Some applications of matching theorems

Vaughan, Emil Richard

Some Applications of Matching Theorems

Emil Richard Vaughan

School of Mathematical Sciences,
Queen Mary,
University of London

Submitted for the degree of Doctor of Philosophy.

February 10, 2010

Abstract

This thesis contains the results of two investigations. The first concerns the 1-factorizability of regular graphs of high degree. Chetwynd and Hilton proved in 1989 that all regular graphs of order $2n$ and degree $2n\lambda$ where

$$\lambda > \tfrac{1}{2}(\sqrt{7}-1) \approx 0.82288$$

are 1-factorizable. We show that all regular graphs of order $2n$ and degree $2n\lambda$ where λ is greater than the second largest root of

$$4x^6 - 28x^5 - 71x^4 + 54x^3 + 88x^2 - 62x + 3$$

(≈ 0.81112) are 1-factorizable. It is hoped that in the future our techniques will yield further improvements to this bound. In addition our study of barriers in graphs of high minimum degree may have independent applications.

The second investigation concerns partial latin squares that satisfy Hall's Condition. The problem of completing a partial latin square can be viewed as a list-colouring problem in a natural way. Hall's Condition is a necessary condition for such a problem to have a solution. We show that for certain classes of partial latin square, Hall's Condition is both necessary and sufficient, generalizing theorems of Hilton and Johnson, and Bobga and Johnson. It is well-known that the problem of deciding whether a partial latin square is completable is NP-complete. We show that the problem of deciding whether a partial latin square that is promised to satisfy Hall's Condition is completable is NP-hard.

Contents

Acknowledgements

I would like to thank my supervisor Peter J. Cameron for his support and for being a constant source of encouragement and good ideas, and for introducing me to many interesting questions.

I would also like to thank Anthony J. W. Hilton for his support and for his pioneering work in some of the areas of graph theory in which I have worked.

Chapter 1

Introduction

This thesis contains the results of two investigations. The first concerns the 1-factorizability of regular graphs of high degree and is presented in Chapter 3. This investigation continues the recent work of Cariolaro and Hilton. The second investigation concerns partial latin squares that satisfy Hall's Condition and the results are presented in Chapter 4.

Both chapters need some results from matching theory and the theory of edge-colourings, and this material has been placed in Chapter 2. This chapter also contains some more specialized results from the investigation into 1-factorizability. However, these results are somewhat independent of those in Chapter 3 and may be of more general applicability. For these reasons they are given in Chapter 2.

Throughout, some familiarity with the basic definitions of graph theory is assumed. The reader is referred to the books of Bondy and Murty [4] and Diestel [14] for explanations of any undefined terms. However, we will try to provide definitions of all but the most basic terms from graph theory.

We will give a few definitions right away. A *matching* in a graph is a set of pairwise non-adjacent edges. Vertices incident with edges of a matching are said to be *covered* by the matching. Similarly, a set of vertices is covered by a matching if all its members are covered. A *1-factor* in a graph is a matching that covers every vertex. A *1-factorization* is a decomposition of the edge set into disjoint 1-factors. Note that a matching always covers an even number of vertices, so graphs of odd order cannot have 1-factors.

1.1 The 1-factorizability of regular graphs of high degree

The following is a a long-standing conjecture, due to Chetwynd and Hilton [8].

Conjecture 1.1. *Any regular graph of order $2n$ and degree at least n is 1-factorizable.*

There is plenty of evidence of the truth of this conjecture, not least the following theorem that shows that it holds "asymptotically". This result was obtained by Häggkvist in the 1980s but the proof was never published. It was proved independently some time later by Perković and Reed.

Theorem 1.2. (Häggkvist [25]; Perković and Reed, 1997 [33]) *For any $\epsilon > 0$ there is an integer $N = N(\epsilon)$ such that for all $n \geq N$ any regular graph on $2n$ vertices and degree at least $(1+\epsilon)n$ is 1-factorizable.*

Various authors have looked at verifying Conjecture 1.1 for graphs of order $2n$ and degree $2n-k$ for some fixed k. It is a folklore result that if G is complete, then G is 1-factorizable (see Lemma 2.3), so Conjecture 1.1 is true when $k = 1$. It is also true when $k = 2$ or 3, and again this result is folklore (see Lemma 3.2). The case $k = 4$ was settled partially by Rosa and Wallis [36] and then completely by Chetwynd and Hilton [8]. In the same paper, Chetwynd and Hilton also dealt with the case $k = 5$. The case $k = 6$ was settled by Song and Yap [39].

However, Chetwynd and Hilton in 1985 gave the first proof that Conjecture 1.1 holds if the degree is sufficiently large as a fraction of the order.

Theorem 1.3. (Chetwynd and Hilton [8]) *Let G be a regular graph of order $2n$ and degree $2n\lambda$ where $\lambda \geq \frac{6}{7}$. Then G is 1-factorizable.*

In subsequent work they obtained the following improvement. (Coincidentally, the same bound was obtained independently by Niessen and Volkmann.)

Theorem 1.4. (Chetwynd and Hilton [10]; Niessen and Volkmann [32]) *Let G be a regular graph of order $2n$ and degree $2n\lambda$ where*

$$\lambda > \tfrac{1}{2}(\sqrt{7} - 1) \approx 0.82288.$$

Then G is 1-factorizable.

We will give an improved proof of this theorem. Recently, Cariolaro and Hilton looked at whether it could be improved further. They showed that if G is a regular graph of order $2n$ and degree $2n\lambda$ where

$$\lambda > \tfrac{1}{6}(\sqrt{57} - 3) \approx 0.75831$$

then G is either 1-factorizable or one of two special cases holds [5, 7]. Furthermore, they showed that if λ is greater than the second largest root of

$$\lambda^4 - \lambda^3 - 4\lambda^2 + 2\lambda + 1$$

(≈ 0.78526) then the first of these cases cannot occur. However, they did not rule out the possibility that the second case could occur.

We will show that for sufficiently large λ the second case cannot occur. As a consequence we obtain an improvement to Theorem 1.4.

Theorem 1.5. *Let G be a regular graph of order $2n$ and degree $2n\lambda$ where λ is greater than the second largest root of*

$$4x^6 - 28x^5 - 71x^4 + 54x^3 + 88x^2 - 62x + 3$$

(≈ 0.81112). *Then G is 1-factorizable.*

1	2	3	4	5	6
3	6	1	2	4	5
5	4	2	6	3	1
2	5				
4	1				
6	3				

Figure 1.1: Goldwasser's square

1.2 Partial latin squares and Hall's Condition

A *partial latin square* of order n is an $n \times n$ array where each cell is either empty or contains a symbol from $\{1, \ldots, n\}$, with the condition that no symbol occurs more than once in any row or column. A *latin square* is a partial latin square with no empty cells. If it is possible to turn a partial latin square into a latin square by filling the empty cells with entries from $\{1, \ldots, n\}$ then we say that the partial latin square is *completable*.

We say that a set of cells of a partial latin square is *independent* if no two of the cells are in the same row or column. We say that a cell *supports* a symbol σ if it either contains σ, or is empty and σ does not occur in the same row or column. (In other words, the cell either contains σ, or *could* contain it.)

Let T be a set of cells of a partial latin square P of order n. Let σ be a symbol from $\{1, \ldots, n\}$. Then $\alpha(\sigma, T)$ is defined to be the maximum size of an independent subset of the set of cells in T that support σ.

We say that a partial latin square P satisfies *Hall's Condition* if for all sets of cells T we have

$$\sum_{\sigma \in \{1,\ldots,n\}} \alpha(\sigma, T) \geq |T| \,. \tag{1.1}$$

It is not hard to show that Hall's Condition is a necessary condition for a partial latin square to be completable (see Lemma 4.2). Hilton and Johnson [27] observed that in the case of partial latin squares where the filled cells form a rectangle, Hall's condition is both a necessary and a sufficient condition for completability:

Theorem 1.6. *Let P be a partial latin square of order n whose filled cells are those in the upper left $r \times s$ rectangle, for some $r, s \in \{1, \ldots, n\}$. P is completable if and only if P satisfies Hall's Condition.*

Cropper [12] asked if this was true more generally: is Hall's Condition a sufficient condition for the completability of partial latin squares? John Goldwasser showed that it is not, by exhibiting a partial latin square that satisfies Hall's Condition but is not completable [24]. The square is shown in Figure 1.1.

