

Vertex coloring acyclic digraphs and their corresponding hypergraphs*

Geir Agnarsson [†] Ágúst S. Egilsson [‡] Magnús M. Halldórsson [§]

Abstract

We consider vertex coloring of an acyclic digraph \vec{G} in such a way that two vertices which have a common ancestor in \vec{G} receive distinct colors. Such colorings arise in a natural way when bounding space for various genetic data for efficient analysis. We discuss the corresponding *down-chromatic number* and derive an upper bound as a function of $D(\vec{G})$, the maximum number of descendants of a given vertex, and the degeneracy of the corresponding hypergraph. Finally we determine an asymptotically tight upper bound of the down-chromatic number in terms of the number of vertices of \vec{G} and $D(\vec{G})$.

2000 MSC: 05C15, 05C20, 05C65, 05C85, 05D90, 06A06.

Keywords: Genetic databases, digraph, ancestor, down-set, vertex coloring, hypergraph, block design.

1 Introduction

The purpose of this article is to discuss a special kind of vertex coloring for acyclic digraphs, where vertices with a common ancestor must receive distinct colors. We discuss some properties of such colorings, similarity and differences with strong hypergraph colorings, and derive an upper bound which, in addition, yields an efficient coloring procedure.

Digraphs representing various biological phenomena and knowledge are ubiquitous in the life sciences and in drug discovery research, e.g. the gene ontology digraph maintained by the Gene Ontology Consortium [7]. An overview of several projects relating to indexing of semistructured data (i.e. acyclic digraphs) can be found in [1]. In these biological digraphs it is important to be able to access the ancestors of nodes in a fast and efficient manner.

Consider the problem of finding a representation an acyclic digraph in a database to allow for fast access to the set of ancestors of a given node. The ancestors of a node are its in-neighbors in the transitive closure of the digraph. If the digraph is sparse and shallow,

*Some of the results presented here appeared in a preliminary form in [3].

[†]Department of Mathematical Sciences, George Mason University, MS 3F2, 4400 University Drive, Fairfax, VA 22030, geir@math.gmu.edu.

[‡]Science Institute, University of Iceland IS-107 Reykjavik, Iceland, egilsson@hi.is.

[§]Department of Computer Science, University of Iceland, IS-107 Reykjavik, Iceland, mmh@hi.is.

the transitive closure is also sparse. Thus, an adjacency matrix representation would be neither efficient nor fast. On the other hand, a matrix has the advantage of corresponding nicely to the relational representation of modern databases. In a database relation that corresponds to an adjacency matrix, the non-empty elements in each column correspond to the in-neighbors of the node indexing that column. This vertex set can then be combined by joins with other tables that are also indexed by vertices, giving an effective language of querying based on graphical properties. Thus, it would be preferable to find a representation that both has a matrix structure, yet consists of relatively small rows.

One compact matrix representation would be to store the adjacency lists in compacted array form, where a list with k elements is stored in the first k array elements. In this case, however, there is no easy way of accessing all the edges entering a given vertex. While this could be alleviated by storing the inverted adjacency matrix, note that in the context of database access, rows are conceptually different from columns. Instead, we seek a compacted representation where all in-edges of a given node are stored in the same column. We say that a many-to-one mapping of vertices to columns that preserves adjacency lists, has the *AC-property*. By recording the mapping of nodes to their respective column storing their in-neighbors, one obtains the same desirable properties of adjacency matrices in the context of a relational database. If the graph is sparse, the possibilities of storage reduction are significant. The actual improvement is related to the number of colors needed in a certain coloring of the digraph, which we now briefly discuss.

A proper *down-coloring* of a digraph is a vertex coloring where vertices with a common ancestor receive different colors. The *down-chromatic number* of a digraph is the minimum number of colors in a down-coloring of the digraph. In a compacted matrix representation of the transitive closure of a digraph, we assign multiple vertices to the same column, but in such a way that their in-adjacency lists must be disjoint. Two vertices have disjoint sets of in-neighbors in the transitive closure if, and only if, they have no common ancestor. Therefore, a down-coloring of a digraph corresponds to a valid compacted representation of its transitive closure, and the down-chromatic number is the minimum number of columns needed in such a representation.

EXAMPLE: Consider the digraph \vec{G} , on $n = 6$ vertices representing genes, where a directed edge from one vertex to a second one indicates that the first gene is an ancestor of the second gene.

$$\begin{aligned} V(\vec{G}) &= \{g_1, g_2, g_3, g_4, g_5, g_6\}, \\ E(\vec{G}) &= \{(g_1, g_4), (g_1, g_5), (g_2, g_4), (g_2, g_6), (g_3, g_5), (g_3, g_6)\}. \end{aligned}$$

In the adjacency matrix representation of this 6 node digraph, we assign a column to each vertex g_i . As we see in the left diagram of Table 1, most of the entries of this 6×6 matrix are empty. By reducing the number of columns in such a way that the AC-property still holds, we obtain a smaller and more compact 6×3 matrix representation as seen on the right diagram of Table 1. There, the i -th row still contains the descendants of g_i and the ancestors of g_i are those g_j 's whose rows g_i appears in. Note that, (i) each g_i appears in

	g_1	g_2	g_3	g_4	g_5	g_6	→		1	2	3
g_1	1	0	0	1	1	0		g_1	g_1	g_4	g_5
g_2	0	1	0	1	0	1		g_2	g_6	g_4	g_2
g_3	0	0	1	0	1	1		g_3	g_6	g_3	g_5
g_4	0	0	0	1	0	0		g_4	–	g_4	–
g_5	0	0	0	0	1	0		g_5	–	–	g_5
g_6	0	0	0	0	0	1		g_6	g_6	–	–

Table 1: An example for $n = 6$.

exactly one column and, (ii) two genes appear in the same column only if their sets of ancestors are disjoint. Here we can view the column numbers 1, 2, and 3 as distinct colors assigned to each vertex. Note further that in this example the transitive closure of \vec{G} is simply \vec{G} itself. An explicit example of how such a coloring can speed up queries in the gene ontology digraph can be found in the appendix of [6].

