Combinatorics, Probability and Computing (2008) 17, 577–589. © 2008 Cambridge University Press
doi:10.1017/S0963548308009140 Printed in the United Kingdom

Graph Colouring with No Large Monochromatic Components

NATHAN LINIAL,[1†] JIŘÍ MATOUŠEK[2],
OR SHEFFET[1‡] and GÁBOR TARDOS[3§]

[1]School of Computer Science and Engineering, Hebrew University, Jerusalem, Israel
(e-mail: nati@cs.huji.ac.il; or.sheffet@gmail.com)

[2]Department of Applied Mathematics and Institute of Theoretical Computer Science (ITI),
Charles University, Malostranské nám. 25, 118 00 Praha 1, Czech Republic
(e-mail: matousek@kam.mff.cuni.cz)

[3]School of Computing Science, Simon Fraser University, Burnaby, BC, Canada
and
Rényi Institute, Budapest, Hungary
(e-mail: tardos@cs.sfu.ca)

Received 16 March 2007; revised 1 May 2008; first published online 13 June 2008

For a graph G and an integer t we let $mcc_t(G)$ be the smallest m such that there exists a colouring of the vertices of G by t colours with no monochromatic connected subgraph having more than m vertices. Let $\mathcal{F}$ be any non-trivial minor-closed family of graphs. We show that $mcc_2(G) = O(n^{2/3})$ for any n-vertex graph $G \in \mathcal{F}$. This bound is asymptotically optimal and it is attained for planar graphs. More generally, for every such $\mathcal{F}$, and every fixed t we show that $mcc_t(G) = O(n^{2/(t+1)})$. On the other hand, we have examples of graphs G with no K_{t+3} minor and with $mcc_t(G) = \Omega(n^{2/(2t-1)})$.

It is also interesting to consider graphs of bounded degrees. Haxell, Szabó and Tardos proved $mcc_2(G) \leqslant 20000$ for every graph G of maximum degree 5. We show that there are n-vertex 7-regular graphs G with $mcc_2(G) = \Omega(n)$, and more sharply, for every $\varepsilon > 0$ there exists $c_\varepsilon > 0$ and n-vertex graphs of maximum degree 7, average degree at most $6 + \varepsilon$ for all subgraphs, and with $mcc_2(G) \geqslant c_\varepsilon n$. For 6-regular graphs it is known only that the maximum order of magnitude of mcc_2 is between $\sqrt{n}$ and n.

We also offer a Ramsey-theoretic perspective of the quantity $mcc_t(G)$.

† Supported by the Israel Science Foundation.
‡ This paper is partially based on this author's undergraduate *Amirim* Honors Program project.
§ Supported by NSERC grant 611470 and the Hungarian Foundation for Scientific Research Grant (OTKA) Nos T037846, T046234, AT048826 and NK62321.

https://doi.org/10.1017/S0963548308009140 Published online by Cambridge University Press

1. Introduction

In the classical graph colouring problem we assign a colour to each vertex so that no two vertices of the same colour are adjacent. In other words, each *monochromatic connected component* must be a single vertex. In the problems that we study here, this requirement is relaxed and we only demand that monochromatic connected components should have small cardinality. Concretely, for a graph G and an integer t we define $mcc_t(G)$ as the smallest integer m such that the vertices of G can be t-coloured so that no monochromatic connected component has cardinality exceeding m. In particular, $mcc_t(G) = 1$ if and only if G can be properly t-coloured.

Here are some technicalities, before we survey earlier work on this subject. When we consider a graph G, n always denotes the number of vertices. For a set $S \subseteq V(G)$, we let $G[S]$ denote the subgraph of G induced by S. We use the standard asymptotic language and conveniently ignore integrality issues in our computations. Some of the examples we consider are line graphs. We recall that the line graph of a graph H is defined as $L(H) = (E(H), \{\{e, e'\} : e \cap e' \neq \emptyset\})$. Clearly, $mcc_t(L(H))$ is the smallest m such that there is a t-colouring of the edges of H so that every monochromatic connected subgraph of H has $\leqslant m$ edges.

The simplest form the well-known HEX lemma states that, in any two-colouring of the triangulated m by m grid G_m, we either have a path from the left edge of the grid to the right edge in one of the colours or else we have a path from the top edge to the bottom edge in the other colour. As a consequence we have that $mcc_2(G_m) \geqslant m$. It is easy to see that equality holds.

The earliest reference we are aware of that investigates the parameter mcc_t directly is Kleinberg, Motwani, Raghavan and Venkatasubramanian [15], where the question was motivated by a problem in computer science concerning dynamically evolving databases. Among other results, the authors prove that mcc_3 is unbounded for planar graphs, and that there is a constant $\varepsilon > 0$ such that, for all (sufficiently large) d, there are d-regular graphs G with $mcc_{\varepsilon\sqrt{d}}(G) = \Omega(n)$.

Apparently independently, the possibility of bounding $mcc_t(G)$ by a constant for graphs of bounded degree has been investigated by graph theorists. The results mainly concern the case $t = 2$. It is easy to see that $mcc_2(G) \leqslant 2$ for any graph G of maximum degree 3. Alon, Ding, Oporowski and Vertigan [1] proved that $mcc_2(G) \leqslant 57$ for every graph G of maximum degree 4. Haxell, Szabó and Tardos [9] improved this to $mcc_2(G) \leqslant 6$ and proved that $mcc_2(G) \leqslant 20000$ for every graph G of maximum degree 5. Robert Berke [3] improved the latter bound to 1908. On the other hand, the HEX lemma shows that that for graphs G with maximum degree 6, $mcc_2(G)$ can be arbitrarily large. For graphs G of maximum degree 3, Berke and Szabó [4] also showed that they admit two-colouring where one colour induces an independent set, while the other colour induces components of size at most 189. Earlier work on this subject [7, 11] mainly focused on more specific questions concerning line graphs of 3-regular graphs. These investigations culminated in Thomassen's paper [24], showing that the edges of every 3-regular graph can be 2-coloured so that each monochromatic component is a path of length at most 5.