Bobga and Johnson found that Hall's Condition is also a sufficient condition for completability in the case that the filled region is a rectangle with one empty cell inside [3]. We present the following generalization of this result:

Theorem 1.7. *Let P be a partial latin square of order n whose filled cells are those in the upper left $r \times s$ rectangle, for some $r, s \in \{1, \ldots, n\}$, except for t cells in this region that are empty, with the condition that there is at most one of these empty cells in each column. P is completable if and only if P satisfies Hall's Condition.*

An obvious computational problem to study is:

Problem 1.8. Let P be a partial latin square. Decide if P satisfies Hall's Condition.

Unfortunately we are not able to say much about the complexity of this problem. We conjecture that it can be solved in polynomial time, but at present we cannot even show that it is in NP. It is however possible to check each Hall's Inequality in polynomial time (see Lemma 4.7), and so if we have a partial latin square that does *not* satisfy Hall's Condition, we can verify this fact in polynomial time. For this reason, Problem 1.8 is in co-NP.

The following problem was shown to be NP-complete by Colbourn [11] (see also [17]):

Problem 1.9. Let P be a partial latin square. Decide if P is completable.

We consider the following variant of Problem 1.9:

Problem 1.10. Let P be a partial latin square that satisfies Hall's Condition. Decide if P is completable.

The set of partial latin squares that are "yes" instances of Problem 1.9 is the same as the set of partial latin squares that are "yes" instances of Problem 1.10. The difference is that in Problem 1.10, the input is restricted to partial latin squares that satisfy Hall's Condition. Thus Problem 1.10 is an example of a *promise problem*, as the input partial latin square is "promised" to satisfy Hall's Condition. (See [23] for a survey of promise problems.) We shall prove the following.

Theorem 1.11. *Problem 1.10 is NP-hard.*

Because of Theorem 1.11, knowing that a partial latin square satisfies Hall's Condition probably does not help one determine if it is completable or not. In Section 4.6 we shall give a reduction from an NP-complete hypergraph colouring problem to Problem 1.9, with the property that it maps to partial latin squares that satisfy Hall's Condition. This proves Problem 1.10 is NP-hard, as well as providing a new proof that Problem 1.9 is NP-complete.

Chapter 2

Matchings in graphs

In this chapter we provide some results that will be needed later. Section 2.1 covers some background material on edge-colourings. Section 2.2 covers matchings in bipartite graphs and Section 2.3 deals with the Gallai-Edmonds Theorem which describes the structure of maximum matchings in general graphs. The last two Sections, 2.4 and 2.5, deal with consequences of the Gallai-Edmonds Theorem for graphs of high minimum degree. This material is specifically included for Chapter 3, although it is somewhat independent and may be of wider applicability.

2.1 Edge-colourings

Let G be a graph and $\mathcal{C}$ a set of colours. An *edge-colouring* of G is a map from $E(G) \to \mathcal{C}$ such that adjacent edges receive different colours. The *chromatic index* $\chi'(G)$ is the least number of colours needed for an edge-colouring of G.

As adjacent edges must be assigned different colours, the set of edges in a graph coloured by a particular colour is a matching. So an edge-colouring can be viewed as a partition of $E(G)$ into a set of disjoint matchings. Note that in a graph with an edge-colouring, we say that a vertex *misses* a colour $c \in \mathcal{C}$ if it is incident with no edges of that colour.

We will use $\Delta(G)$ to denote the maximum degree of G, and later on we will

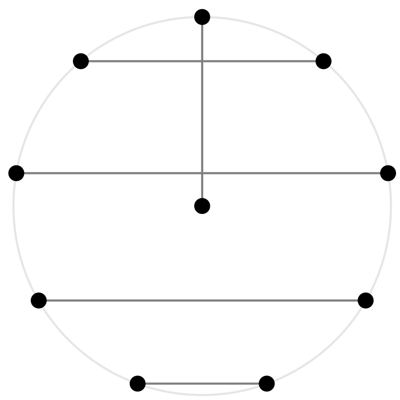


Figure 2.1: A 1-factorization of K_n.

use $\delta(G)$ to denote the minimum degree. One of the most well-known theorems in edge-colouring is the following:

Theorem 2.1. (Vizing, 1965 [4, 14]) *Let G be a graph. The chromatic index $\chi'(G)$ is either $\Delta(G)$ or $\Delta(G)+1$.* □

If $\chi'(G) = \Delta(G)$ then G is said to be of *class 1*, otherwise it is of *class 2*. A graph G that is regular and class 1 is 1-factorizable, as an edge-colouring with $\Delta(G)$ colours partitions $E(G)$ into disjoint 1-factors. And as a 1-factorization is a colouring with $\Delta(G)$ colours, we have:

Remark 2.2. *A graph G is 1-factorizable if and only if it is regular and class 1.*

Note that a regular graph of odd order cannot be class 1, as in any graph of odd order there must be at least one vertex missing from any matching, so more than $\Delta(G)$ colours are needed for an edge-colouring.

Lemma 2.3. *The complete graph K_n is class 1 if and only if n is even.*

Proof. By what has just been observed, K_n must be class 2 if n is odd. If n is even, we can draw the vertices of K_n in a circle with one vertex in the centre, as shown in Figure 2.1. In the diagram, one 1-factor is shown—rotating it gives $n-1$ disjoint 1-factors, showing that K_n is 1-factorizable, and so class 1, when n is even. □

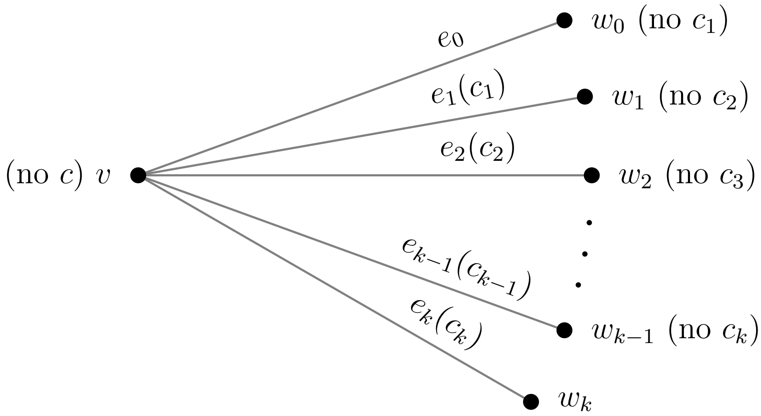


Figure 2.2: Pivot and fan.

Let G be a graph with an edge-colouring using the set of colours $\mathcal{C}$. Suppose $c_1, c_2 \in \mathcal{C}$ are two colours. Then a c_1c_2-*alternating path* is a path in G whose edges are coloured alternately c_1 and c_2. To *exchange* a c_1c_2-alternating path is to recolour the edges coloured c_1 with c_2 and the edges coloured c_2 with c_1.

The following lemma was used by Vizing in the original proof of his theorem [4].

Lemma 2.4. *Let G be a graph and let $\mathcal{C}$ be a set of colours. Let $e = vw$ be an edge of G and suppose that there is an edge-colouring of $G - e$ using the colours from $\mathcal{C}$. If there is a colour missing at v, and if each of the neighbours of v has a colour missing, then there is an edge-colouring of G using the colours from $\mathcal{C}$.*

Proof. Suppose v is missing the colour $c \in \mathcal{C}$. We can construct a sequence of edges, called a *fan*, using the following procedure. The vertex v is said to be the *pivot* of the fan.

(1) Let $i = 0$, and $w_0 = w$.

(2) Stop if w_i is missing c.

(3) If there is a colour in $\mathcal{C}$ that is missing from w_i and is distinct from $\{c_1, \ldots, c_i\}$ then let c_{i+1} be this colour. If not, stop.

(4) If there is an edge of colour c_{i+1} incident with v, let e_{i+1} be this edge, and let w_{i+1} be its other end. If not, stop.

(5) Increment i. Goto (2).

It is clear that this procedure will always terminate, as the number of coloured edges incident with v is finite. Suppose it creates a fan with edges $e_0, \ldots, e_k$. (See Figure 2.2.) Upon termination, one of the following holds:

(i) Colour c is missing from w_k.

(ii) There is no colour c_{k+1} distinct from $\{c_1, \ldots, c_k\}$ that is missing at w_k.

(iii) There is no edge of colour c_{k+1} incident with v.

We will show how we can recolour the edges of the fan, to allow us to colour the edge e.