Hence, it seems worthwhile to discuss two aspects of such colorings: (1) How can we assign reasonably few colors to the vertices/columns efficiently, and (2) how large can the discrepancy theoretically be between the actual minimum number of colors needed and an obvious lower bound of needed colors?

Our Results

The contributions of this paper are threefold. First, we establish a close link between down coloring digraphs and strong coloring hypergraphs. Second, we give efficiently computable bounds on the down chromatic number in terms of the inductiveness of the related hypergraph and $D(\vec{G})$, the maximum number of descendants of a given vertex. And thirdly, we give a tight bound on the discrepancy between the down chromatic number and the lower bound $D(\vec{G})$. This also has independent interest as characterizing the largest ratio of the strong chromatic number of a hypergraph to the sum of the number of edges and the number of vertices.

Related Work

Some special classes of such acyclic digraphs are studied in [2], in particular those of height two in which every vertex has an in-degree of two. For a brief introduction and additional references to the ones mentioned here, we refer to [2].

Note that acyclic digraphs are often called directed acyclic graphs or DAG's by computer scientists, as is the case in [13, p. 194].

A straight forward condition of a vertex coloring of a digraph \vec{G} is to insist that two vertices u and v receive distinct color if there is a directed edge from u to v in \vec{G} . Such a coloring is, of course, the same as coloring the vertices of the underlying graph G of \vec{G} (by forgetting the orientation of the directed edges) in the usual sense.

Another vertex coloring of digraphs that relies on the direction of the edges is the *dichromatic number* of a digraph \vec{G} , as studied in [11] and [14], which is defined as the minimum number of colors needed to vertex color \vec{G} in such a way that no monochromatic directed cycle is created.

Strong colorings of hypergraphs have been studied, but not quite to the extent of various other types of colorings of hypergraphs. Since strong colorings are generalizations of the usual vertex colorings of graphs, the determination of the exact strong chromatic number of a hypergraph is in general a daunting task. Most results in this direction in the literature on strong colorings are restricted to some very special types of hypergraphs. In [16] a nice survey of various aspects of hypergraph coloring theory is found, containing almost all fundamental results in the past three decades. What we are concerned here in this section though, is not necessarily an *exact* computation of the strong chromatic number, but rather a good theoretical upper bound that is valid for all possible corresponding digraphs \vec{G} . In Section 4 however, we discuss the asymptotics of how large the exact down-chromatic number can be.

2 Basic definitions

We attempt to be consistent with standard graph theory notation in [17], and the notation in [15] when applicable.

For a natural number $n \in \mathbb{N}$ we let $[n] = \{1, \dots, n\}$. A *simple digraph* is a finite simple directed graph $\vec{G} = (V, E)$, where $V = V(\vec{G})$ is a finite set of vertices and $E = E(\vec{G}) \subseteq V \times V$ is a set of directed edges. The digraph \vec{G} is said to be *acyclic* if \vec{G} has no directed cycles. Henceforth \vec{G} will denote an acyclic digraph in this section.

The binary relation \leq on $V(\vec{G})$ defined by

$$u \leq v \Leftrightarrow u = v, \text{ or there is a directed path from } v \text{ to } u \text{ in } \vec{G}, \quad (1)$$

is reflexive, antisymmetric and transitive and therefore a partial order on $V(\vec{G})$. Hence, whenever we talk about \vec{G} as a *poset*, the partial order will be the one defined by (1).

The *transitive closure* of \vec{G} is the poset \vec{G} viewed as a digraph, that is the digraph \vec{G}^* on $V(\vec{G})$ where $(v, u) \in E(\vec{G}^*)$ iff $u < v$. By the *height* of \vec{G} as a poset, we mean the number of vertices in the longest directed path in \vec{G} . We denote by $\max\{\vec{G}\}$ the set of maximal vertices of \vec{G} with respect to the partial order \leq .

For vertices $u, v \in V(\vec{G})$ with $u \leq v$, we say that u is a *descendant* of v , and v is an *ancestor* of u . The *closed principal down-set* or simply the *down-set* $D[u]$ of a vertex $u \in V(\vec{G})$ is the set of descendants of u in \vec{G} , that is, $D[u] = \{x \in V(\vec{G}) : x \leq u\}$. Likewise, the *open principal down-set* or the *open down-set* of a vertex u is $D(u) = D[u] \setminus \{u\}$.

Definition 2.1 A down-coloring of \vec{G} is a map $c : V(\vec{G}) \rightarrow [k]$ satisfying

$$u, v \in D[w] \text{ for some } w \in V(\vec{G}) \Rightarrow c(u) \neq c(v)$$

for every $u, v \in V(\vec{G})$. The down-chromatic number of \vec{G} , denoted by $\chi_d(\vec{G})$, is the least k for which \vec{G} has a proper down-coloring $c : V(\vec{G}) \rightarrow [k]$.

Clearly, in an undirected graph G the vertices in a clique must all receive distinct colors in a proper vertex coloring of G . Therefore $\omega(G) \leq \chi(G) \leq |V(G)|$ where $\omega(G)$ denotes the clique number of G . Similarly, if $D(\vec{G}) = \max_{u \in V(\vec{G})} \{|D[u]|\}$ for our acyclic digraph \vec{G} , we clearly have $D(\vec{G}) \leq \chi_d(\vec{G}) \leq |V(\vec{G})|$. Hence, when considering down-colorings, it can be useful to map the problem to one on undirected graphs. Given an acyclic digraph \vec{G} , the corresponding simple undirected *down-graph* G' has the same set of vertices, with each pair of vertices connected that are contained in the same principal down-set:

$$\begin{aligned} V(G') &= V(\vec{G}), \\ E(G') &= \{\{u, v\} : u, v \in D[w] \text{ for some } w \in V(\vec{G})\}. \end{aligned}$$

In this way we have transformed the problem of down-coloring the digraph \vec{G} to the problem of vertex coloring the simple undirected graph G' in the usual sense, and we have $\chi_d(\vec{G}) = \chi(G')$. Hence, from the point of down-colorings, both \vec{G} and G' are equivalent.