The Hadwiger conjecture is one of the most intriguing open problems on graph colouring. In a recent paper, Kawarabayashi and Mohar [14] consider an approach to it that is based on colouring with no large monochromatic connected components for minor closed families of graphs. Their main result says that there exist a function f such that $mcc_{\lceil 15.5k \rceil}(G) \leqslant f(k)$ for any $k > 1$ and any graph G having no K_k minor.

https://doi.org/10.1017/S0963548308009140 Published online by Cambridge University Press

We should also mention that there is a fairly rich literature that deals with the notion of t-vertex colouring where each monochromatic connected component has a small *diameter*: see, *e.g.*, [16]. This line of research originated in the field of distributed computing.

Here are the main results of the present paper. First we show that the triangulated grid of the HEX lemma does *not* give maximal mcc_2 value among planar graphs with a fixed number of vertices, and find the order of magnitude of this maximum.

Theorem 1.1. *For every planar graph G we have $mcc_2(G) = O(n^{2/3})$, and this bound is tight.*

More generally, for every non-trivial minor-closed family of graphs $\mathcal{F}$ (i.e., $\mathcal{F}$ does not contain all graphs) and $t \geqslant 2$, there exists a constant $C = C_{\mathcal{F},t}$ such that if G belongs to $\mathcal{F}$ then $mcc_t(G) \leqslant Cn^{2/(t+1)}$. On the other hand, for every t there exist graphs with no K_{t+3} minor and with $mcc_t(G) \geqslant C_0 n^{2/(2t-1)}$). Here $C_0 > 0$ is an absolute constant.

For bounded tree-width families of graphs, we find how large $mcc_t(G)$ can become for every $t \geqslant 2$.

Theorem 1.2. *If the tree-width of G is bounded by a constant* (*equivalently, if G excludes a fixed* planar *minor*), *then $mcc_t(G) = O(n^{1/t})$. This bound is asymptotically optimal for every fixed t.*

The following theorem contrasts with the result of Haxell, Szabó and Tardos [9] on graphs of maximum degree 5 mentioned above. It would be interesting to find how big $mcc_2(G)$ can become for graphs of maximum degree 6. The HEX lemma proves that it can be as large as $\sqrt{n}$ but we do not know whether it can be linear.

The following theorem contrasts the result of [9] of maximum degree 5 graphs mentioned above. It would be interesting to find the how big $mcc_2(G)$ can become with maximum degree 6 graphs. The HEX lemma proves that it can be as large as $\sqrt{n}$ but we do not know whether it can be linear.

Theorem 1.3. *For every $\varepsilon > 0$ there exists a constant $c_\varepsilon > 0$ and arbitrarily large graphs G so that:*

- *every vertex in G has degree at most 7,*
- *every subgraph of G has average degree at most $6 + \varepsilon$,*
- *$mcc_2(G) \geqslant c_\varepsilon n$.*

As already mentioned, the questions we consider here have independently originated in computer science and in graph theory. Graph colouring is, of course, one of the most fascinating parts of graph theory. Due to its great significance and associated famous open questions, many different variations on the basic theme are being investigated (see, *e.g.*, [13]), and the present problems can be viewed as part of this ongoing research effort.

What is less obvious is the connection between the graph invariants we consider here and Ramsey Theory. The usual perception is that Ramsey-type theorems express the fact that large systems necessarily contain 'highly regular islands'. We suggest that many Ramsey-type results can be viewed as 'sum theorems'. Specifically, let $\mathcal{G}$ be a class of graphs closed under taking

https://doi.org/10.1017/S0963548308009140 Published online by Cambridge University Press

subgraphs. Given a graph G, we ask for the smallest number of members in $\mathcal{G}$ whose union is G. Thus, when $\mathcal{G}$ consists of all graphs without a k-clique we encounter the classical Ramsey problem. When it is the class of all graphs not containing a given subgraph we recover the so-called Graph Ramsey Theory. Finally, if $\mathcal{G}$ contains all graphs in which each connected component has small cardinality, we arrive at our present problem. We believe that this perspective deserves further research. Needless to say, this concept extends beyond graphs and unions. Other mathematical objects and other appropriate operations can be considered. Given an object G and a class $\mathcal{G}$, one seeks the most economical way of expressing G as a 'sum' of members in $\mathcal{G}$.

There are certain classes of graphs for which the study of colouring with small monochromatic connected components is particularly interesting. Let D_m^d denote the d-dimensional grid with all diagonals; that is, D_m^d is the graph with vertex set $\{1, 2, \ldots, m\}^d$ where the two vertices u, v are adjacent if $\|u - v\|_\infty = \max_i |u_i - v_i| \leqslant 1$. The study of *mcc* for this graph and its relatives leads to very interesting problems which bring together combinatorics, geometry and topology. The d-dimensional version of the well-known HEX lemma (see Gale [8] and Linial and Saks [16]) implies that $mcc_d(D_m^d) \geqslant m$ (and a simple colouring shows an $O(m)$ upper bound for every fixed d). More colours allow for constant size components, *i.e.*, we have $mcc_{d+1}(D_m^d) = O(1)$ for every fixed d. For two colours, Matoušek and Přívětivý [19] proved $mcc_2(D_m^d) \geqslant m^{d-1} - d^2 m^{d-2}$, which nearly matches the obvious upper bound (layer-by-layer 2-colouring) of m^{d-1}. The behaviour of $mcc_t(D_m^d)$ for $3 \leqslant t < d$ is still unknown and remains an intriguing open problem.

