertices and define S by putting the set $\{1\}$ on all but the last vertex and the set $\{2\}$ on the last vertex, then if $k \geq 4$, (P_k, S) is not p-tradeable for any p. There are not enough 2's available. This also shows that Equation 4.1 is not sufficient for p-tradeability for some p. The following problem, denoted as problem $(G, c_1, c_2, \ldots, c_r)$, is clearly related to p-tradeability: Given graph G and positive integers $c_1, c_2, \ldots, c_r$, is there a graph coloring of G so that for $i = 1, 2, \ldots, r$, the number of vertices receiving color i is at most c_i? If c_i is the number of times that i occurs in some $S(x)$, then (G, S) is p-tradeable for some p if and only if the problem $(G, c_1, c_2, \ldots, c_r)$ has a positive answer. The problem $(G, c_1, c_2, \ldots, c_r)$ arises in "timetabling" applications in scheduling. In [**2**] and [**4**], it is shown that the problem $(G, c_1, c_2, \ldots, c_r)$ is NP-complete, even for the class of line graphs of bipartite graphs. Hansen, Hertz, and Kuplinsky [**14**] study this problem for the special case where all c_i are the same. An edge coloring version of the same problem was shown to be NP-complete even for bipartite graphs with maximum degree 3 in [**7**]. The papers [**2**], [**3**], [**4**] consider the problem $(G, S, c_1, c_2, \ldots, c_r)$ where S is a list assignment for the graph G and we ask if there is a list coloring for (G, S) solving problem $(G, c_1, c_2, \ldots, c_r)$. It is shown that the problem $(G, S, c_1, c_2, \ldots, c_r)$ can be solved in polynomial time when G is a union of vertex-disjoint cliques. Dror, et al. [**5**] study the problem $(G, S, c_1, c_2, \ldots, c_r)^*$ where $\sum c_i$ is the number of vertices of G and each color i must be used *exactly* c_i times. The former is a necessary condition for the problem to have a positive answer. Dror, et al. show that problem $(G, S, c_1, c_2, \ldots, c_r)^*$ is NP-complete even if the graph is a path or a union of paths with each $|S(v)| \leq 2$. This shows that the problem $(G, S, c_1, c_2, \ldots, c_r)$ is also NP-complete in this case, since if $\sum c_i$ is the number of vertices, then problem $(G, S, c_1, c_2, \ldots, c_r)^*$ has a positive answer if and only if problem $(G, S, c_1, c_2, \ldots, c_r)$ does. Dror, et al. also show that if the total number of colors in all of the lists is fixed (not part of the input), the problem $(G, S, c_1, c_2, \ldots, c_r)^*$ is solvable in polynomial time on paths and vertex-disjoint unions of paths. The special case where $| \cup S(x)| = 3$ had previously been solved in [**28**].

If $|\cup S(x)| = s \geq 3$, it is clearly NP-complete to determine if (G, S) is p-tradeable for some p. For, given G and s, define $S(x) = \{1, 2, \ldots, s\}$ for all x in G. Then $\chi(G) \leq s$ iff (G, S) is p-tradeable for some p iff (G, S) is list-colorable. However, we have the following observation.

PROPOSITION 5.1. *If $|\cup S(x)| = 2$, then it is polynomial to determine whether or not (G, S) is p-tradeable for some p.*

PROOF. It suffices to consider the case where G is connected. Note that if (G, S) is p-tradeable for some p, then G is 2-colorable, i.e., bipartite. By connectedness, if G is bipartite, there is a unique bipartition into classes A and B. Then it is easy to see that the necessary and sufficient condition for p-tradeability for some p is that G is bipartite with classes A and B and there are at least $|A|$ 1's and at least $|B|$ 2's in $\cup S(x)$ or vice versa. $\square$

Analogous to the definition of the inflexibility $I(G, S)$, we define the *trade inflexibility* $I_t(G, S)$ to be the smallest p so that (G, S) is p-tradeable, with $I_t(G, S)$ undefined if there is no such p. We note that if $I_t(G, S)$ is defined, then $I_t(G, S) \leq |V(G)|$. For, suppose S' is obtained from S by a sequence of p trades and (G, S') has a list coloring f. Then each $f(x)$ is in some $S(y)$ and the number of x's for which $f(x) = i$ is at most the number of times i is in some $S(y)$. So, we can arrange to trade the required number of i's to sets assigned to vertices x for which $f(x) = i$. There is at most one incoming trade into each set this way.

The following theorem has the same interpretation as the analogous theorem for $I(G, S)$.