As observed in [2, Obs. 2.3] we have:

Observation 2.2 *There is no function $f : \mathbb{N} \rightarrow \mathbb{N}$ with $\chi_d(\vec{G}) \leq f(D(\vec{G}))$ for all acyclic digraphs \vec{G} .*

However, although not a function of $D(\vec{G})$ alone, there are computable parameters such that $\chi_d(\vec{G})$ can be bounded by functions in terms of these parameters. That will be the purpose of the following section.

3 Hypergraph representations

In this section we discuss alternative representations of our digraph \vec{G} , and define some parameters which we will use to bound the down-chromatic number $\chi_d(\vec{G})$.

We first consider the issue of the height of digraphs. We say that two digraphs on the same set of vertices are *equivalent* if every down-coloring of one is also a valid down-coloring of the other, that is, if they induce the same undirected down-graph. We show that for any acyclic digraph \vec{G} there is an equivalent acyclic digraph \vec{G}_2 of height two with $\chi_d(\vec{G}) = \chi_d(\vec{G}_2)$.

Lemma 3.1 *Any down-graph G' of an acyclic digraph \vec{G} is also a down-graph of an acyclic digraph \vec{G}_2 of height two.*

Proof. The derived digraph \vec{G}_2 has the same vertex set as \vec{G} , while the edges all go from $\max\{\vec{G}\}$ to $V(\vec{G}) \setminus \max\{\vec{G}\}_2$, where $(u, v) \in E(\vec{G}_2)$ if, and only if, $v \in D(u)$. In this way we see that two vertices in \vec{G} have a common ancestor if, and only if, they have a common ancestor in \vec{G}_2 . Hence, we have the proposition. \square

Therefore, when considering down-colorings of digraphs, we can by Lemma 3.1 assume them to be of height two.

Recall that a *hypergraph* is H is set system on V , that is $H = (V, \mathcal{E})$ where V is a set of vertices and \mathcal{E} is a set (possibly a multiset) of subsets of V called *hyperedges*. A hypergraph is *simple* if \mathcal{E} is not a proper multiset (that is, $\mathcal{E} \subseteq \mathbb{P}(V)$, the power set of V), and each hyperedge has cardinality 2 or more. For a given hypergraph H , simple or not, denote by $V(H)$ the set of its vertices and $\mathcal{E}(H)$ the set of its hyperedges. To every simple hypergraph H there is an associated simple *clique graph* G on the same vertices as H where two vertices are connected iff they are contained in the same hyperedge. Note that two distinct simple hypergraphs can have identical clique graphs.

There is a natural correspondence between acyclic digraphs and certain hypergraphs.

Definition 3.2 For a digraph \vec{G} , the corresponding down-hypergraph $H_{\vec{G}}$ of \vec{G} is defined by:

$$\begin{aligned} V(H_{\vec{G}}) &= V(\vec{G}) \setminus \max\{\vec{G}\}, \\ \mathcal{E}(H_{\vec{G}}) &= \{D(u) : u \in \max\{\vec{G}\}\}. \end{aligned}$$

Conversely, for a hypergraph H the corresponding up-digraph \vec{G}_H of H is defined by:

$$\begin{aligned} V(\vec{G}_H) &= V(H) \cup \{w_e : e \in \mathcal{E}(H)\}, \\ E(\vec{G}_H) &= \{(w_e, u) : u \in e \in \mathcal{E}(H)\}. \end{aligned}$$

Note that with the notation from above we have for any digraph \vec{G} with no isolated vertices that $\vec{G}_{H_{\vec{G}}} = \vec{G}_2$, the equivalent digraph of height two from here above. We summarize in the following:

Observation 3.3 For any hypergraph H we have $H_{\vec{G}_H} = H$ and for any digraph \vec{G} of height two with no isolated vertices we have $\vec{G}_{H_{\vec{G}}} = \vec{G}$.

Hence, for our down-coloring purposes, digraphs are equivalent to digraphs of height two with no isolated vertices, which then again are equivalent to hypergraphs, where vertices in the same hyperedge receive different colors. This is precisely a *strong coloring* of a hypergraph H , that is a map $\Psi : V(H) \rightarrow [k]$ such that $u, v \in e$ for some $e \in \mathcal{E}(H)$, implies $\Psi(u) \neq \Psi(v)$. The *strong chromatic number* $\chi_s(H)$ is the least number k of colors for which H has a proper strong coloring $\Psi : V(H) \rightarrow [k]$. Just as for graphs, when considering strong colorings of hypergraphs, we can, with no loss of generality, restrict to simple hypergraphs.

For an acyclic digraph \vec{G} we see that an optimal strong coloring of $H_{\vec{G}}$ will yield and optimal down-coloring of \vec{G} , simply by completing the colorings of $\max\{\vec{G}\}$ in a greedy fashion. In the case where $\chi_s(H_{\vec{G}}) = \sigma(H_{\vec{G}})$, then since $D(\vec{G}) = \sigma(H_{\vec{G}}) + 1$, we have $\chi_d(\vec{G}) = \chi_s(H_{\vec{G}}) + 1$. Otherwise, when $\chi_s(H_{\vec{G}}) > \sigma(H_{\vec{G}})$, we always have at least one available color from the set $\{1, 2, \dots, \chi_s(H_{\vec{G}})\}$ to complete the down-coloring of \vec{G} in a legitimate and optimal fashion. Hence we have $\chi_d(\vec{G}) = \chi_s(H_{\vec{G}})$ in this case. We summarize in the following.