- *If case (i) occurs.* We recolour $e_k, e_{k-1}, \ldots, e_1$ with $c, c_k, \ldots, c_2$ and colour e_0 with c_1. This gives an edge-colouring of G using the colours from $\mathcal{C}$. (This recolouring procedure is called "rotating" the fan.)
- *If case (iii) occurs.* In this case, c_{k+1} is also missing at v, so we recolour e_k, $e_{k-1}, \ldots,$ e_1 with $c_{k+1}, c_k, \ldots, c_2$ and colour e_0 with c_1. This gives an edge-colouring of G using the colours from $\mathcal{C}$.
- *If case (ii) occurs.* Let c_{k+1} be a colour missing at w_k. So there is an $i \in \{1, \ldots, k-2\}$ such that $c_i = c_{k+1}$. Let P be the maximal alternating cc_i path starting at w_k. We need to consider four cases:

  (A) P ends at v.

  (B) P ends at w_0.

  (C) P ends at one of $w_1, \ldots, w_{k-1}$.

  (D) P ends somewhere else.

- *If case (D) occurs.* In this case we can exchange P and recolour $e_k, e_{k-1}, \ldots, e_1$ with $c, c_k, \ldots, c_2$ and colour e_0 with c_1. This gives an edge-colouring of G using the colours from $\mathcal{C}$.
- *If case (A) occurs.* As v is missing c, the last edge of P must be coloured c_i, which means it must be e_i. Again we exchange P. P includes e_i which is recoloured with c. We then recolour $e_{i-1}, \ldots, e_1$ with $c_i, \ldots, c_2$ and colour e_0 with c_1. This gives an edge-colouring of G using the colours from $\mathcal{C}$.
- *If case (B) occurs.* This is the simplest case. We exchange on P and colour e_0 with c. This gives an edge-colouring of G using the colours from $\mathcal{C}$.
- *If case (C) occurs.* Note that P cannot end at w_i, as w_i has edges of both colour c_i and colour c incident with it. The last edge of P must be coloured c because all the vertices $w_0, \ldots, w_k$ have an edge of colour c incident with them. Suppose P ends at w_j. Again we exchange P. Then we recolour $e_j, \ldots, e_1$ with $c, c_j, \ldots, c_2$ and colour e_0 with c_1. This gives an edge-colouring of G using the colours from $\mathcal{C}$.

$\square$

Proof of Theorem 2.1. It is immediate that for any graph G we must have $\chi'(G) \geq \Delta(G)$, so it will suffice to prove that any graph G has an edge colouring with $\Delta(G) + 1$ colours. Let $e = vw$ be an edge of G. By induction on the number of edges, we may assume that $G - e$ can be coloured with $\Delta(G - e) + 1$ colours, so it can certainly be coloured with $\Delta(G) + 1$ colours. As we have at most $\Delta(G)$ edges incident with any vertex, and we have $\Delta(G) + 1$ colours, there must surely be a colour missing at every vertex. So by Lemma 2.4 we can colour G with $\Delta(G) + 1$ colours. $\square$

The following two results are consequences of Lemma 2.4.

Lemma 2.5. *Let G be a graph and $v \in V(G)$. Suppose $\Delta(G-v) = \Delta(G)$ and that v is adjacent to at most one vertex of maximum degree. Then $\chi'(G) = \chi'(G-v)$.*

Proof. Since $\chi'(G) \geq \chi'(G-v)$ is immediate, it will suffice to show that $\chi'(G) \leq \chi'(G-v)$. Suppose we have an edge-colouring of $G-v$ with $\chi'(G-v)$ colours. We can extend this edge-colouring to G, by colouring the edges incident with v one at a time. If there is an edge joining v to a vertex of maximum degree we leave this edge until last. In this way, at all times there is at least one colour missing from v and each of its neighbours, because each neighbour either has degree less than $\Delta(G)$ or is incident with at least one uncoloured edge. So by Lemma 2.4 each edge can be coloured, using v as the pivot. □

Let G be a graph. The *core* of G, denoted G_Δ, is the subgraph of G induced by the vertices of degree $\Delta(G)$. Fournier gave the following sufficient condition for a graph to be class 1. Note that a *forest* is a graph where each connected component is a tree.

Theorem 2.6. (Fournier, 1973 [20]) *If G_Δ is a forest, then G is class 1.*

Proof. We will show that G can be coloured with $\Delta(G)$ colours. First we colour all the edges that are incident with at most one vertex of G_Δ. In the case that an edge is incident with a vertex $v \in G_\Delta$, we use v as the pivot. Lemma 2.4 says that we can always colour these edges, as any vertex joined to the pivot by a coloured edge will be a vertex of degree less than $\Delta(G)$, and will therefore have at least one colour missing. (The pivot will naturally have a colour missing, as it is incident with an uncoloured edge.)

It remains to colour the edges that are incident with two vertices of maximum degree. Let T be a connected component of G_Δ; by assumption T is a tree. Let v_0 be an arbitrary vertex in T, and let k be the maximum distance of a vertex in T from v_0. We can colour the edges of T in k stages. In stage 1 we colour all the edges of T incident with v_0, using the neighbours of v_0 as pivots. In stage i ($1 < i \leq k$) we colour all the edges that join vertices at distance $i-1$ from v_0 to vertices at distance i from v_0, always choosing the vertex furthest from v_0 as the pivot. By colouring in this order, we ensure that the only edges joining our pivots to other vertices in G_Δ are uncoloured edges, and hence these neighbouring vertices each have at least one colour missing. So by Lemma 2.4 all these edges can be coloured. We repeat this process for each component of G_Δ. $\square$

We will need the following well-known theorem of Dirac, that provides a sufficient condition for a graph to have a Hamilton cycle.

Theorem 2.7. (Dirac, 1952 [15, 14]) *Let G be a graph of order $n \geq 3$. If $\delta(G) \geq \frac{1}{2}n$ then G has a Hamilton cycle.*

Sadly, we only use Theorem 2.7 to obtain the following corollary.

Corollary 2.8. *Let G be a graph of order n where n is even. If $\delta(G) \geq \frac{1}{2}n$ then G has a 1-factor.*

Proof. If $n = 2$ then the result clearly holds. Otherwise $n \geq 4$ and by Theorem 2.7 G has a Hamilton cycle. A 1-factor can be obtained by taking every other edge of this Hamilton cycle. $\square$

We conclude this section with the following easy lemma.

Lemma 2.9. *Let G be a graph and F a 1-factor of G. Then G is class 1 if $G-F$ is class 1.*

Proof. If $G-F$ is class 1 then it has an edge-colouring with $\Delta(G-F) = \Delta(G)-1$ colours. So we can colour the edges of G with $\Delta(G)$ colours by colouring the edges of $G-F$ with $\Delta(G)-1$ colours and colouring the edges of F with another colour. $\square$

2.2 Matchings in bipartite graphs

In this section we give three classical theorems that will be needed in Chapter 4. The third theorem is less well known than the other two, so we provide a proof.

Theorem 2.10. (Hall, 1936 [26, 1]) *Let G be a bipartite graph with bipartition (A, B). There is a matching in G which covers A if and only if each subset $A' \subseteq A$ has at least $|A'|$ neighbours.*

If each non-empty subset $A' \subseteq A$ has at least $|A'| + k$ neighbours, then we say that A has *surplus* k. So Theorem 2.10 states that there is a matching that covers A if and only if A has non-negative surplus. If a set $A' \subseteq A$ has exactly $|A'|$ neighbours it is said to be *tight*, and if it has exactly $|A'| + 1$ neighbours it is said to be *nearly-tight*.

König's Theorem states that all bipartite graphs are class 1:

Theorem 2.11. (König, 1916 [1, 4, 14]) *Let G be a bipartite graph. Then $\chi'(G) = \Delta(G)$.*

We will also need the following theorem, due to Dulmage and Mendelsohn.

Theorem 2.12. (Dulmage and Mendelsohn, 1958 [16, 1]) *Let G be a bipartite graph with bipartition (A, B) and suppose M_1 and M_2 are two matchings in G. Then there is a matching $M \subseteq M_1 \cup M_2$ such that M covers all the vertices of A covered by M_1 and all the vertices of B covered by M_2.*

Proof. Let A' be those vertices in A covered by M_1 and let B' be those vertices in B covered by M_2. Let $M_1 \triangle M_2$ denote the symmetric difference of M_1 and M_2. Let $G_1, \ldots, G_k$ be the components of the subgraph of G consisting of the edges $M_1 \triangle M_2$. Each of these components is a path or an even cycle. If G_i is an even cycle, then both $M_1 \cap E(G_i)$ and $M_2 \cap E(G_i)$ cover all the vertices of $A' \cup B'$ that belong to G_i. If G_i is a path, then it may be the case that $M_1 \cap E(G_i)$ misses a vertex in B', or it may be the case that $M_2 \cap E(G_i)$ misses a vertex in A', but at most one of these situations can occur.