THEOREM 5.2. *There are (G, S) such that $I_t(G, S)/|V(G)|$ is arbitrarily close to 1.*

PROOF. Suppose that $m > 3p+2$. Consider the graph $G = G(m, p) = K_{m-p} + pI_{m-p}$, where pI_{m-p} is $I_{m-p} + I_{m-p} + \ldots + I_{m-p}$, with p copies of I_{m-p}. Define S as follows. On K_{m-p}, use the sets

$$\{i, i+1, m-p+1, m-p+2, \ldots, m\}$$

for $i = 1, 2, \ldots, m-p-1$ and

$$\{m-p, 1, m-p+1, m-p+2, \ldots, m\}.$$

On each copy of I_{m-p}, use the sets $\{i, i+1\}, i = 1, 2, \ldots, m-p-1$, and $\{m-p, 1\}$.

Let f be a list coloring obtained after trades give a new list assignment S'. Each $i, 1 \leq i \leq m-p$, appears in two sets $S(x)$ on K_{m-p} and two sets $S(x)$ on each copy of I_{m-p}, so in $2(p+1)$ sets in all. Since $m > 3p+2$, and therefore $m-qp > 2(p+1)$, any list coloring f uses a color different from $1, 2, \ldots, m-p$ on each copy of I_{m-p}. Thus, on each such copy, f uses a color in the set $\{m-p+1, m-p+2, \ldots, m\}$. It follows that none of these colors are used by f on K_{m-p}. Thus, f on K_{m-p} must use the colors $1, 2, \ldots, m-p$ and none of these colors can be used by f on any copy of I_{m-p}.

We are left with p colors to be used by f on the p copies of I_{m-p}. It follows that f uses color $m-p+1$ on all vertices of one copy of I_{m-p}, color $m-p+2$ on all vertices of a second copy of $I_{m-p}, \ldots$, and color m on all vertices of a p^{th} copy of I_{m-p}. To obtain S', we must trade $m-p$ copies of each of $m-p+1, m-p+2, \ldots, m$

to sets on copies of I_{m-p}. Hence, we need a minimum of $p(m-p)$ trades. This number suffices. Thus,

$$I_t(G,S)/|V(G)| = p(m-p)/(p+1)(m-p) \to 1$$

as $p \to \infty$. □

There might be some situations in which we only wish to allow trades to go from a vertex to an adjacent (neighboring) vertex. This might occur in channel assignment, for example, if adjacency (interference) corresponds to physical proximity. Adjacency had been interpreted as interference, which in turn could correspond to physical closeness. In this case, we might only be willing or able to trade with a nearby person. It is not clear what this means in the physical mapping application. We shall say that (G,S) is *p-neighbor-tradeable* if there is a sequence of p trades, each from a vertex to an adjacent one, that results in a list colorable list assignment. Let $I_{t,n}(G,S)$ be the smallest p so that (G,S) is p-neighbor-tradeable and be undefined if there is no such p. In contrast to the situation with p-tradeability, it is possible for $I_{t,n}(G,S)$ to be larger than $|V(G)|$. In fact, $I_{t,n}(G,S)/|V(G)|$ can be arbitrarily large. This is easy to see and we shall indicate why below.

6. The Exchange Model

A variant on the trade model arises when we do not use one-way trades, but instead use two-way exchanges. Here, a color from $S(x)$ and a color from $S(y)$ are interchanged at each step. In physical mapping, this might correspond to the case where labels of two clones are transposed. We think of exchanges as taking place in a series of steps. Formally, an *exchange* between x and y means that we find a color c in $S(x)$ and a color d in $S(y)$ and move c to $S(y)$ and d to $S(x)$ by taking $S'(x) = S(x) \cup \{d\} - \{c\}$, $S'(y) = S(y) \cup \{c\} - \{d\}$, and otherwise $S'(u) = S(u)$.

We shall ask how many exchanges are required in order to obtain a list assignment that has a list coloring. We say that (G,S) is *p-exchangeable* if this can be done in p exchanges. Note that (G,S) is p-exchangeable for some p iff it is q-tradeable for some q. For, if (G,S) is p-exchangeable, then since every exchange can be broken into two trades, it is $2p$-tradeable. If (G,S) is p-tradeable for some p, then, as we have observed before, we can accomplish this with p trades so that every vertex receives a trade at most once. Every vertex that receives a trade can therefore trade back one of its colors without affecting the result. So, this is p-exchangeable. Thus, the complexity of determining whether or not (G,S) is p-exchangeable for some p is the same as that of determining whether or not (G,S) is p-tradeable for some p.

Analogous to the definition of trade inflexibility $I_t(G,S)$, we define the *exchange inflexibility* $I_e(G,S)$ to be the smallest p so that (G,S) is p-exchangeable, with $I_e(G,S)$ undefined if there is no such p. By the above reasoning, we note that $I_e(G,S) \le I_t(G,S)$ and so if $I_e(G,S)$ is defined, $I_e(G,S) \le |V(G)|$.