Theorem 3.4 For an acyclic digraph \vec{G} we have

$$\chi_d(\vec{G}) = \begin{cases} \chi_s(H_{\vec{G}}) + 1 & \text{if } \chi_s(H_{\vec{G}}) = \sigma(H_{\vec{G}}), \\ \chi_s(H_{\vec{G}}) & \text{if } \chi_s(H_{\vec{G}}) > \sigma(H_{\vec{G}}). \end{cases}$$

We can also characterize the down chromatic number precisely in terms of the strong chromatic number of related hypergraph. The *closed down hypergraph* $\hat{H}_{\vec{G}}$ has the same vertex set as $H_{\vec{G}}$ but the edgeset $E(\hat{H}_{\vec{G}}) = \{D[u] : u \in \max\{\vec{G}\}\}$.

Observation 3.5 For an acyclic digraph \vec{G} , we have

$$\chi_d(\vec{G}) = \chi_s(\hat{H}_{\vec{G}}).$$

The down-graph G' of \vec{G} is precisely the clique-graph of the closed down-hypergraph $\hat{H}_{\vec{G}}$.

Computable bounds

Two vertices of a hypergraph H are *neighbors* in H if they are contained in the same edge in $\mathcal{E}(H)$. An edge in $\mathcal{E}(H)$ containing just one element is called *trivial*. The largest cardinality of a hyperedge of H will be denoted by $\sigma(H)$. The *degree* $d_H(u)$, or just $d(u)$, of a vertex $u \in V(H)$ is the number of non-trivial edges containing u . The minimum and maximum degree of H are given by $\delta(H) = \min_{u \in V(H)} \{d_H(u)\}$ and $\Delta(H) = \max_{u \in V(H)} \{d_H(u)\}$ respectively. The subhypergraph $H[S]$ of H , induced by a set S of vertices, is given by

$$\begin{aligned} V(H[S]) &= S, \\ \mathcal{E}(H[S]) &= \{X \cap S : X \in \mathcal{E}(H) \text{ and } |X \cap S| \geq 2\}. \end{aligned}$$

Definition 3.6 Let H be a simple hypergraph. The *degeneracy* or the *inductiveness* of H , denoted by $\text{ind}(H)$, is given by

$$\text{ind}(H) = \max_{S \subseteq V(H)} \{\delta(H[S])\}.$$

If $k \geq \text{ind}(H)$, then we say that H is *k-degenerate* or *k-inductive*.

Note that Definition 3.6 is a generalization of the degeneracy or the inductiveness of a usual undirected graph G , given by $\text{ind}(G) = \max_{H \subseteq G} \{\delta(H)\}$. Note that the degeneracy of a (hyper)graph is always greater than or equal to the degeneracy of any of its sub(hyper)graphs.

To illustrate, let us for a brief moment discuss the degeneracy of an important class of simple graphs, namely that of simple planar graphs. Every subgraph of a simple planar graph is again planar. Since every planar graph has a vertex of degree five or less, the degeneracy of every planar graph is at most five. This is the best possible for planar graphs, since the graph of the icosahedron is planar and 5-regular. That a planar graph has degeneracy of five, implies that it can be vertex colored in a simple greedy fashion with at most six colors. The degeneracy has also been used to bound the chromatic number of

the square G^2 of a planar graph G , where G^2 is a graph obtained from G by connecting two vertices of G if, and only if, they are connected in G or they have a common neighbor in G (see [4].) In general, the degeneracy of an undirected graph G yields an ordering $\{u_1, u_2, \dots, u_n\}$ of $V(G)$, such that each vertex u_i has at most $\text{ind}(G)$ neighbors among the previously listed vertices u_1, \dots, u_{i-1} . Such an ordering provides a way to vertex color G with at most $\text{ind}(G) + 1$ colors in an efficient greedy way, and hence we have in general that $\chi(G) \leq \text{ind}(G) + 1$.

The degeneracy of a simple hypergraph is also connected to a greedy vertex coloring of it, but not in such a direct manner as for a regular undirected graph, since, as noted, the number of neighbors of a given vertex in a hypergraph is generally much larger than its degree.

Theorem 3.7 *If the simple undirected graph G is the clique graph of the simple hypergraph H then $\text{ind}(G) \leq \text{ind}(H)(\sigma(H) - 1)$.*

Proof. For each $S \subseteq V(G) = V(H)$, let $G[S]$ and $H[S]$ be the subgraph of G and the subhypergraph of H induced by S , respectively. Note that for each $u \in S$, each hyperedge in $H[S]$ which contains u , has at most $\sigma(H[S]) - 1 \leq \sigma(H) - 1$ other vertices in addition to u . By definition of $d_{H[S]}(u)$, we therefore have that $d_{G[S]}(u) \leq d_{H[S]}(u)(\sigma(H) - 1)$, and hence

$$\delta(G[S]) \leq \delta(H[S])(\sigma(H) - 1). \quad (2)$$

Taking the maximum of (2) among all $S \subseteq V(G)$ yields the theorem. \square

Recall that the *intersection graph* of a collection $\{A_1, \dots, A_n\}$ of sets, is the simple graph with vertices $\{u_1, \dots, u_n\}$, where we connect u_i and u_j if, and only if, $A_i \cap A_j \neq \emptyset$.

Directly by definition of the inductiveness we have the following.

Observation 3.8 *For a simple connected hypergraph H , then $\text{ind}(H) = 1$ if, and only if, the intersection graph of its hyperedges $\mathcal{E}(H)$ is a tree.*

What Observation 3.8 implies, is that edges of H can be ordered as $\mathcal{E}(H) = \{e_1, \dots, e_m\}$, such that each e_i intersects exactly one edge from the set $\{e_1, \dots, e_{i-1}\}$. If now G is the clique graph of H , this implies that $\text{ind}(G) = \sigma(H) - 1$ and hence $\chi(G) = \sigma(H)$. Therefore, by Theorem 3.4, we have in general the following:

If $\text{ind}(H_{\vec{G}}) = 1$, then $\chi_s(H_{\vec{G}}) = \chi(G) = \sigma(H_{\vec{G}})$, and hence $\chi_D(\vec{G}) = \sigma(H_{\vec{G}}) + 1 = D(\vec{G})$. Otherwise, if $\text{ind}(H_{\vec{G}}) > 1$, then by Theorem 3.7 we have

$$\chi_s(H_{\vec{G}}) = \chi(G) \leq \text{ind}(G) + 1 \leq \text{ind}(H_{\vec{G}})(\sigma(H_{\vec{G}}) - 1) + 1.$$

Since now $D(\vec{G}) = \sigma(H_{\vec{G}}) + 1$ we have therefore the following corollary.