2. Excluded minors, separators, and colouring

In this section we prove the upper bounds in Theorems 1.1 and 1.2.

A subset $C \subseteq V$ of the vertex set of a graph $G = (V, E)$ is called a *separator* if no component of $G[V \setminus C]$ has more than $\frac{2}{3}|V|$ vertices (the choice of the constant $\frac{2}{3}$ is somewhat arbitrary). By the well-known planar separator theorem of Lipton and Tarjan [17], every planar graph G has a separator with at most $O(\sqrt{n})$ vertices. More generally, for every h-vertex graph H, every G containing no minor isomorphic to H has a separator with at most $h^{3/2}\sqrt{n}$ vertices [2]. We will also need that graphs of bounded tree width have bounded size separators. In particular, a graph of tree width w has a separator of size $w + 1$: see [21].

In view of these results, the upper bounds in Theorems 1.1 and 1.2 all follow from the next proposition.

Proposition 2.1. *Let $\mathcal{G}$ be a class of graphs closed under taking induced subgraphs such that every $G \in \mathcal{G}$ has a separator with at most Kn^γ vertices, where $\gamma \in [0, 1)$ and K are constants depending only on $\mathcal{G}$. Then, for every $G \in \mathcal{G}$ we have*

$$mcc_2(G) = O(n^{1/(2-\gamma)}),$$

and more generally, for any $t \geqslant 2$,

$$mcc_t(G) = O(n^{1/(t-(t-1)\gamma)}),$$

where the hidden constants of proportionality depend on K, γ, and (in the second case) on t. □

https://doi.org/10.1017/S0963548308009140 Published online by Cambridge University Press

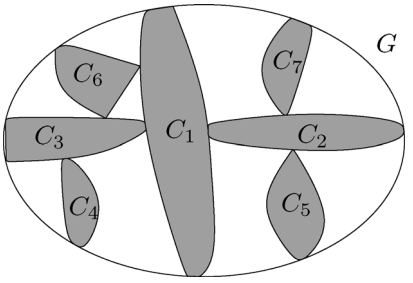


Figure 1. The recursive algorithm for 2-colouring via separators.

Proof. First we deal with the special case $t = 2$. Given an n-vertex graph G, let $n_0 := \lfloor n^{1/(2-\gamma)} \rfloor$ be a threshold parameter. We present a simple algorithm producing a 2-colouring of $V(G)$. The algorithm maintains a set S of vertices, initialized to $\emptyset$ and the list $\mathscr{L}$ of the connected components of $G[V(G) \setminus S]$ initialized to $\{G\}$. While $\mathscr{L}$ contains at least one graph with more than n_0 vertices, we select one such graph $G_i \in \mathscr{L}$ arbitrarily, we remove it from $\mathscr{L}$, we find a separator C_i of G_i of size at most $K|V(G_i)|^\gamma$, we set $S := S \cup C_i$, and we add all of the components of $G_i[V(G_i) \setminus C_i]$ to $\mathscr{L}$. The algorithm ends when $\mathscr{L}$ contains only graphs of size at most n_0; at this moment, we colour the vertices of S blue and all remaining vertices (*i.e.*, the vertices of all graphs in $\mathscr{L}$) red and we finish. The algorithm is illustrated in Figure 1.

By construction, no red component in this colouring has more than n_0 vertices, and it suffices to show $|S| \leqslant 3Kn^{1/(2-\gamma)}$ at the end of the algorithm (S can form a single blue component at worst). We use the following charging scheme. Whenever we select a separator C_i in a graph G_i, we let each vertex of G_i pay $K|V(G_i)|^{\gamma-1}$ units. Since $|C_i| \leqslant K|V(G_i)|^\gamma$, the total paid by all vertices of G_i at this step is at least $|C_i|$. To bound the size of S it is enough to bound the total charge paid by all vertices in all steps. We consider an individual vertex $v \in V(G)$ and we bound the total charge paid by it throughout the whole algorithm. There may be several successive charges, since v first pays as a vertex of G, and then possibly as a vertex of some of the G_i. Let x_j be the amount paid when v is charged the jth time, $j = 1, 2, \ldots, q$. We observe that this is a fast-increasing sequence as $x_{j-1} \leqslant (2/3)^{1-\gamma} x_j$ for all j, since the component of G_i containing v always has at most two-thirds of the vertices of G_i. The largest charge paid is $x_q \leqslant K n_0^{\gamma-1}$, since only graphs G_i with at least n_0 vertices get partitioned, and hence the x_j are bounded from above by a geometric series. Thus the total charge paid by v is $O(n_0^{\gamma-1})$. So $|S| \leqslant O(nn_0^{\gamma-1}) = O(n^{1/(2-\gamma)})$.

Next, we consider the case of $t > 2$ colours. We proceed by induction on t, assuming that for every n-vertex graph $G \in \mathscr{G}$ we can construct a colouring with $t-1$ colours witnessing $mcc_{t-1}(G) = O(n^{1/(t-1-(t-2)\gamma)})$.

For the induction step, we consider a $G \in \mathscr{G}$ and we apply to it the algorithm above with the following modifications. This time we let the threshold be $n_0 := \lfloor n^{1/(t-(t-1)\gamma)} \rfloor$, and, at the end, we colour $G[S]$ by $t-1$ colours using the inductive assumption, while the vertices not belonging to S get colour t.