So for $i \in \{1, \ldots, k\}$, we let

$$P_i = \begin{cases} M_1 \cap E(G_i) & \text{if } M_2 \cap E(G_i) \text{ misses a vertex in } A', \\ M_2 \cap E(G_i) & \text{if } M_1 \cap E(G_i) \text{ misses a vertex in } B', \\ M_1 \cap E(G_i) & \text{otherwise.} \end{cases}$$

Then $M = (M_1 \cap M_2) \cup P_1 \cup \cdots \cup P_k$ gives the required matching. □

2.3 The Gallai-Edmonds Theorem

In this section we will state the celebrated Gallai-Edmonds Theorem, which gives a canonical partition of the vertices of a graph into three sets. We first need some more definitions.

A matching in G that misses exactly one vertex is said to be a *near 1-factor* of G. If G has near 1-factors missing every vertex it is said to be *factor-critical*. (Note that a graph with a near 1-factor is necessarily of odd order.) For a subset $S \subseteq V(G)$ we define $\mathrm{Od}(S)$ to be the set of components of $G - S$ that have odd order (these will be referred to as "odd components" from now on). A *barrier* is a subset $S \subseteq V(G)$ that maximizes $|\mathrm{Od}(S)| - |S|$.

Let $S \subseteq V(G)$. A matching M can cover all the vertices in an odd component only if M contains an edge joining a vertex in the odd component to a vertex in the barrier S. So no matching can miss fewer than $|\mathrm{Od}(S)| - |S|$ vertices. In fact we have the following theorem.

Theorem 2.13. (Berge, 1958 [31]) *The number of vertices missed by a maximum matching in G is*

$$\max \{|\mathrm{Od}(S)| - |S| : S \subseteq V(G)\}.$$

(Note that this quantity is never negative, as we can take $S = \emptyset$ to get

$|\mathrm{Od}(S)| - |S| = |\mathrm{Od}(S)| \geq 0$.)

We will now describe a way to partition $V(G)$ into three parts: $D(G)$, $A(G)$ and $C(G)$. This partition is called the *Gallai-Edmonds decomposition* of G, and is defined as follows:

(i) Let $D(G)$ be the set of vertices $v \in V(G)$ such that there is a maximum matching that misses v.

(ii) Let $A(G)$ be those vertices in $V(G) - D(G)$ adjacent to at least one vertex in $D(G)$.

(iii) Let $C(G) = V(G) - D(G) - A(G)$.

Note that we have defined $D(G)$, $A(G)$ and $C(G)$ as *sets*, although we will sometimes need to consider the subgraphs induced by these sets, and for convenience these will also be referred to as $D(G)$, $A(G)$ and $C(G)$.

Let $B(G)$ denote the graph obtained from G by deleting the vertices of $C(G)$, contracting the components of $D(G)$ to single vertices and deleting edges with both ends in $A(G)$. Thus $B(G)$ is a bipartite graph with bipartition ($A(G)$, $\mathrm{Od}(G)$).

Note that there are two degenerate cases:

(A) If G has a 1-factor, then all the vertices belong to $C(G)$, and $A(G)$ and $D(G)$ are empty.

(B) If G is factor-critical, all vertices belong to $D(G)$, and $A(G)$ and $C(G)$ are empty.

The following theorem is known as the *Gallai-Edmonds Structure Theorem.*

Theorem 2.14. (Gallai, Edmonds [31]) *If G is a graph and $D(G)$, $A(G)$ and $C(G)$ are defined as above, then*

(1) The subgraph induced by $C(G)$ has a 1-factor.

(2) Each component of the subgraph induced by $D(G)$ is factor-critical.

(3) $A(G)$ is a barrier.

(4) Any maximum matching in G consists of a 1-factor of $C(G)$, near 1-factors of each component of $D(G)$ and matches all vertices of $A(G)$ with vertices in distinct components of $D(G)$.

(5) In $B(G)$, $A(G)$ has surplus 1.

Theorem 2.13 follows as a corollary of Theorem 2.14. A consequence of condition (5) is the following.

Remark 2.15. *For any $v \in \mathrm{Od}(G)$ there is a matching in $B(G)$ that covers $A(G)$ but misses v.*

2.4 Barriers in graphs of high minimum degree

The results in this section will be used in Chapter 3. However, these results may have wider applicability. The results concern graphs of order at most $2n$ where the minimum degree is at least $\frac{1}{3}(2n)$. The first two lemmas are due to Cariolaro and Hilton [7, 5], although we give them here in more generality.

Lemma 2.16. *Let G be a graph of even order on at most $2n$ vertices with $\delta(G) \geq \frac{1}{3}(2n)$. Suppose that G has no 1-factor and that $S \subset V(G)$ is a barrier in G. Then either $S = \emptyset$ or $|S| \geq \delta(G)$.*

Proof. Since G is of even order, $|\mathrm{Od}(S)|$ and $|S|$ must have the same parity. Let $s = |S|$ and $z = |\mathrm{Od}(S)|$. We cannot have $z = s$ as G does not have a 1-factor, so we must have $z \geq s + 2$. Suppose the odd components have sizes $q_1 \leq \cdots \leq q_z$. Let v be a vertex in an odd component of minimal size. The degree of v cannot exceed $s + q_1 - 1$, as v can only be adjacent to other vertices in its odd component or vertices in S. So we have

$$\delta(G) \leq s + q_1 - 1. \tag{2.1}$$

The number of vertices in G is at least

$$\begin{aligned} s + \sum_{j=1}^{z} q_j &\geq s + zq_1 \\ &\geq s + z(\delta(G) - s + 1) && \text{(by (2.1))} \\ &\geq s + (s+2)(\delta(G) - s + 1). \end{aligned}$$

This last expression can be viewed as a quadratic in s. For $s = 1$ and $s = \delta(G) - 1$ it takes the value $3\delta(G) + 1$ and exceeds this in between. But

$$3\delta(G) + 1 > 2n,$$

which contradicts G having at most $2n$ vertices, so in fact s must either be 0 or at least $\delta(G)$. □

Suppose G is a graph of even order on at most $2n$ vertices with $\delta(G) \geq \frac{1}{3}(2n)$ that does not have a 1-factor. We can apply Lemma 2.16 to the barrier $A(G)$ given by the Gallai-Edmonds decomposition. So there are two scenarios:

1. If $|A(G)| = 0$ we say that $A(G)$ is *case 1.*

2. If $|A(G)| \geq \delta(G)$ we say that $A(G)$ is *case 2.*

We shall see that if $A(G)$ is case 1 then $D(G)$ consists of two large odd components and $C(G)$ is empty. (So G itself consists of two large connected components.) On the other hand, if $A(G)$ is case 2 then $A(G)$ and $D(G)$ are large, and $D(G)$ consists of many small odd components.

Lemma 2.17. *Suppose that $A(G)$ is case 1. Then G consists of two components of odd order each containing at least $\delta(G) + 1$ vertices.*

Proof. Since $A(G)$ is empty, a vertex in an odd or even component of $C(G) \cup D(G)$ must have all its neighbours in the same component. This means that each component must have at least $\delta(G) + 1$ vertices. We know that there must be at least two odd components. And there cannot be any other odd or even components as $\delta(G) \geq \frac{1}{3}(2n)$. Hence $D(G)$ has two odd components and $C(G)$ is empty. $\square$

For the rest of this section we will assume that G has no 1-factor and that $A(G)$ is case 2. Let $s = |A(G)|$. As observed in the proof of Lemma 2.16, $|\mathrm{Od}(S)|$ and $|S|$ must have the same parity. This means that there are at least $s + 2$ odd components of $D(G)$.

Lemma 2.18. *There are at least $(s + 2)\delta(G)$ edges between $D(G)$ and $A(G)$.*

Proof. An odd component of k vertices must have at least $k(\delta(G) - k + 1)$ edges going to vertices in $A(G)$. This is a quadratic in k with roots at 0 and $\delta(G) + 1$. When k is 1 or $\delta(G)$ it takes the value $\delta(G)$ and exceeds this in between. We claim that k must be less than $\delta(G)$. As $s \geq \delta(G)$ and there are at least $s + 2$ odd components, there are more than $2\delta(G) \geq \frac{2}{3}(2n)$ vertices in the other odd components and $A(G)$, so it must be the case that k is less than $\delta(G)$. Hence each odd component, of which there are at least $s + 2$, has at least $\delta(G)$ edges going to vertices in $A(G)$. $\square$

Lemma 2.19. *Let v be a vertex in $C(G) \cup D(G)$. Then the number of vertices of $A(G)$ adjacent to v is at least $\delta(G) - 2n + 2s + 2$.*

Proof. Besides the component containing v there are at least $s + 1$ other components, accounting for at least this many vertices. So there are at most $2n - s - (s + 1) - 1 = 2n - 2s - 2$ other vertices in the component that contains v. As v has at least $\delta(G)$ neighbours, and at most $2n - 2s - 2$ are in its own component, it must have at least

$$\delta(G) - (2n - 2s - 2) = \delta(G) - 2n + 2s + 2$$

neighbours in $A(G)$. $\square$

A *singleton* is a component of order 1.