In the example in the proof of Theorem 5.2, we have $I_e(G,S) = I_t(G,S)$. The proof applies with only minor modification at the end. Thus, we have the following theorem.

THEOREM 6.1. *There are (G,S) such that $I_e(G,S)/|V(G)|$ is arbitrarily close to 1.*

This has the same interpretation as the analogous theorem for $I(G,S)$.

Just as with the trade model, there might be situations in the exchange model in which we only allow exchanges to go between adjacent vertices. We define *p-neighbor-exchangeable* and $I_{e,n}(G,S)$ analogously to the same concepts for tradeable.

One can show that $I_{e,n}(G,S)/|V(G)|$ can be arbitrarily large. For instance, consider the path P_{2k+1} on $2k+1$ vertices with sets $S(x) = \{1\}$ on the first $k+1$ vertices and sets $S(x) = \{2\}$ on the last k vertices. The only way to color this path with colors from the set $\cup S(x) = \{1,2\}$ is to alternate colors. Thus, we must move 2's to the left and 1's to the right. Doing this by a series of exchanges between neighbors is analogous to changing the identity permutation into another permutation by transpositions of the form $(i \ \ i+1)$. The number of transpositions needed to do this is readily established. According to a well-known formula (see for example [**17**]), the number of such transpositions required to switch an identity permutation into the permutation π of $\{1,2,\ldots,n\}$ is given by

$$J(\pi) = |\{(i,j) : 1 \leq i < j \leq n \ \ \& \ \ \pi(i) > \pi(j)\}.$$

Here, π is the permutation

$$1 \ \ k+2 \ \ 2 \ \ k+3 \ \ 3 \ \ k+4 \ \ldots \ k \ \ 2k+1 \ \ k+1$$

and $J(\pi) = k(k+1)/2$. Hence,

$$I_{e,n}(P_{2k+1},S)/|V(P_{2k+1})| = k(k+1)/2(2k+1) \to \infty$$

as $k \to \infty$. An analogous proof shows that for this example,

$$I_{t,n}(P_{2k+1},S)/|V(P_{2k+1})| \to \infty$$

as $k \to \infty$.

7. Concluding Remarks

We have presented three consensus procedures individuals might use to modify their acceptable sets in order for the group to achieve a list colorable situation. So far, very little is known about these procedures. For instance, it would be very helpful to describe conditions under which (G,S) is p-tradeable or p-exchangeable for some p, at least for special classes of G and S. Very little is known about the values of or bounds for the parameters $I(G,S)$, $I_t(G,S)$, $I_{t,n}(G,S)$, $I_e(G,S)$, and $I_{e,n}(G,S)$ for specific graphs or classes of graphs and specific list assignments or classes of list assignments. Also, we might ask about values of or bounds for these parameters under the extra restriction that all sets $S(x)$ have the same fixed cardinality. Finally, we might ask for good algorithms for finding optimal ways to modify list assignments so that we obtain a list colorable assignment under the different consensus models, again at least for special classes of graphs and classes of list assignments.

Specifically motivated by the physical mapping problem, we have posed the following questions. Given (G,S), can we subtract edges from G to obtain an interval graph H so that (H,S) is list colorable? If so, what is the smallest number of edges we can remove to accomplish this?

Also, given graphs $G_1 = (V,E_1)$ and $G_2 = (V,E_2)$, with $E_1 \subset E_2$, and given a list assignment S on V, is there an interval graph $H = (V,E)$ with $E_1 \subset E \subset E_2$ so that (H,S) is list colorable?

Of course, we also want to ask whether the consensus models are useful in physical mapping, and to this point these ideas are preliminary and have not been applied in practice.

Acknowledgement

The authors thank Paul Dreyer for his helpful comments.

References

[1] N. Alon, Restricted colorings of graphs, in (K. Walker, ed.), *Surveys in Combinatorics*, *Proc. 14^{th} British Combinatorial Conference*, London Math. Soc. Lecture Notes Series **187**, Cambridge University Press, Cambridge, 1993, 1-33.

[2] D. de Werra, Restricted coloring models for timetabling, *Discrete Math.* **165** (1997), 161-170.

[3] D. de Werra, On a multiconstrained model for chromatic scheduling, *Discrete Applied Math.* **94** (1999), 171-180.

[4] D. de Werra, Restricted graph coloring: Some mathematical programming models, *CRM Proceedings and Lecture Notes*, Centre de Recherches Mathematiques, Montreal **25** (1999), 135-148.