By the above analysis, we have $|S| = O(nn_0^{\gamma-1})$, and by induction, the monochromatic components in colours 1 to $t-1$ have size at most $O(|S|^{1/(t-1-(t-2)\gamma)}) = O(n^{1/(t-(t-1)\gamma)})$. The components in colour t have size at most n_0 by construction. This finishes the proof of Proposition 2.1. □

https://doi.org/10.1017/S0963548308009140 Published online by Cambridge University Press

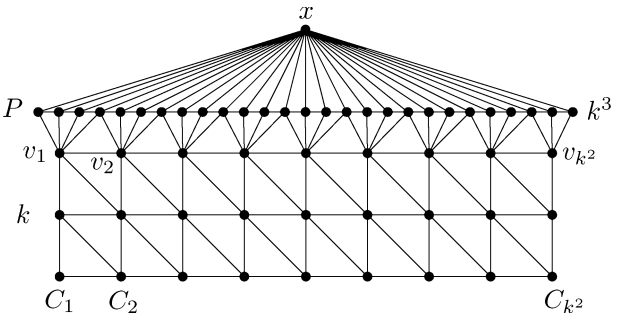


Figure 2. The lower bound construction for planar graphs, drawn for $k = 3$.

3. Lower bounds for planar graphs and for excluded minors

To prove that the bound in Theorem 1.1 for $mcc_2(G)$ is tight for planar graphs G, we construct planar graphs G with $mcc_2(G) = \Omega(n^{2/3})$. For every integer k we construct $G = G_k$ on $n = 2k^3 + 1$ vertices, as indicated in Figure 2. This G is constructed from a k by k^2 grid Z, a path P of k^3 vertices, and an extra vertex x. For $1 \leqslant i \leqslant k^2$ we denote the ith column of Z by C_i, and we let v_i be the top vertex of C_i. We break the path P into consecutive intervals $I_1, \ldots, I_{k^2}$ of k vertices each and we connect the vertices of I_i with v_i. We let $R_i = C_i \cup I_i$ and call this a *rib* of G. We connect x with all vertices in P. Finally, we add diagonals to all quadrilateral faces of the planar graph constructed so far, so that it becomes a *triangulated polygon*, that is, a planar graph where all faces, except possibly one, are triangles.

Our main tool is the following lemma on triangulated polygons (a very similar lemma appears in [19]). For a set S of vertices in a graph, we denote by ∂S the set of vertices that are not in S but have a neighbour in S.

Lemma 3.1. *Let G be a triangulated polygon and let $S \subseteq V(G)$ be such that $G[S]$ is connected. Suppose that two vertices $u, v \in \partial S$ are not separated by S in G. Then there is a path between u and v that is entirely included in ∂S.*

Proof. This is a simple consequence of the planar HEX lemma. Since $G[S]$ is connected and $u, v \in \partial S$, there is an u–v path P_1 with all internal vertices in S. Since S does not separate u from v, there is another u–v path P_2 that avoids S. We consider a subgraph of H consisting of the cycle $P_1 \cup P_2$ plus the part of G that triangulates the interior of this cycle (assuming that the single non-triangular face of G is the outer face). We add two new vertices z and t and we connect z to all vertices of P_1 and t to all vertices of P_2. The resulting graph H is a triangulation of the cycle $uzvt$.

We colour blue all vertices in ∂S including u and v, and we colour red all other vertices of H including z and t. By the HEX lemma (as stated, *e.g.*, in [18]) we have either a blue u–v path, which is what we want, or a red z–t path Q. We want to exclude the latter possibility. Let us imagine that we follow the red path Q from t to z and we watch the distance to S in H. Since t is adjacent only to P_2, whose vertices are not in S, initially at t this distance is at least 2. On the other hand, since the red vertices connected to z are inner vertices of P_1 and thus in S, the

https://doi.org/10.1017/S0963548308009140 Published online by Cambridge University Press

penultimate vertex in Q is in S. Consequently, there is a vertex in Q at distance 1 from S, but such a vertex was coloured blue: a contradiction. □

We need the following consequence of Lemma 3.1.

Corollary 3.2. *Consider a red–blue vertex colouring of the graph $G = G_k$, where x is red. Let S be the connected component of x in the red subgraph. If there is a connected component of $G \setminus S$ containing at least r complete ribs, then G has a blue connected subgraph with at least rk vertices.*

Proof. If a rib R_i is S-free (that is, contains no vertex in S), then the k vertices in the interval I_i are contained in ∂S and are therefore blue. For any two S-free ribs R_i, R_j contained in the same connected component of $G \setminus S$, we choose vertices $u \in I_i$ and $v \in I_j$ (arbitrarily). The previous lemma now shows that u and v are connected by a blue path. □

We can now show that any two-colouring of G has a monochromatic connected subgraph of at least $k^2/2$ vertices. As in the corollary, we assume that x is red and we let S be the connected component of x in the red subgraph. We may assume $|S| \leqslant k^2/2$, for otherwise we have a large red component. Hence there are at least $k^2/2$ S-free ribs. We want to show that at least $k/2$ of them are in the same connected component of $G \setminus S$; then we will be done, by the corollary.

Since $|S| < k^2$, at least one of the k *rows* of the grid Z contains fewer than k vertices from S. It follows that the S-free ribs live in at most k connected components of $G \setminus S$. So there must be at least $k/2$ of them in the same connected component, as claimed. □

For $m \geqslant 1$ and a graph G, let $cone(mG)$ be a graph constructed by taking the union of m isomorphic and pairwise disjoint copies of G (no edges between the copies) and connecting all of their vertices to an additional new vertex, called the *apex*.