Corollary 3.9 *If \vec{G} is an acyclic digraph, then its down-chromatic number satisfies the following:*

1. *If $\text{ind}(H_{\vec{G}}) = 1$ then $\chi_d(\vec{G}) = D(\vec{G})$.*

2. If $\text{ind}(H_{\vec{G}}) > 1$ then $\chi_d(\vec{G}) \leq \text{ind}(H_{\vec{G}})(D(\vec{G}) - 2) + 1$.

Moreover, in both cases the given upper bound of colors can be used to down-color \vec{G} in an efficient greedy fashion.

EXAMPLE: Let $k, m \in \mathbb{N}$, let A_1, \dots, A_k be disjoint sets, each A_i containing exactly m vertices. Let $H(k, m)$ be the hypergraph with

$$\begin{aligned} V(H(k, m)) &= \bigcup_{i \in [k]} A_i, \\ \mathcal{E}(H(k, m)) &= \{A_i \cup A_j : i \neq j, \{i, j\} \subseteq [k]\}, \end{aligned}$$

Let $\vec{G}(k, m) = \vec{G}_{H(k, m)}$ be the up-digraph of the hypergraph $H(k, m)$. Clearly $\vec{G}(k, m)$ is a simple acyclic digraph on $km + \binom{k}{2}$ vertices and with $\binom{k}{2} \cdot 2m = k(k-1)m$ directed edges. Further, $\sigma(H(k, m)) = 2m$ and so $D(\vec{G}(k, m)) = 2m + 1$. Since each vertex is contained in exactly k hyperedges we have $\text{ind}(H(k, m)) = k - 1$. Hence, by Corollary 3.9, we obtain that $\chi_d(\vec{G}(k, m)) \leq (k-1)(2m-1) + 1 = \Theta(km)$, which agrees with the asymptotic value of the actual down-chromatic number km (also a $\Theta(km)$ function). Hence, up to a constant (of 2), Corollary 3.9 is asymptotically tight.

4 Discrepancy between parameters

So far we have discussed how to approximate the down-chromatic number $\chi_d(\vec{G})$ of an acyclic digraph \vec{G} in terms of $D(\vec{G})$ and $\text{ind}(H_{\vec{G}})$, the inductiveness of the corresponding down-hypergraph. In this section we will discuss the relative discrepancy between $D(\vec{G})$ and the actual down chromatic number $\chi_d(\vec{G})$, and determine a tight asymptotic upper bound for their ratio.

If $H(k, m)$ is the hypergraph defined in the previous section, then for $\vec{G}_{H(k, m)}$ we clearly have

$$\frac{\chi_d(\vec{G}_{H(k, m)})}{D(\vec{G}_{H(k, m)})} = \frac{km}{2m+1} \rightarrow \infty$$

as $k \rightarrow \infty$ and m is fixed. Hence, allowing an unbounded number of vertices of \vec{G} , the above ratio clearly can become arbitrarily large even when $D(\vec{G}(k, m)) = 2m + 1$ is fixed.

The purpose of this last section is to derive a tight upper bound for $\chi_d(\vec{G})/D(\vec{G})$ among all acyclic digraphs \vec{G} with $D(\vec{G})$ bounded and with bounded number of vertices.

Definition 4.1 For $n, \delta \in \mathbb{N}$ define the relative down-coloring discrepancy, or simply the rdcd , $d_d(\delta, n)$ by

$$d_d(\delta, n) = \max_{|V(\vec{G})| \leq n, D(\vec{G}) \leq \delta} \left\{ \frac{\chi_d(\vec{G})}{D(\vec{G})} \right\},$$

where the maximum is among all acyclic digraphs \vec{G} satisfying the stated conditions.

Note that for a hypergraph H , it holds that $D(\vec{G}_H) = \sigma(H) + 1$ and $|V(\vec{G}_H)| = |V(H)| + |\mathcal{E}(H)|$. Hence, for $n, \sigma \in \mathbb{N}$ we define the *relative strong-coloring discrepancy*, or simply the *rscd*, $d_s(\sigma, n)$ by

$$d_s(\sigma, n) = \max_{|V(H)| + |\mathcal{E}(H)| \leq n, \sigma(H) \leq \sigma} \left\{ \frac{\chi_s(H)}{\sigma(H) + 1} \right\}, \quad (3)$$

where the maximum is taken among all hypergraphs H . By Lemma 3.1 and Observation 3.3 we have the following.

Observation 4.2 *For $n, \delta, \sigma \in \mathbb{N}$ we have $d_d(\sigma + 1, n) = d_s(\sigma, n)$.*

Although our original motivation for the relative discrepancy $d_d(\delta, n)$ is given by Definition 4.1, by Observation 4.2 it suffices to (and in some ways is more natural to) determine a tight upper bound of $d_s(\sigma, n)$ for given $n, \sigma \in \mathbb{N}$ from (3).

Definition 4.3 *For $n, \sigma \in \mathbb{N}$, let $r^+(\sigma, n)$ denote the positive root of the quadratic polynomial $x + \frac{x(x-1)}{\sigma(\sigma-1)} = n$ in terms of x .*

Using Definition 4.3 we now can state our first theorem.