[5] M. Dror, G. Finke, S. Gravier, and W. Kubiak, On the complexity of a restricted list-coloring problem, *Discrete Math.* **195** (1999), 103-109.

[6] P. Erdös, A. L. Rubin, and N. Taylor, Choosability in graphs, *Congressus Numerantium* **XXVI** (1979), 125-157.

[7] S. Even, A. Itai, and A. Shamir, On the complexity of timetable and multicommodity flow problems, *SIAM J. Computing* **5** (1976), 691-703.

[8] M. R. Fellows, M. T. Hallett, and H. T. Wareham, DNA physical mapping: Three ways difficult, in *Proceedings of the First Annual European Symposium on Algorithms*, Lecture Notes in Computer Science **726**, Springer-Verlag, Berlin, 1993.

[9] P. C. Fishburn, *Interval Orders and Interval Graphs*, Wiley, New York, 1985.

[10] P. W. Goldberg, M. C. Golumbic, H. Kaplan, and R. Shamir, *Three Strikes Against Physical Mapping of DNA*, unpublished manuscript, 1993.

[11] M. C. Golumbic, *Algorithmic Graph Theory and Perfect Graphs*, Academic Press, New York, 1980.

[12] M. C. Golumbic, H. Kaplan, and R. Shamir, On the complexity of physical mapping, *Advances in Appl. Math.* **15** (1994), 251-261.

[13] M. C. Golumbic and R. Shamir, R., Complexity and algorithms for reasoning about time: A graph-theoretic approach, *J. ACM* **40** (1993), 1108-1133.

[14] P. Hansen, A. Hertz, and J. Kuplinksy, Bounded vertex colorings of graphs, *Discrete Math.* **111** (1993), 305-312.

[15] K. Jansen, and P. Scheffler, Generalized colorings for tree-like graphs, *Discrete Applied Mathematics* **75** (1997), 135-155.

[16] T. R. Jensen and B. Toft, *Graph Coloring Problems*, Wiley Interscience, New York, 1995.

[17] M. R. Jerrum, The complexity of finding minimum-length generator sequences, *Theoretical Computer Science* **36** (1985), 265-289.

[18] H. Kaplan, R. Shamir, and R. E. Tarjan, Tractability of parametrized completion problems on chordal and interval graphs: Minimum fill-in and physical mapping, in *Proceedings of the IEEE Thirty-fifth Annual Symposium on Foundations of Computer Science*, 1994, 780-791.

[19] J. Kratochvíl, Z. Tuza, and M. Voigt, New trends in the theory of graph colorings: Choosability and list coloring, in (R. L. Graham, J. Kratochvíl, J. Nešetřil, and F. S. Roberts eds.), *Contemporary Trends in Discrete Mathematics*, DIMACS Series **49**, American Mathematical Society, Providence, RI, 1999, 183-197.

[20] M. Kubale, Some results concerning the complexity of restricted colorings in graphs, *Discrete Applied Math.* **36** (1992), 35-46.

[21] F. S. Roberts, *Discrete Mathematical Models, with Applications to Social, Biological, and Environmental Problems*, Prentice-Hall, Englewood Cliffs, NJ, 1976.

[22] F. S. Roberts, *Graph Theory and its Applications to Problems of Society*, CBMS-NSF Monograph No. 29, SIAM, Philadelphia, 1978.

[23] F. S. Roberts, *Applied Combinatorics*, Prentice-Hall, Englewood Cliffs, NJ, 1984.

[24] F. S. Roberts, From garbage to rainbows: Generalizations of graph coloring and their applications, in (Y. Alavi, G. Chartrand, O.R. Oellermann, and A.J. Schwenk, eds.), *Graph Theory, Combinatorics, and Applications* **2**, Wiley, New York, 1991, 1031-1052.

[25] J. Setubal and J. Meidanis, *Introduction to Computational Molecular Biology*, Brooks/Cole, Pacific Grove, CA, 1997.

[26] Z. Tuza, Graph colorings with local constraints – a survey, *Discussiones Mathematicae – Graph Theory* **17** (1997), 161-228.

[27] V. G. Vizing, Coloring the vertices of a graph in prescribed colors, *Metody Diskret. Anal. v Teorii Kodov i Schem* **29** (1976), 3-10. (In Russian)

[28] K. Xu, *A Special Case of the Edge-chromatic Scheduling Problem*, Report ORWP 96/03, Ecole Polytechnique Federale de Lausanne, 1996.

Department of Computer Science, Fitchburg State College
E-mail address: nmahadev@fsc.edu

Department of Mathematics and DIMACS, Rutgers University
E-mail address: froberts@dimacs.rutgers.edu