Lemma 3.3. *Let $t \geqslant 1$, let G be a graph, and let $m = mcc_t(G)$. Then $mcc_{t+1}(cone(mG)) \geqslant m$.*

Proof. Let us consider a colouring of the vertices of $cone(mG)$ with $t+1$ colours, and let us assume that the apex has colour $t+1$. Clearly, all vertices of colour $t+1$ form a connected subgraph, so if there are at least m of them we have our large monochromatic connected subgraph. Otherwise, one of the copies of G lacks colour $t+1$ and the claim follows. □

Notice that as we pass from G to $cone(mG)$, the number of vertices grows (approximately) m times, but other parameters grow slowly: the tree width grows by at most one, the size of the largest clique minor grows by one, and if G is outerplanar, then $cone(mG)$ is planar.

To prove the existence statement in Theorem 1.1, we simply take the planar graph G_k constructed at the beginning of this section and we apply the above lemma $t-2$ times with $m = k^2/2$. Then mcc_t of the resulting graph is at least $k^2/2$, and the number of vertices is $O(k^{2t-1})$. This finishes the proof of Theorem 1.1. □

To prove the existence statement statement in Theorem 1.2 we need a different base graph: let F_k be the 'fan' consisting of a k-vertex path and an additional vertex adjacent to all vertices

https://doi.org/10.1017/S0963548308009140 Published online by Cambridge University Press

of this path. Clearly, F_k is an outerplanar graph of tree width 2. A straightforward computation shows that $mcc_2(F_k) = \Theta(\sqrt{k})$ (this also appears in [1]). Applying the above lemma to F_k $t-2$ times with $m = mcc_2(F_k)$, we obtain a graph of tree width at most t on $O(k^{t/2})$ vertices with $mcc_t = \Omega(\sqrt{k})$. This finishes the proof of Theorem 1.2. □

The case $t = 3$ of the construction for the proof of Theorem 1.2 yields n-vertex planar graphs with mcc_3 at least $\Omega(n^{1/3})$. These very graphs were used in [15] to show that mcc_3 is not bounded for planar graphs.

4. Edge expansion and degree $6+\varepsilon$

In this section we prove Theorem 1.3. The graphs we construct are line graphs $G = L(H)$. So the property that we need is that in every 2-colouring of $E(H)$ there are monochromatic connected components containing a positive fraction of the edges of H. To this end, it suffices to show that small subgraphs of H have small average degrees, as the next observation shows.

Lemma 4.1. *Let H be a graph with average degree $\overline{d}$. Suppose that every subgraph on p or fewer vertices in H has average degree strictly smaller than $\overline{d}/t$. Then $mcc_t(L(H)) \geqslant p$.*

Proof. If $F \subseteq E(H)$ is the largest colour class in a t-colouring of $E(H)$, then the graph $(V(H), F)$ has average degree at least $\overline{d}/t$. Hence, some connected component of $(V(H), F)$ has average degree at least $\overline{d}/t$. Such a connected component must have more than p vertices, and thus at least p edges. □

In the proof of Theorem 1.3, we will use a suitable random graph (with bounded vertex degrees) for H. The basic idea is that random graphs are typically very good *expanders*. This means that every sufficiently small set $S \subset V(H)$ has many neighbours, hence most of the edges incident to S leave S, and consequently the average degree of $H[S]$ is small. More precisely, it turns out that if $|S|$ is sufficiently small and if $H[S]$ is connected, then this subgraph is nearly a tree, in the sense that the average degree in $H[S]$ is just a little bigger than 2. Here 'sufficiently small' means that $|S| < \beta|V(H)|$, where $\beta > 0$ is a suitable small constant depending on how close we want to get to average degree 2.

A result on this almost-tree behaviour of small sets in random *regular* graphs appears explicitly in Hoory, Linial and Wigderson [10]. In particular, Theorem 4.16(1) in [10] tells us that, for every $d \geqslant 3$ and every $\delta > 0$, there exists $\beta = \beta(d, \delta) > 0$ such that almost every d-regular graph H on m vertices has average degree of $H[S]$ at most $2 + \delta$ for all S with at most βm vertices. This, together with Lemma 4.1, immediately yields the following weaker analogue of Theorem 1.3: *There exist arbitrarily large 8-regular graphs G with $mcc_2(G) = \Omega(n)$.* Indeed, we choose H as a 5-regular graph on m vertices (thus, $G = L(H)$ is 8-regular) satisfying the conclusion of the statement quoted above with $\delta = 0.4$. Then, for $\beta = \beta(5, 0.4)$, every $S \subseteq V(H)$ with at most βm vertices induces a subgraph average degree at most 2.4, and hence Lemma 4.1 with $t = 2$ and $\overline{d} = 5$ shows $mcc_2(L(H)) \geqslant \beta m$.

https://doi.org/10.1017/S0963548308009140 Published online by Cambridge University Press

In order to lower the maximum degree of G to 7 and the maximum average degree to $6+\varepsilon$, we will use a random H where a small fraction of vertices have degree 5, all others have degree 4, and no two degree-5 vertices are connected.

The random graph model. It is easier to deal with random *bipartite* graphs. Most of the literature in this area deals with regular random graphs, but we need a suitable mixture of vertex degrees, and so we prescribe the degree individually for each vertex. That is, we have two disjoint sets A and B of vertices, $|A|+|B|=m$, and for every $v \in A \cup B$ we specify a number $d(v) \in \{1,2,\dots,D\}$, where D is a constant. These degrees and the sizes of A and B are related by the condition $d(A)=d(B)$, where we use the notation $d(S)=\sum_{v\in S} d(v)$. Hence $|A|=c_A m$, $|B|=c_B m$ for constants c_A, c_B. To generate the random graph, every vertex v starts with $d(v)$ 'half-edges', and then the half-edges of all vertices in A are matched at random to the half-edges of all vertices in B (this is a *configuration model* of generating random bipartite graphs). We note that the resulting H may have multiple edges, but the line graph $L(H)$ we are interested in is still a simple graph.[†]

The following lemma speaks about number of *vertices* adjacent to S; the number of edges is then obtained as a simple consequence.