Lemma 2.20. *The number of singletons in $D(G)$ is at least $2s - n + 3$.*

Proof. Let t be the number of singletons in $D(G)$. All other odd components have at least 3 vertices. So there are at least $s - t + 2$ odd components of $D(G)$ with at least 3 vertices each. But as there are at most $2n - s$ vertices in $D(G)$, we have

$$t + 3(s - t + 2) \leq 2n - s,$$

which gives $t \geq 2s - n + 3$. $\square$

Lemma 2.21. *If a and b are two vertices in different components of $C(G) \cup D(G)$ then they have a common neighbour in $A(G)$.*

Proof. Suppose the numbers of vertices in the components containing a and b are σ_a and σ_b respectively. The number of neighbours of a in $A(G)$ is at least $\delta(G) - \sigma_a + 1$ and the number of neighbours of b in $A(G)$ is at least $\delta(G) - \sigma_b + 1$. Now a and b will certainly have a common neighbour in $A(G)$ if

$$(\delta(G) - \sigma_a + 1) + (\delta(G) - \sigma_b + 1) > s,$$

which can be rearranged to give

$$2\delta(G) - (\sigma_a + \sigma_b - 2) > s. \tag{2.2}$$

Since there are at most $2n$ vertices, and there are s vertices in $A(G)$ and at least $s + 2$ components in $D(G)$, the number of vertices in the components containing a and b, besides these vertices themselves, cannot exceed $2n - 2s - 2$. So (2.2) is implied by

$$2\delta(G) - (2n - 2s - 2) > s,$$

which can be rearranged to give

$$2\delta(G) + s - 2n + 2 > 0. \tag{2.3}$$

But we know that

$$\delta(G) \geq \tfrac{1}{3}(2n),$$

which implies

$$3\delta(G) > 2n - 2,$$

and then (2.3) follows from the fact that $s \geq \delta(G)$. □

Let M be a matching in $B(G)$. An M*-alternating path* in $B(G)$ is a path where every other edge belongs to M. Note that the following lemma requires a stronger assumption about $\delta(G)$ than what we have assumed up until now: we need to assume $\delta(G) \geq \frac{3}{8}(2n)$ instead of $\delta(G) \geq \frac{1}{3}(2n)$.

Lemma 2.22. *Suppose $\delta(G) \geq \frac{3}{8}(2n)$. Let M be a matching in $B(G)$ of all the vertices in $A(G)$ into $\mathrm{Od}(G)$. Suppose that all singletons in $\mathrm{Od}(G)$ are covered by M. Then any unmatched vertex in $\mathrm{Od}(G)$ contains an M-alternating path to a singleton in $\mathrm{Od}(G)$.*

Proof. Let v be a vertex of $D(G)$ that is in an odd component $c(v)$. Suppose that $c(v)$ is a missed by M in $B(G)$. By Lemma 2.19, v is adjacent to at least $\delta(G) - 2n + 2s + 2$ vertices in $A(G)$. It follows that $c(v)$ is adjacent in $B(G)$ to at least this many vertices of $A(G)$, all of which are matched to vertices in $\mathrm{Od}(G) - c(v)$. By Lemma 2.20 there are at least $2s - n + 3$ singletons, and so there are at most $s + 2 - (2s - n + 3) = n - s - 1$ odd components that are not singletons.

There will be an M-alternating path of two edges from $c(v)$ to a singleton component if

$$\delta(G) - 2n + 2s + 2 > n - s - 1,$$

Figure 2.3: Exchanging a type-1 quadrilateral.

which can be rearranged to give

$$3s - 3n + 3 + \delta(G) > 0. \tag{2.4}$$

We have

$$\delta(G) \geq \tfrac{3}{8}(2n),$$

which implies

$$4\delta(G) - 3n + 3 > 0,$$

which together with the fact that $s \geq \delta(G)$ implies (2.4). $\square$

Lemma 2.23. *Let u and v be two vertices in $D(G)$. Then there are at least s vertices of G that are non-adjacent to both u and v.*

Proof. Each vertex in $D(G)$ must be non-adjacent to the vertices in the other odd components, and there are at least $s + 2$ odd components. $\square$

2.5 Quadrilaterals

As with the last section, this section will be concerned with graphs of high minimum degree; specifically graphs of order at most $2n$ where the minimum degree is at least $\frac{1}{3}(2n)$. We will consider such graphs that do not have 1-factors, and look at how the size of a maximum matching changes when they are "perturbed" by the operation of *exchanging a quadrilateral*.

Again, the results in this section are principally required in Chapter 3, but they may have other applications.

Let G be a graph of even order on at most $2n$ vertices with $\delta(G) \geq \frac{1}{3}(2n)$ that does not have a 1-factor. We will consider the Gallai-Edmonds decomposition $A(G)$, $C(G)$ and $D(G)$. Recall that by Lemma 2.16, $A(G)$ is either *case 1* or *case 2*.

A *quadrilateral* is a 4-tuple of vertices (a, b, c, d) where ac, bd are edges and ab, cd are non-edges. We will define two specializations of this definition: *type-1* and *type-2* quadrilaterals. A graph G can have a type-1 quadrilateral only if $A(G)$ is case 1, and it can have a type-2 quadrilateral only if $A(G)$ is case 2.

Suppose that $A(G)$ is case 1. In this case, by Lemma 2.17 $D(G)$ consists of two large odd components. We define a *type-1 quadrilateral* to be a 4-tuple of vertices (a, b, c, d) of G such that

(i) a, c belong to one component of $D(G)$ and b, d belong to the other.

(ii) ac, bd are edges in G (certainly ab, cd are not).

Let $G^{\sim} = G \cup \{ab, cd\} - \{ac, bd\}$. We say that $G^{\sim}$ is the graph obtained from G by *exchanging* the quadrilateral (a, b, c, d). (See Figure 2.3.)

We can imagine that we have changed the edges ac, bd into non-edges and the non-edges ab, cd into edges. The following lemma shows that exchanging a type-1 quadrilateral always increases the size of a maximum matching, and thus $G^{\sim}$ has a 1-factor.

Lemma 2.24. *$G^{\sim}$ has a 1-factor containing the edge ab.*

Proof. By part (2) of Theorem 2.14 (Gallai-Edmonds Theorem), the two components of $D(G)$ are factor-critical, so in particular they have near 1-factors missing a and b. Clearly these do not use the edges ac and bd. So we can form a 1-factor of $G^{\sim}$ by taking these two near 1-factors together with the edge ab. □

Figure 2.4: Exchanging a type-2 quadrilateral.

Suppose that $A(G)$ is case 2. In this case, we define a *type-2 quadrilateral* to be a set of four vertices (a, b, c, d) of G such that

(i) a, b belong to different odd components of $D(G)$.

(ii) c, d belong to $A(G)$.

(iii) ac, bd are edges in G and cd is not (certainly ab is not).

As before, we can exchange the quadrilateral. Let $G^{\sim} = G \cup \{ab, cd\} - \{ac, bd\}$. (See Figure 2.4.) The situation with type-2 quadrilaterals is more complicated than with type-1; exchanging does not always increase the size of a maximum matching. However, we will see that one of two things happens: either the size stays the same or is increased by 1. The rest of this section is concerned with what happens when a type-2 quadrilateral is exchanged.

Note that in what follows we will consider G and $G^{\sim}$ to be two graphs on the same set of vertices. For this reason, we will sometimes refer to $A(G)$ and $D(G)$ as subsets of $V(G^{\sim})$. When doing so, it is hoped that the reader will not confuse them with the sets $A(G^{\sim})$ and $D(G^{\sim})$ given by the Gallai-Edmonds decomposition of $G^{\sim}$.

Let $c(a)$ and $c(b)$ denote the components of $D(G)$ containing a and b respectively.

Lemma 2.25. *Let ν be the size of a maximum matching in G, and ν' the size of a maximum matching in $G^{\sim}$. Then*

(1) ν' is either ν or $\nu + 1$.