Theorem 4.4 *For $n, \sigma \in \mathbb{N}$ the rscd $d_s(\sigma, n)$ satisfies*

$$d_s(\sigma, n) \leq \frac{r^+(\sigma, n)}{\sigma + 1}$$

Proof. Let H be a hypergraph with $|V(H)| + |\mathcal{E}(H)| \leq n$, $\chi_s(H) = x \in \mathbb{N}$ and $\sigma(H) = \sigma$. In this case there is an optimal strong x -coloring of the vertices of H . Let V_1, \dots, V_x be corresponding partition of $V(H)$ into color classes. For each i and j with $1 \leq i < j \leq x$, there is at least one hyperedge $e_{ij} \in \mathcal{E}(H)$ that contains one vertex from V_i and one vertex from V_j . Since $|e| \leq \sigma$ for each $e \in \mathcal{E}(H)$, each hyperedge can cover at most $\binom{\sigma}{2}$ sets of two vertices that are colored by distinct pairs of colors. Since there are $\binom{x}{2}$ pairs of colors, the number of hyperedges of H must satisfy

$$|\mathcal{E}(H)| \geq \frac{\binom{x}{2}}{\binom{\sigma}{2}} = \frac{x(x-1)}{\sigma(\sigma-1)}.$$

Since each color class V_i is nonempty, we must have $|V(H)| \geq \chi_s(H) = x$. Combining the last two inequalities we obtain, in particular, that

$$n \geq |V(H)| + |\mathcal{E}(H)| \geq x + \frac{x(x-1)}{\sigma(\sigma-1)}. \quad (4)$$

Viewing n and σ as arbitrary but fixed, we obtain by (3) and (4) that

$$d_s(\sigma, n) \leq \max_{x + \frac{x(x-1)}{\sigma(\sigma-1)} \leq n} \left\{ \frac{x}{\sigma + 1} \right\}. \quad (5)$$

By solving the corresponding quadratic inequality $x + \frac{x(x-1)}{\sigma(\sigma-1)} \leq n$ in terms of x , keeping in mind that x is positive, we have that $0 < x < r^+(\sigma, n)$. The maximum value of the fraction $x/(\sigma + 1)$ is clearly taken when x is at maximum, that is for $x = r^+(\sigma, n)$. Hence, by (5) we obtain

$$d_s(\sigma, n) \leq \frac{r^+(\sigma, n)}{\sigma + 1},$$

which completes the proof. \square

By solving the quadratic equation $x + x(x - 1)/(\sigma(\sigma - 1)) = n$ for x we get

$$r^+(\sigma, n) = \frac{\sqrt{\sigma(\sigma - 1)n}}{\sqrt{1 + \frac{(\sigma(\sigma - 1) - 1)^2}{4\sigma(\sigma - 1)n} + \frac{\sigma(\sigma - 1) - 1}{\sqrt{4\sigma(\sigma - 1)n}}}},$$

it is immediate that $r^+(\sigma, n) \leq \sqrt{\sigma(\sigma - 1)n}$, and

$$\lim_{n \rightarrow \infty} \frac{r^+(\sigma, n)}{\sqrt{n}} = \sqrt{\sigma(\sigma - 1)}. \quad (6)$$

Hence, by Theorem 4.4 we obtain

Corollary 4.5 *For $n, \sigma \in \mathbb{N}$ the rscd $d_s(\sigma, n)$ satisfies*

$$d_s(\sigma, n) \leq \frac{\sqrt{\sigma(\sigma - 1)}}{\sigma + 1} \sqrt{n}.$$

Note that for a fixed σ the upper bounds for $d_s(\sigma, n)$ in Theorem 4.4 and Corollary 4.5 are by (6) asymptotically the same as $n \rightarrow \infty$. We now argue that the upper bound from Corollary 4.5 is asymptotically tight in the sense that

$$\limsup_{n \rightarrow \infty} \frac{d_s(\sigma, n)}{\sqrt{n}} = \frac{\sqrt{\sigma(\sigma - 1)}}{\sigma + 1} \quad (7)$$

for infinitely many values of σ . More specifically, we will show that there is an infinite collection $(\sigma_i)_{i \geq 1}$ such that for each i there is again an infinite collection $(n_{ij})_{j \geq 1}$ with the property that there exists a hypergraph H_{ij} with $|V(H_{ij})| + |\mathcal{E}(H_{ij})| = n_{ij}$ and $\sigma(H_{ij}) = \sigma_i$ that matches the upper bound of Theorem 4.4, that is

$$\frac{\chi_s(H_{ij})}{\sigma(H_{ij}) + 1} = \frac{r^+(\sigma_i, n_{ij})}{\sigma_i + 1}$$

for each i and j . This together with Theorem 4.4 and Corollary 4.5 will yield (7). For this we need some additional terminology for hypergraphs.

Recall that a *balanced incomplete block design* or a *BIBD* for short, is a simple hypergraph $H = (V, \mathcal{B})$ where V is a finite set of vertices and $\mathcal{B} \subseteq \mathbb{P}(V)$ is a collection of hyperedges where (i) all the hyperedges have the same cardinality that is strictly less than

that of V , (ii) each vertex is contained in the same $r > 0$ number of hyperedges, and (iii) each pair of vertices is contained in exactly $\lambda > 0$ hyperedges. In this case the vertices are sometimes called *varieties* and the hyperedges *blocks*. In a series of three papers [18], [19] and [20], R. M. Wilson proved that for any given $k, \lambda \in \mathbb{N}$ there exists a constant $C = C(k, \lambda) \in \mathbb{N}$ such that for any $v \geq C$ satisfying (i) $\lambda(v - 1) \equiv 0 \pmod{k - 1}$ and (ii) $\lambda v(v - 1) \equiv 0 \pmod{k(k - 1)}$, then there exists a BIBD on v vertices, where each hyperedge has cardinality k and where each vertex is contained in λ hyperedges. In particular, for $\lambda = 1$, we have with our notation and terminology from above the following.