Lemma 4.2. *Let D be a fixed integer. Then, for every $\delta>0$ there exists $\beta=\beta(D,\delta)>0$ such that, if H is generated according to the above model (with an arbitrary choice of the $d(v)$s), then with probability $1-o(1)$ (as $m\to\infty$), every $S\subseteq A$ with $|S|\leqslant \beta m$ has at least $d(S)-(1+\delta)|S|$ neighbours in B.*

Proof. Let us write $w(S)=d(S)-(1+\delta)|S|$. Since every S has at least $|S|/D$ neighbours, it suffices to consider only the S with $w(S)\geqslant |S|/D$. The calculation is a variation of that in [10].

For sets $S\subseteq A$, $|S|=s\leqslant \beta m$, and $W\subseteq B$, $|W|=\lfloor w(S)\rfloor$, let $X_{S,W}$ be the event 'all neighbours of S lie in W'. If no $X_{S,W}$ occurs, then H satisfies the conclusion of the lemma. We have

$$\Pr[X_{S,W}]=\frac{d(W)(d(W)-1)\cdots(d(W)-d(S)+1)}{d(B)(d(B)-1)\cdots(d(B)-d(S)+1)}\leqslant\left(\frac{d(W)}{d(B)}\right)^{d(S)}\leqslant\left(\frac{D^2w(S)}{m}\right)^{d(S)}.$$

Thus, for S fixed, the probability of $X_{S,W}$ occurring for some w is at most

$$\binom{|B|}{w(S)}\left(\frac{D^2w(S)}{m}\right)^{d(S)}.$$

Estimating the binomial coefficient as $\binom{x}{y}\leqslant(ex/y)^y$ and using $s/D\leqslant w(S)\leqslant d(S)\leqslant Ds$, this can be bounded by

$$\left(\frac{e|B|}{w(S)}\right)^{w(S)}\left(\frac{D^2w(S)}{m}\right)^{d(S)}\leqslant\left(\frac{e|B|D}{s}\right)^{w(S)}\left(\frac{D^3s}{m}\right)^{d(S)}\leqslant C_1^s\left(\frac{s}{m}\right)^{(1+\delta)s},$$

[†] It is well known that for random regular graphs of fixed degree, the configuration model yields a simple graph with probability bounded away from 0, and consequently, any property that holds almost surely in the configuration model also holds for almost all simple regular graphs of the given degree: see, *e.g.*, [12]. By slightly modifying the proof of this fact, we could also get a similar result for our model with mixed degrees, and hence have H simple.

https://doi.org/10.1017/S0963548308009140 Published online by Cambridge University Press

where $C_1 = e^D D^{4D}$. The probability of any $X_{S,W}$ occurring at all is bounded by

$$\sum_{1 \leqslant s \leqslant \beta m} \binom{|A|}{s} C_1^s \left(\frac{s}{m}\right)^{(1+\delta)s} \leqslant \sum_s C_2^s \left(\frac{s}{m}\right)^{\delta s},$$

where $C_2 = eC_1$. The term for $s = 1$ is $O(m^{-\delta}) = o(1)$, and the ratio of consecutive terms is at most $C_2\beta^\delta$, which can be made smaller than $\frac{1}{2}$, say, by fixing β small enough. Then the entire sum is $o(1)$, as claimed. □

Corollary 4.3. *In the setting of Lemma 4.2, the following holds almost surely: the average degree of the subgraph of H induced by any set of at most βm vertices is at most $2 + 2\delta$.*

Proof. Let us consider the subgraph of H induced by $S \cup T$, $S \subseteq A$, $T \subseteq B$, $|S \cup T| \leqslant \beta m$. By Lemma 4.2, we may assume that S has at least $w(S)$ neighbours. Hence at least $w(S) - |T|$ neighbours of S do not lie in T, and each such neighbour 'consumes' at least one edge among the $d(S)$ edges incident to S. Thus the number of edges in $H[S \cup T]$ is at most $d(S) - w(S) + |T| = (1 + \delta)|S| + |T|$, and the average degree is at most $2 + 2\delta$ (actually, at most $2 + \delta$, if we use symmetry and assume $|S| \geqslant |T|$). □

Proof of Theorem 1.3. We let $\rho = \rho(\varepsilon) > 0$ be a sufficiently small constant, and we set the parameters of our random graph model as follows: we let $d(v) = 5$ for some ρm vertices in A, and all remaining vertices in $A \cup B$ have $d(v) = 4$. Clearly, the average degree of H is $4 + \Omega(\rho)$, and thus if we choose δ sufficiently small in terms of ρ and let $\beta = \beta(5, \delta)$, then Lemma 4.1 and Corollary 4.3 guarantee $mcc_2(L(H)) \geqslant \beta m$ almost surely.

The maximum degree of $L(H)$ is 7; the degree-7 vertices of the line graph correspond to $5\rho m$ edges of H incident to the ρm vertices of degree 5. It remains to bound the maximum average degree of $L(H)$.

To this end, we apply Corollary 4.3 once again, this time with $\delta = \frac{1}{2}$, say, and we let $\beta_0 = \beta(5, \frac{1}{2})$ be the corresponding parameter. We note that β_0 is independent of ρ, and hence we can assume that ρ/β_0 is sufficiently small.