(2) In the case that $\nu' = \nu$ any matching in $B(G)$ that covers $A(G)$ and misses $c(a)$, must cover $c(b)$.

(3) In the case that $\nu' = \nu + 1$ there is a matching in $B(G)$ that covers $A(G)$ and misses both $c(a)$ and $c(b)$. Furthermore, each maximum matching in $G^{\sim}$ consists of a 1-factor of $C(G)$, the edge ab, near 1-factors of each component of $D(G)$ and matches all vertices of $A(G)$ with vertices in distinct components of $D(G)$.

(Note that condition (3) means that no vertex in $A(G)$ is matched with a vertex in the components $c(a)$ or $c(b)$.)

Proof. By Remark 2.15, there is a matching M in $B(G)$ that covers $A(G)$ and misses $c(a)$. Let M' be the matching consisting of all the edges of M, except for the edge that covers $c(b)$, should there be one. We can then form a matching of size at least ν in $G^{\sim}$ by taking a set of edges that correspond to M' (there may be some choice as to which edges are chosen, in the case that a vertex of $A(G)$ is adjacent to more than one vertex in an odd component of $D(G)$), together with the edge ab, near 1-factors of all the odd components and a 1-factor of $C(G)$. Clearly M' will not use the deleted edges ac, bd and will have size at least that of M. So $\nu' \geq \nu$.

Adding k edges to a graph increases the size of a maximum matching by at most k, and equality is obtained if and only if the set of new edges is independent and there is a maximum matching in G that is independent of the new edges.

Since there are only two edges in $G^{\sim}$ that are not present in G, the size of a maximum matching in $G^{\sim}$ is at most $\nu + 2$. But by part (4) of Theorem 2.14 (Gallai-Edmonds Theorem), the vertices c and d are covered by every maximum matching in G. So the size of a maximum matching in $G^{\sim}$ must be less than $\nu + 2$. So we have established claim (1).

Suppose that $\nu' = \nu + 1$. Let F' be a maximum matching in $G^{\sim}$. Clearly F' must use at least one of the edges ab, cd (for otherwise F' would also be a matching of size $\nu + 1$ in G). Such a matching cannot contain cd but not ab, because cd is not disjoint from any maximum matching in G. So F' contains ab. If F' contains both ab and cd then by replacing $\{ab, cd\}$ with $\{ac, bd\}$ we get a matching of size $\nu + 1$ in G, which is not possible. So the only remaining possibility is that F' contains ab but not cd. And so clearly if F' is to have size $\nu + 1$ it must match all the vertices of $A(G)$ into vertices in distinct odd components of $D(G)$ that do not include the components containing a and b. This is possible if and only if there is a matching of $A(G)$ into $\mathrm{Od}(G)$ that misses both $c(a)$ and $c(b)$. This proves claims (2) and (3). $\square$

Lemma 2.26. *Suppose that the size of a maximum matching in $G^{\sim}$ is $\nu + 1$. Then $C(G) \subseteq C(G^{\sim})$. Moreover, if P is a 1-factor of $C(G)$ then there is a 1-factor P' of $C(G^{\sim})$ that contains P.*

Proof. By part (3) of Lemma 2.25 a maximum matching in $G^{\sim}$ of size $\nu+1$ must contain a 1-factor of $C(G)$ and match all the vertices of $A(G)$ into $D(G)$. So the only vertices in $G^{\sim}$ that can be missed by a maximum matching are vertices in $D(G)$. None of the vertices in $C(G)$ have neighbours in $D(G)$, so we have $C(G) \subseteq C(G^{\sim})$.

By part (3) of Lemma 2.25 a maximum matching in $G^{\sim}$ of size $\nu + 1$ must use the edge ab. So the vertices in the components containing a and b belong to $C(G^{\sim})$. Let $U = D(G) \cap C(G^{\sim})$ and let W be the neighbours of U that belong to $A(G)$. If a vertex $w \in W$ is adjacent to a vertex $v \in D(G^{\sim})$ then there is a maximum matching in $G^{\sim}$ that misses a vertex in U, which is a contradiction. So the vertices of W have no neighbours in $D(G) - U$, and so $W \subset C(G^{\sim})$.

So $C(G^{\sim})$ contains $C(G)$, U, W and the vertices in the components containing a and b. We can form the required 1-factor P' of $C(G^{\sim})$ by taking a matching of W into U, near 1-factors of the components of U and the components containing a and b, the edge ab, and P. $\square$

The final lemma in this section is used in Chapter 3 to show that exchanging a type-2 quadrilateral cannot result in a graph $G^{\sim}$ where $A(G^{\sim})$ is case 1. (Unfortunately we do not have a more elegant way of doing this.)

Lemma 2.27. *If $2n \geq 10$ both G and $G^{\sim}$ are connected.*

Proof. Lemma 2.21 says that, in G, any two vertices in different components of $C(G) \cup D(G)$ have a common neighbour in $A(G)$. So for any such pair of vertices we can find a path of two edges connecting them via their common neighbour in $A(G)$. Let E_1 be the set of edges obtained by taking the union of these paths for each pair of vertices in different components of $C(G) \cup D(G)$. Let E_2 be a maximum matching in G.

Let H be the subgraph of G consisting of the edges $E_1 \cup E_2$. In H there is a path between any two vertices of $C(G) \cup D(G)$, and each vertex in $A(G)$ is adjacent to a vertex in $D(G)$. So G is connected, and H is a spanning subgraph of G.

Now consider $G^{\sim}$. Suppose that E_2 contains the edges av and bw for some vertices $v, w \in A(G)$. Then the subgraph of H induced by the vertices $V(G) - \{a, b, v, w\}$ is a subgraph of $G^{\sim}$, and spans the vertices $V(G) - \{a, b, v, w\}$. Hence $G^{\sim}$ has a connected component of size at least $2n - 4$. But $\delta(G^{\sim}) = \delta(G) \geq \frac{1}{3}(2n) \geq \frac{10}{3}$, and the minimum degree must be an integer, so $\delta(G^{\sim}) \geq 4$, and so any connected component must have at least 5 vertices. Hence there can only be one connected component in $G^{\sim}$, and so it is connected. $\square$

Chapter 3

The 1-factorizability of regular graphs of high degree

In this chapter we present our investigation into the 1-factorizability of regular graphs of high degree. Sections 3.1, 3.2 and 3.3 all contain preliminary material for the proof of Theorem 1.4, which is given in Section 3.4. Sections 3.5, 3.6, 3.7 and 3.8 contain the proof of Theorem 1.5. In Section 3.9 we look at the Overfull Conjecture.

3.1 Non-regular graphs from regular graphs

Sometimes we can show that a regular graph is class 1 by showing that an associated non-regular graph is class 1. The simplest example of this is the following lemma.

Lemma 3.1. *Let G be a regular graph of even order and degree d, and let x be any vertex of G. Then $\chi'(G) = \chi'(G - x)$.*

Proof. If G is complete, then the result follows from Lemma 2.3. If G is not complete, we have $\Delta(G - x) = d$. So by Theorem 2.1 (Vizing's Theorem) it will suffice to show that G is d-edge-colourable if and only if $G-x$ is d-edge-colourable. An edge-colouring of G with d colours clearly gives an edge-colouring of $G - x$

with d colours, when restricted to the edges of $G - x$. Conversely, suppose that we have an edge-colouring of $G-x$ with d colours. Because $G-x$ is of odd order, and every colour necessarily appears at an even number of vertices, each colour must be missing at an odd number of vertices. But the only vertices that miss colours are the d vertices adjacent to x in G, and these all miss exactly one colour (as they have degree $d - 1$). It follows that each of these vertices must miss a different colour, and so the edge-colouring of $G - x$ can be easily extended to an edge-colouring of G. □

One use of this lemma is the following theorem.

Theorem 3.2. *Let G be a regular graph on $2n$ vertices of degree $2n-2$ or $2n-3$. Then G is 1-factorizable.*

Proof. Let x be a vertex of G. The core of $G-x$ must certainly be a forest, as it is a graph on one or two vertices. By Theorem 2.6 (Fournier's Theorem) $G - x$ is class 1, and so G is class 1 by Lemma 3.1. □

Note that it is quite easy to show that any regular graph of order $2n$ and degree $2n-2$ is 1-factorizable without the use of this lemma: such a graph is just a complete graph minus a 1-factor.

In general, if x and y are two vertices of G, the chromatic index of $G - x - y$ does not determine the chromatic index of G. However, we have the following result for the special case where x and y have the same—or nearly the same—set of neighbours.