Corollary 4.6 *For each $\sigma \in \mathbb{N}$ there exists a $K = K(\sigma) \in \mathbb{N}$ such that for all $k \geq K$ with $k^2 \equiv k \pmod{\sigma}$, we have*

$$d_s \left(\sigma, k(\sigma - 1 + k) + 1 - \frac{k(k - 1)}{\sigma} \right) = \frac{k(\sigma - 1) + 1}{\sigma + 1}.$$

Proof. By R. M. Wilson, there is a $C = C(\sigma) \in \mathbb{N}$ such that for all $v \geq C$ satisfying $v - 1 \equiv 0 \pmod{\sigma - 1}$ and $v(v - 1) \equiv 0 \pmod{\sigma(\sigma - 1)}$, there is a BIBD, call it H , on v vertices such that each hyperedge has exactly σ vertices, each pair of vertices is contained in exactly one hyperedge. In particular, the number of hyperedges of H is given by

$$|\mathcal{E}(H)| = \frac{\binom{v}{2}}{\binom{\sigma}{2}} = \frac{v(v - 1)}{\sigma(\sigma - 1)}.$$

The conditions on v mean that $v = k(\sigma - 1) + 1$ where $k^2 \equiv k \pmod{\sigma}$ and $k \in \mathbb{N}$ is large enough, say $k \geq K = K(\sigma)$. Since each pair of the v vertices is contained in a hyperedge, we clearly have $\chi_s(H) = v$. We also clearly have $\sigma(H) = \sigma$ and hence, in this case we have $\chi_s(H)/(\sigma(H) + 1) = v/(\sigma + 1) = (k(\sigma - 1) + 1)/(\sigma + 1)$. Also, for $v = k(\sigma - 1) + 1$ we have

$$n = v + \frac{v(v - 1)}{\sigma(\sigma - 1)} = k(\sigma - 1 + k) + 1 - \frac{k^2 - k}{\sigma}.$$

By (3) this implies that

$$d_s \left(\sigma, k(\sigma - 1 + k) + 1 - \frac{k(k - 1)}{\sigma} \right) \geq \frac{k(\sigma - 1) + 1}{\sigma + 1}.$$

Since $r^+ \left(\sigma, k(\sigma - 1 + k) + 1 - \frac{k(k - 1)}{\sigma} \right) = k(\sigma - 1) + 1$, we have by Theorem 4.4 the corollary. \square

We conclude this section by an explicit and self contained construction of a class of hypergraphs for which the asymptotic value in (7) can also be reached. First note that for $\sigma \in \mathbb{N}$ and $v = \sigma^k$ then $v - 1 \equiv 0 \pmod{\sigma - 1}$ and $v(v - 1) \equiv 0 \pmod{\sigma(\sigma - 1)}$ hold for all $k \in \mathbb{N}$.

Before proving Proposition 4.7 we recall some notations and results: Every finite field has cardinality of a prime power $q = p^n$. If \mathbb{Z}_p denotes the integers modulo p , then the unique field \mathbb{F}_q of cardinality q can be given as the quotient $\mathbb{F}_q = \mathbb{Z}_p[X]/(X^q - X)$ which

turns out to be the splitting field of the polynomial $X^q - X$ over \mathbb{Z}_p . In particular, \mathbb{F}_r is a subfield of \mathbb{F}_q whenever $r = p^m$ and $m \leq n$ (See [10, p. 278].) The *affine plane* over a field F is the tuple (F^2, \mathcal{L}) where \mathcal{L} is the collection of all lines $\{(x, y) : ax + by = c\} \subseteq F^2$. The affine plane over \mathbb{F}_q is a BIBD with $\lambda = 1$ (See [12, p. 199].)

Proposition 4.7 *If $\sigma = p^k$ is a prime power, then for any $m \in \mathbb{N}$ there is a BIBD on σ^m vertices, where each hyperedge has cardinality σ and where each pair of vertices is contained in exactly one hyperedge.*

Proof. (Sketch) The hypergraph on σ vertices with exactly one hyperedge containing all the vertices proves the basic case for $m = 1$.

We proceed by induction on m . Assume there is a BIBD $H = (V(H), \mathcal{E}(H))$ on σ^m vertices, where each hyperedge has cardinality σ and where each pair of vertices is contained in exactly one hyperedge. Without loss of generality we may assume $V(H) = \mathbb{F}_q$ where $q = p^{km}$. In the affine plane of \mathbb{F}_q the lines either have the form of $x = \gamma$ or $y = \alpha x + \beta$ for some $\alpha, \beta, \gamma \in \mathbb{F}_q$. Call the lines of the latter form *non-vertical*. Consider $H' = (\mathbb{F}_\sigma \times \mathbb{F}_q, \mathcal{L}')$ where \mathcal{L}' consists of all non-vertical lines from the affine plane of \mathbb{F}_q restricted to $\mathbb{F}_\sigma \times \mathbb{F}_q \subseteq \mathbb{F}_q^2$, together with $\{\{a\} \times L : a \in \mathbb{F}_\sigma \text{ and } L \in \mathcal{E}(H)\}$. Then H' is a BIBD on σ^{m+1} vertices, where each hyperedge has cardinality σ and where each pair of vertices is contained in exactly one hyperedge. This completes the proof. \square

REMARKS: (i) The condition that each pair of vertices is contained in exactly one hyperedge is a natural geometric condition called the *Euclid's first postulate*, when vertices are viewed as points and hyperedges as lines. (ii) Note that the affine plane over any field F , in particular for the finite field \mathbb{F}_q , satisfies the *Euclidean parallel postulate* (aka *Euclid's fifth postulate*), that for any three vertices, not all contained in a hyperedge, there is precisely one hyperedge containing the third vertex, that is disjoint from the unique hyperedge containing the first two vertices. (See [8].) However, this does not hold in H' from the above proof, since by restricting/truncating we introduce in general many additional hyperedges containing the third vertex that are disjoint from the unique hyperedge containing the first two vertices. In fact, the well-known conjecture, whether or not there exists an affine plane on n vertices when n is not a power of a prime, is still open. (See [5].)

From Proposition 4.7 we deduce the following corollary.