Now let $F \subseteq E(H)$ be an arbitrary subset of edges, and let U be the set of all vertices incident to edges of F. If $|U| \leqslant \beta_0 m$, then by Corollary 4.3 the graph $K := (U, F)$ has average degree at most 2.5. The average degree of $L(K)$ is

$$\bar{d}(L(K)) = \frac{1}{|F|} \sum_{\{u,v\} \in F} (\deg_K(u) + \deg_K(v) - 2) = \frac{1}{|F|} \sum_{u \in U} \deg_K(u)(\deg_K(u) - 1).$$

If we denote by x_i the fraction of vertices $u \in U$ with $\deg_K(u) = i$, $i = 1, 2, \ldots, 5$, then the x_i satisfy the constraints $\sum_{i=1}^5 x_i = 1$ and $\sum_{i=1}^5 i x_i \leqslant 2.5$ (this reflects the bound on the average degree of K), and we have $\bar{d}(L(K)) = 2(2x_2 + 6x_3 + 12x_4 + 20x_5)/(x_1 + 2x_2 + 3x_3 + 4x_4 + 5x_5)$. One can use, *e.g.*, linear programming to verify that the above constraints imply $\bar{d}(L(K)) \leqslant 6$ as needed.

If, on the other hand, $|U| > \beta_0 m$, then since H has at most ρm vertices of degree 5, these vertices constitute at most ρ/β_0 fraction of U, and all other vertices have degree at most 4. Hence at most a small fraction of the vertices of $L(K)$ can have degree 7, while others have

https://doi.org/10.1017/S0963548308009140 Published online by Cambridge University Press

degrees at most 6, and it follows that the average degree of $L(K)$ can be pushed below $6+\varepsilon$ by making ρ sufficiently small. Theorem 1.3 is proved. □

5. The Hamming cube

An interesting example, where mcc_2 can be determined exactly, is the line graph of the Hamming cube. Let Q_d denote the d-dimensional Hamming cube with vertex set $\{0,1\}^d$ and with two vectors $u, v \in \{0,1\}^d$ adjacent in Q_d if they differ in exactly one coordinate. Let $L(Q_d)$ be the line graph of Q_d.

Proposition 5.1. *For every even d we have $mcc_2(L(Q_d)) = \frac{d}{4}2^{d/2}$.* □

Proof. The upper bound is witnessed by the following colouring. Colour an edge $\{u,v\} \in E(Q_d)$ red if u and v differ in one of the first $d/2$ coordinates, and blue otherwise. Then the monochromatic components are $(d/2)$-dimensional subcubes.

For the lower bound, it suffices to show that whenever the edges of Q_d are coloured red and blue, there exists a monochromatic connected subgraph with at least $\frac{d}{4}2^{d/2}$ edges. Let us assume that, *e.g.*, blue is the majority colour; that is, at least $\frac{d}{4}2^d$ edges are blue. Let $B_1, B_2, \ldots, B_k$ be the connected components of the blue subgraph, and let m_i be the number of blue edges in B_i. Hence $\sum_{i=1}^k m_i \geqslant \frac{1}{2}|E(Q_d)| = \frac{d}{4}2^d$.

We recall the following formulation of the edge-isoperimetric inequality for the cube (*e.g.*, [5, Chapter 16]). Every subgraph of Q_d on v vertices has at most $\frac{1}{2}v\log_2 v$ edges. Let $\beta_i > 0$ be the real number satisfying $\frac{1}{2}\beta_i \log_2 \beta_i = m_i$ (where $\beta_i = 1$ for $m_i = 0$). Thus, β_i is a lower bound for the number of vertices of B_i, and consequently, $\sum_{i=1}^k \beta_i \leqslant 2^d$. Assuming for contradiction that $m_i < \frac{d}{4}2^{d/2}$ for all i, we have $\beta_i < 2^{d/2}$ for all i, and thus

$$\sum_{i=1}^k m_i = \sum_{i=1}^k \frac{1}{2}\beta_i \log_2 \beta_i < \frac{d}{4}\sum_{i=1}^k \beta_i \leqslant \frac{d}{4}\sum_{i=1}^k |V(B_i)| \leqslant \frac{d}{4}2^d.$$

But, as was noted above, $\sum_{i=1}^k m_i \geqslant \frac{d}{4}2^d$, and this contradiction establishes the proposition. □

The proof also shows that the monochromatic components in any extremal colouring have to be $(d/2)$-dimensional subcubes.

6. Open problems

There are quite a few interesting open questions suggested by the present paper. Here are some of them.

(1) How large can $mcc_2(G)$ be for graphs of maximum degree 6? By the planar HEX lemma, the triangulated planar grid is an example with $mcc_2(G) = \Theta(\sqrt{n})$, but this is at present the best we know.

(2) A special case of the previous question, which seems interesting in its own right, is when $G = L(H)$ for some 4-regular H. The best lower bound we know is $\Omega(\log n)$, from a construction by Alon, Ding, Oporowski and Vertigan [1], where H is a 4-regular graph of logarithmic girth.

https://doi.org/10.1017/S0963548308009140 Published online by Cambridge University Press

(3) The examples we know for planar graphs G with large $mcc_2(G)$ and $mcc_3(G)$ have at least one vertex of high degree. Can anything better be said if we assume that G has bounded degrees? More specifically, the following was asked in [15]: Is there a function f such that, for every planar graph G of maximum degree Δ, we have $mcc_3(G) \leqslant f(\Delta)$?

(4) For two colours we cannot hope for constant monochromatic component size in bounded-degree planar graphs, as shown by a triangulated planar grid. However, as in the previous question, we can ask if there exists a function g such that every n-vertex planar graph G of maximum degree Δ satisfies $mcc_2(G) \leqslant g(\Delta)\sqrt{n}$.

(5) There is still a gap between the best bounds we know for $mcc_t(G)$ for graphs from minor-closed families of graphs. Can this gap be closed?