Lemma 3.3. *Let G be a regular graph of even order and degree d, and let x and y be two distinct vertices of G that may or may not be adjacent. Suppose that there is at most one vertex in $G - x - y$ that is adjacent to x but not y, and at least one vertex that is adjacent to neither x nor y. Then $\chi'(G) = \chi'(G - x - y)$.*

Note that since G is regular, the number of vertices adjacent to x but not y is equal to the number of vertices adjacent to y but not x.

Proof. As there is a vertex that is non-adjacent to both x and y, we have $\Delta(G - x - y) = d$. As in the proof of Lemma 3.1 it will suffice to show that G is d-edge-colourable if and only if $G - x - y$ is d-edge-colourable. Suppose that we have an edge-colouring of G with d colours. This colouring, when restricted to the edges of $G - x - y$, gives an edge-colouring of $G - x - y$ with d colours. Conversely, suppose that we have an edge-colouring of $G - x - y$ with d colours. The vertex y in $G - x$ is adjacent to at most one vertex of maximum degree, so by Lemma 2.5 we can extend our edge-colouring of $G - x - y$ to an edge-colouring of $G - x$. And then by Lemma 3.1 we can extend this edge-colouring to an edge-colouring of G. □

Note that Lemma 3.3 is true whether x is adjacent to y or not, and it is not necessary to distinguish between these cases in the proof.

3.2 A sufficient condition for a graph to be class 1

By considering the core, Theorem 2.6 (Fournier's Theorem) gives a sufficient condition for a graph to be class 1. We will need another sufficient condition, which is due to Chetwynd and Hilton [10]. It was presented slightly differently in the work of Cariolaro and Hilton [6, 5, 7]. Our presentation is different to both.

Lemma 3.4. *Let G be a graph. Suppose the vertices of G_Δ can be partitioned into two sets L and R such that the following conditions hold:*

(i) The vertices of L and R can be numbered $l_1, \ldots, l_m$ and $r_t, \ldots, r_m$ (where $m \geq 0$ and $t \leq 1$ are integers) such that the edges $l_1r_1, \ldots, l_mr_m$ are all present in G_Δ.

(ii) There are no edges between vertices of L.

(iii) There are no edges between L and R of the form $l_i r_j$ where $i > j$.

(iv) There are no edges between vertices of $\{r_t, \ldots, r_0\}$.

Then G is class 1.

Proof. We will show that G can be coloured with $\Delta(G)$ colours. As in the proof of Theorem 2.6 (Fournier's Theorem), we first colour all the edges that are incident with at most one vertex of G_Δ. Suppose we have the required partition of the vertices of G_Δ into the sets L and R. Next we colour the edges that have both ends in R. If one of the vertices is in $\{r_t, \ldots, r_0\}$, we use this vertex as the pivot. The vertices in G_Δ that are adjacent to the pivots are all in $\{l_1, \ldots, l_m, r_1, \ldots, r_m\}$ and have at least one uncoloured edge, as the edges $l_1 r_1, \ldots, l_m r_m$ are uncoloured at this stage. So by Lemma 2.4 they can all be coloured.

Next we colour the edges incident with vertices in L. We first colour the edges incident with l_1, using l_1 as a pivot and colouring the edge $l_1 r_1$ last. Then for $i = 2, \ldots, m$ we colour the edges incident with l_i, using l_i as a pivot, and colouring the edge $l_i r_i$ last. Each neighbour to l_i has at least one uncoloured edge, since l_i is not adjacent to any r_j where $i > j$ and the edges $l_i r_i, \ldots, l_m r_m$ are uncoloured. So again by Lemma 2.4 all the edges can be coloured. □

Note that, in contrast to Theorem 2.6 (Fournier's Theorem), Lemma 3.4 actually requires the presence of certain edges. Lemma 3.4 is very useful, as it gives a sufficient condition for a graph to be class 1 that covers graphs that have $\Omega(n^2)$ edges in their core, where n is the number of vertices of maximum degree. This is in contrast to the condition of Theorem 2.6 which can only be satisfied by graphs with at most $n - 1$ edges in their cores.

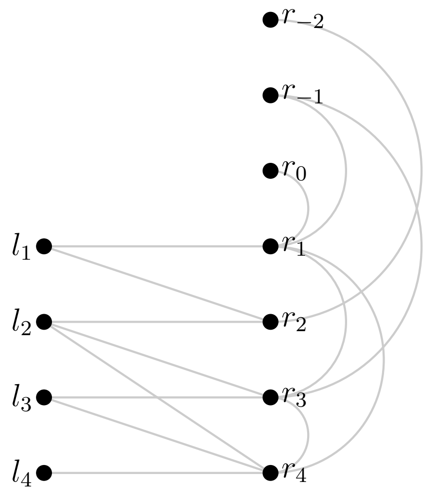


Figure 3.1: An example of a core that satisfies the conditions of Lemma 3.4.

3.3 Chetwynd-Hilton decompositions

We will now define a certain graph that depends on two parameters, and a certain edge-colouring of this graph. We will regard the edge-colouring as a decomposition of the edge set into matchings. This decomposition will be essential to the proof of Theorems 1.4 and 1.5.

The graph $\mathcal{H}_{m,t}$ depends on two parameters m and t: both are integers, $m \geq 0$ and $t \leq 1$. Let $q = 2m - t + 1$. Then we define the graph $\mathcal{H}_{m,t}$ on q vertices as follows:

(1) Take the complete graph on q vertices, and label the vertices $l_1, \ldots, l_m$ and $r_t, \ldots, r_m$.

(2) Delete all the edges with both ends in $\{r_t, \ldots, r_m\}$.

(3) Delete all edges of the form $l_i r_j$ where $i \leq j$.

(Note that $\mathcal{H}_{m,t}$ contains precisely those edges that are excluded by conditions (ii) and (iii) of Lemma 3.4.)

The *Chetwynd-Hilton decomposition* is a set of q matchings $M_1^+, \ldots, M_q^+$ in $\mathcal{H}_{m,t}$. (It can be thought of as an edge-colouring of $\mathcal{H}_{m,t}$.) We provide a formal definition below, but Figure 3.3 should make the situation clear.

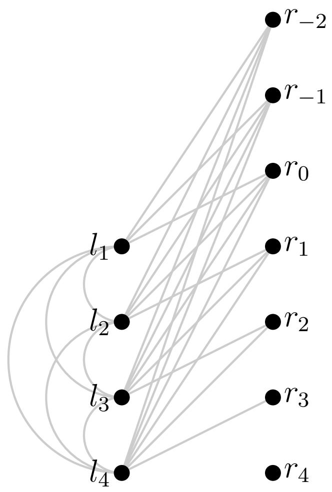


Figure 3.2: The graph $\mathcal{H}_{4,-2}$.

For $1 \le i \le 1-t$ we let

$$M_i^+ = \{l_1 r_{1-i}, l_2 r_{2-i}, \ldots, l_m r_{m-i}\}.$$

For $2-t \le i \le m+1-t$ we let

$$M_i^+ = \{l_1 l_{i+t-1}, l_2 l_{i+t-2}, \ldots, l_{\lfloor (i+t-1)/2 \rfloor} l_{\lceil (i+t-1)/2 \rceil +1}\} \\ \cup \{l_{i+t} r_t, l_{i+t+1} r_{t+1}, \ldots, l_m r_{m-i}\}.$$

For $m+2-t \le i \le 2m-t-1 = q-2$ we let

$$M_i^+ = \{l_{i-m+t} l_m, l_{i-m+t+1} l_{m-1}, \ldots, l_{\lfloor (i+t-1)/2 \rfloor} l_{\lceil (i+t-1)/2 \rceil +1}\}.$$

And finally we let

$$M_{q-1}^+ = M_q^+ = \emptyset.$$

It can be verified that each M_i^+ is a matching, and that each edge of $\mathcal{H}_{m,t}$ is contained in some matching. Figure 3.3 illustrates the Chetwynd-Hilton decom-