Corollary 4.8 *If $\sigma = p^k$ is a prime power, then for any $m \in \mathbb{N}$ we have*

$$d_s \left(\sigma, \sigma^m + \frac{\sigma^{m-1}(\sigma^m - 1)}{\sigma - 1} \right) = \frac{\sigma^m}{\sigma + 1}.$$

Proof. We clearly have

$$r^+ \left(\sigma, \sigma^m + \frac{\sigma^{m-1}(\sigma^m - 1)}{\sigma - 1} \right) = \sigma^m,$$

and hence by Theorem 4.4 we have

$$d_s \left(\sigma, \sigma^m + \frac{\sigma^{m-1}(\sigma^m - 1)}{\sigma - 1} \right) \leq \frac{r^+(\sigma, n)}{\sigma + 1} = \frac{\sigma^m}{\sigma + 1},$$

yielding the upper bound.

By Proposition 4.7 there is a BIBD H on σ^m vertices, where each hyperedge has cardinality σ and where each pair of vertices is contained in exactly one hyperedge. In this case we have $\chi_s(H) = \sigma^m$ and $\sigma(H) = \sigma$ and hence

$$d_s \left(\sigma, \sigma^m + \frac{\sigma^{m-1}(\sigma^m - 1)}{\sigma - 1} \right) \geq \frac{\chi_s(H)}{\sigma(H) + 1} = \frac{\sigma^m}{\sigma + 1},$$

yielding the lower bound and completes the proof. \square

REMARK: Throughout this article we have assumed our digraphs to be acyclic. However, we note that the definition of down-coloring can be easily extended to a regular cyclic digraph \vec{G} by interpreting the notion of descendants of a vertex u to mean the set of nodes reachable from u . In fact, if \vec{G} is an arbitrary digraph, then there is an equivalent acyclic digraph \vec{G}' , on the same set of vertices, with an identical down-graph: First form the *condensation* \hat{G} of \vec{G} by shrinking each strongly connected component of \vec{G} to a single vertex. Then form \vec{G}' by replacing each node of \hat{G} which represents a strongly connected component of \vec{G} on a set $X \subseteq V(\hat{G})$ of vertices, with an arbitrary vertex $u \in X$, and then add a directed edge from u to each $v \in X \setminus \{u\}$. This completes the construction.

Observe that each node $v \in X$ has exactly the same neighbors in the down-graph of \vec{G}' as u , as it is a descendant of u and u alone. Further, if node v was in a different strong component of \vec{G} than u but was reachable from u , then it will continue to be a descendant of u in \vec{G}' . Hence, the down-graphs of \vec{G} and \vec{G}' are identical.

Acknowledgments

The authors are grateful to John L. Pfaltz for his interest and encouragements to write this article. Also, sincere thanks to Jim Lawrence for pointing out the very relevant work of Richard M. Wilson on BIBDs.

References

- [1] S. Abitebboul, P. Buneman and D. Suciu. Data on the Web: From Relations to Semistructured Data and XML. *Morgan Kaufmann Publishers*, (1999).
- [2] G. Agnarsson and Á. Egilsson. On vertex coloring simple genetic digraphs. *Congressus Numerantium*, **161**, 117 – 127, (2004).

- [3] Geir Agnarsson, Ágúst Egilsson and Magnús M. Halldórsson. Proper down-coloring simple acyclic digraphs. Second International Workshop, AGTIVE-2003, Charlottesville, VA, USA, September/October 2003. *Lecture Notes of Computer Science LNCS 3062*, 299 – 312, Springer Verlag, (2004).
- [4] G. Agnarsson and M. M. Halldórsson. Coloring powers of planar graphs. *SIAM Journal of Discrete Mathematics*, **16(4)**, 651 – 662, (2003).
- [5] Kenneth P. Bogart. Introductory Combinatorics. *Academic Press*, Third edition, (2000).
- [6] Á. Egilsson and H. Gudbjartsson. Indexing of tables referencing complex structures. preprint. Available at <http://arxiv.org/abs/cs.DB/0309011>, (2003).
- [7] The Gene Ontology Consortium. Gene Ontology: Tool for the unification of biology. *Nature Genet*, **25**, 25 – 29, (2000).
- [8] Marvin Jay Greenberg. Euclidean and Non-Euclidean Geometries. *W. H. Freeman and Company, New York*, 3. ed., (1993).
- [9] C. C. Harner and R. C. Entringer. Arc colorings of digraphs. *Journal of Combinatorial Theory, Series B*, **13**, 219 – 225, (1972).
- [10] T. W. Hungerford. Algebra. Graduate Texts in Mathematics, GTM – 73, *Springer Verlag*, (1974).
- [11] H. Jacob and H. Meyniel. Extensions of Turán’s Brooks’ theorems and new notions of stability and colorings in digraphs. *Combinatorial Mathematics, North-Holland Math. Stud.*, **75**, 365 – 370, (1983).
- [12] J. H. van Lint and R. M. Wilson. A course in Combinatorics. *Cambridge University Press*, (1992).
- [13] C. A. Shaffer. A practical introduction to data structures and algorithm analysis *Java Edition, Prentice Hall*, (1998).
- [14] X. Y. Su. Brooks’ theorem on colorings of digraphs. *Fujian Shifan Daxue Xuebao Ziran Kexue Ban*, **3(1)**, 1 – 2, (1987).
- [15] W. T. Trotter. Combinatorics and Partially Ordered Sets: Dimension Theory. *Johns Hopkins Studies in the Mathematical Sciences, The Johns Hopkins University Press*, (1992).
- [16] W. Wang and K. Zhang. Colorings of hypergraphs. *Adv. Math. (China)*, **29(2)**, 115–136, (2000).
- [17] D. B. West. Introduction to graph theory. *Prentice Hall*, 2nd ed., (2001).

- [18] R. M. Wilson. An existence theory for pairwise balanced designs. I. Composition theorems and morphisms. *J. Combinatorial Theory Ser. A*, **13**, 220–245, (1972).
- [19] R. M. Wilson. An existence theory for pairwise balanced designs. II. The structure of PBD-closed sets and the existence conjectures. *J. Combinatorial Theory Ser. A*, **13**, 246–273, (1972).
- [20] R. M. Wilson. An existence theory for pairwise balanced designs. III. Proof of the existence conjectures. *J. Combinatorial Theory Ser. A*, **18**, 71–79, (1975).

July 6, 2007