(6) A question suggested to us by Emo Welzl concerns the possible behaviour of $mcc_t(G)$ when the chromatic number of G as well as its number of vertices are known.

(7) The proof of Theorem 1.3 can be adapted to show that for any fixed t there exist n-vertex graphs with maximum degree $4t-1$ and with $mcc_t(G) = \Omega(n)$. Note that in [1] it was shown that $mcc_t(G)$ is not bounded by a constant even for graphs of maximum degree $4t-2$; however, that proof gives only a logarithmic lower bound for $mcc_t(G)$ (*cf.* our first open problem). From the other direction, Haxell, Szabó and Tardos [9] show that $mcc_t(G)$ is bounded by a constant for all t and all graphs G with maximum degree at most $3t-1$. The constant 3 here is not optimal, since the same paper shows that for some constant $\varepsilon > 0$ and all sufficiently large t, the value $mcc_t(G)$ is bounded by a constant for all graphs G of maximum degree at most $(3+\varepsilon)t$. It would be interesting to find the asymptotic behaviour of the maximal value of $mcc_t(G)$ for graphs G with maximum degree d in the intermediate range $(3+\epsilon)t < d < 4t-2$. In particular, it would be interesting to know if there exist t and d for which the above maximum is sublinear but not a constant.

(8) There are several natural conjectures pertaining to mcc_t for triangulations of the d-dimensional grid graph. These questions suggest an interesting 'combinatorial dimension theory' waiting to be discovered. More on this subject can be found in [19].

Acknowledgements

We would like to thank Maria Chudnovsky and Tibor Szabó for fruitful conversations and to Eyal Ackerman and Robin Thomas for important references.

References

[1] Alon, N., Ding, G., Oporowski, B. and Vertigan, D. (2003) Partitioning into graphs with only small components. *J. Combin. Theory Ser. B* **87** 231–243.

[2] Alon, N., Seymour, P. and Thomas, R. (1990) A separator theorem for nonplanar graphs. *J. Amer. Math. Soc.* **3,4** 801–808.

[3] Berke, R. (2008) Coloring and transversals of graphs. PhD thesis, ETH Zurich.

[4] Berke, R. and Szabó, T. (2005) Relaxed two-coloring of cubic graphs. In *DMTCS Proc. 2005 European Conference on Combinatorics, Graph Theory and Applications* (EUROCOMB), pp. 341–344.

[5] Bollobás, B. (1986) *Set Systems, Hypergraphs, Families of Vectors and Combinatorial Probability*, Cambridge University Press.

[6] Bollobás, B. and Leader, I. (1991) Edge-isoperimetric inequalities in the grid. *Combinatorica* **11** 299–314.

https://doi.org/10.1017/S0963548308009140 Published online by Cambridge University Press

[7] Ding, G., Oporowski, B., Sanders, D. and Vertigan, D. (1996) Preprint, Louisiana State University, Baton Rouge.
[8] Gale, D. (1979) The game of hex and the Brouwer fixed-point theorem. *Amer. Math. Monthly* **86** 818–827.
[9] Haxell, P., Szabó, T. and Tardos, G. (2003) Bounded size components: Partitions and transversals. *J. Combin. Theory Ser. B* **88** 281–297.
[10] Hoory, S., Linial, N. and Wigderson, A. (2006) Expander graphs and their applications. *Bull. Amer. Math. Soc.* **43,4** 439–561.
[11] Jackson, B. and Wormald, N. (1996) On the linear k-arboricity of cubic graphs. *Discrete Math.* **162** 293–297.
[12] Janson, S., Łuczak, T. and Ruciński, A. (2000) *Topics in Random Graphs*, Wiley.
[13] Jensen, T. R. and Toft, B. (1995) *Graph Coloring Problems*, Wiley–Interscience Series in Discrete Mathematics and Optimization.
[14] Kawarabayashi, K. and Mohar, B. (2007) A relaxed Hadwiger's conjecture for list colorings. *J. Combin. Theory Ser. B* **97** 647–651.
[15] Kleinberg, J., Motwani, R., Raghavan, P. and Venkatasubramanian, S. (1997) Storage management for evolving databases. In *Proc. 38th IEEE Symposium on Foundations of Computer Science* (FOCS), pp. 353–362.
[16] Linial, N. and Saks, M. (1993) Low diameter graph decompositions. *Combinatorica* **13** 441–454.
[17] Lipton, R. and Tarjan, R. (1979) A separator theorem for planar graphs. *SIAM J. Appl. Math.* **2** 177–190.
[18] Matoušek, J. and Nešetřil, J. (1998) *Invitation to Discrete Mathematics*, Oxford University Press.
[19] Matoušek, J. and Přívětivý, A. Large monochromatic components in two-colored grids. *SIAM J. Discrete Math.*, in press.
[20] Reed, B. (1992) Finding approximate separators and computing tree width quickly. In *Proc. 24th Annual ACM Symposium on the Theory of Computing* (STOC), pp. 221–228.
[21] Robertson, N. and Seymour, P. (1986) Graph minors II: Algorithmic aspects of tree-width. *J. Algorithms* **7** 309–322.
[22] Robertson, N. and Seymour, P. (1991) Graph minors X: Obstructions to tree-decomposition. *J. Combin. Theory Ser. B* **52** 153–190.
[23] Robertson, N. and Seymour, P. (1995) Graph minors XIII: The disjoint path problem. *J. Combin. Theory Ser. B* **63** 65–110.
[24] Thomassen, C. (1999) Two-colouring the edges of a cubic graph such that each monochromatic component is a path of length at most 5. *J. Combin. Theory Ser. B* **75** 100–109.

https://doi.org/10.1017/S0963548308009140 Published online by Cambridge University Press