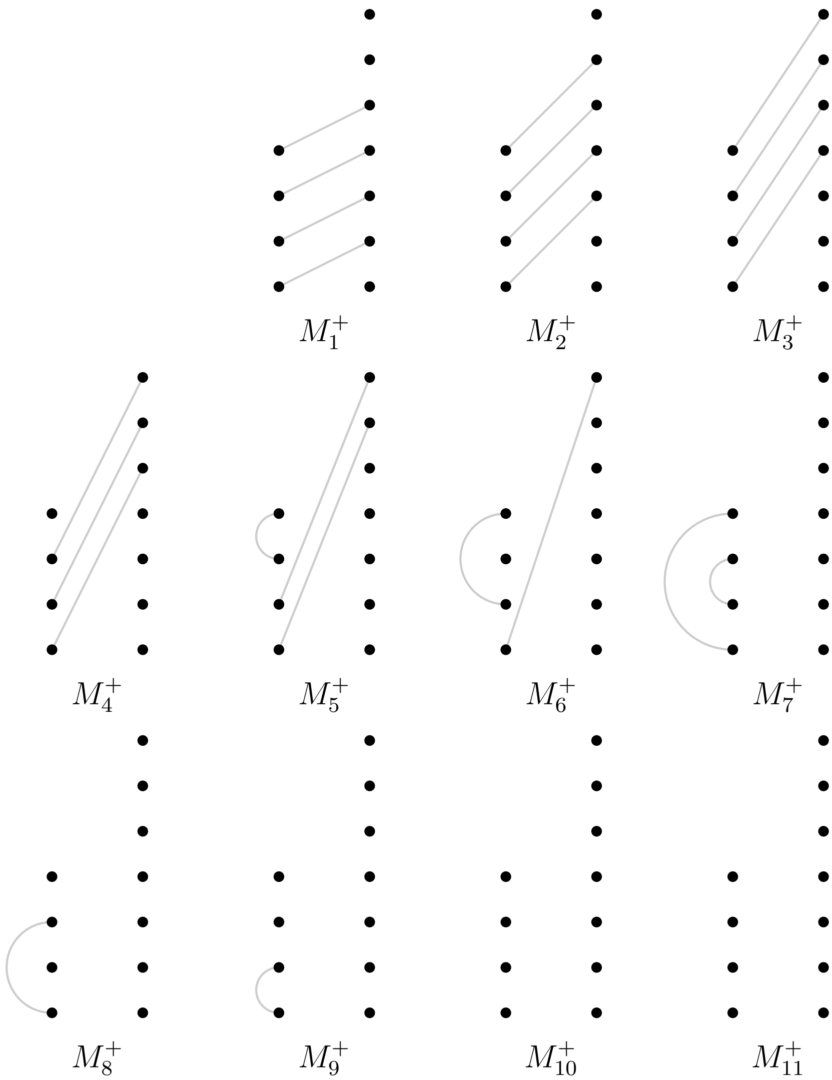


Figure 3.3: The Chetwynd-Hilton decomposition of $\mathcal{H}_{4,-2}$.

position of $\mathcal{H}_{4,-2}$.

The vertices of $\mathcal{H}_{m,t}$ can be written in a sequence

$$\mathcal{A} = r_m, \ldots, r_1, \ldots, r_t, l_1 \ldots, l_m.$$

We will need the following easy observations.

Remark 3.5. *M_i^+ misses the first i vertices of $\mathcal{A}$.*

Remark 3.6. *The number of edges in $\mathcal{H}_{m,t}$ is $m(m-t)$.*

Remark 3.7. *Suppose we fix q and then choose $m \geq 0$ and $t \leq 1$ such that*

$q = 2m - t + 1$. *Then the number of edges in $\mathcal{H}_{m,t}$ is less than $q^2/4$.*

Proof. The number of edges, $m(m-t)$, is maximized when we choose $m = \lfloor q/2 \rfloor$ and

$$t = \begin{cases} 1 & \text{if } q \text{ is even,} \\ 0 & \text{if } q \text{ is odd.} \end{cases}$$

In either case the number of edges is less than $q^2/4$. □

3.4 Proof of Theorem 1.4

In this section we will give a new proof of Theorem 1.4. While the original proof in [10] used near 1-factors, ours is simpler and just uses 1-factors. To begin with, we will prove the following special case.

Theorem 3.8. *Let G be a regular graph of order $2n$ and degree $2n\lambda$ where $\lambda \geq \frac{3}{4}$. Suppose that there are two vertices x, y such that there is at most one vertex in $V(G) - x - y$ that is adjacent to x but not y. Then G is 1-factorizable.*

Suppose we have a graph G with two vertices x and y such that the assumptions of Theorem 3.8 are satisfied. By Theorem 3.2, if the degree of G is at least $2n - 3$ then G is 1-factorizable. So we may assume that $2n\lambda < 2n - 3$. In particular this means that $2n > 12$.

We can partition the vertices of $G - x - y$ into four sets as follows:

$$\begin{aligned} X &= \text{those vertices adjacent to } x \text{ but not } y \\ Y &= \text{those vertices adjacent to } y \text{ but not } x \\ Z &= \text{those vertices adjacent to both } x \text{ and } y \\ W &= \text{those vertices not adjacent to } x \text{ or } y \end{aligned}$$

(See Figure 3.4.) Because G is regular, X and Y will be of the same size, and by our assumption this is 0 or 1. It may or may not be the case that x is adjacent

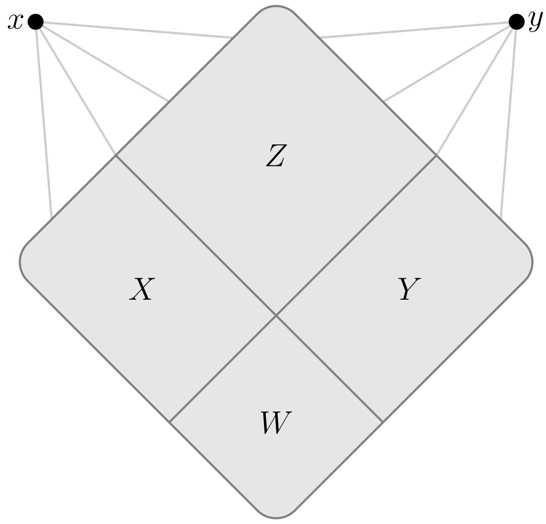


Figure 3.4: X, Y, Z and W.

to y.

Our strategy will be to show that $G - x - y$ is class 1 and then it will follow by Lemma 3.3 that G is also class 1. We will do this by removing some 1-factors from $G - x - y$ to leave a graph that is guaranteed to be class 1 by Lemma 3.4.

In $G-x-y$ the vertices of maximum degree are those belonging to W, the set of vertices that in G are non-adjacent to both x and y. The size of W depends on whether x is adjacent to y or not, and whether $|X|$ is 0 or 1. However, if $q = |W|$, in all cases we have

$$q \leq 2n(1 - \lambda) - 1. \tag{3.1}$$

Let M_0 be a maximum matching in the subgraph of G induced by W. We will define q matchings $M_1, \ldots, M_q$. In the case that M_0 is empty, we will let

$$M_1 = \cdots = M_q = \emptyset.$$

Otherwise, suppose $m = |M_0|$ and let $t = 2m - q + 1$. We can label the vertices of W as $l_1, \ldots, l_m, r_t, \ldots, r_m$ such that the edges of M_0 are $l_1r_1, \ldots, l_mr_m$, and the vertices $r_t, \ldots, r_0$ are the vertices missed by M_0. So we have now labelled the vertices in the same way as those of the graph $\mathcal{H}_{m,t}$ defined in Section 3.3. We now consider the Chetwynd-Hilton decomposition of $\mathcal{H}_{m,t}$. For $i \in \{1, \ldots, q\}$ we let

$$M_i = M_i^+ \cap E(W)$$

where $E(W)$ denotes the set of edges of G with both ends in W. The following lemma contains the core of the proof of Theorem 3.8.

Lemma 3.9. *Suppose that $F_1, \ldots, F_q$ are disjoint 1-factors of $G - x - y$ such that*

(i) No edge of M_0 is contained in any of the 1-factors.

(ii) Each edge in $M_1 \cup \cdots \cup M_q$ is contained in one of the 1-factors.

Then G is 1-factorizable.

Proof. By Lemma 3.3 it will suffice to show that $G - x - y$ is class 1. And by Lemma 2.9 it will suffice to show that the graph

$$G' = G - x - y - (F_1 \cup \cdots \cup F_q)$$

is class 1. The core of G' is the subgraph induced by the vertices in W. These vertices have degree $2n\lambda - q$. (The vertices in Z have degree $2n\lambda - q - 2$ and in the case that X and Y contain a vertex each, these vertices have degree $2n\lambda - q - 1$.)

If M_0 is empty, then the core of G' contains no edges, and so G' is class 1 by Theorem 2.6. Otherwise, we can use Lemma 3.4 to show that G' is class 1. We have already labelled the vertices of W as $l_1, \ldots, l_m, r_t, \ldots, r_m$ so we just need to check that the conditions from the statement of Lemma 3.4 hold. None of the edges of M_0 are contained in any of the 1-factors $F_1, \ldots, F_q$, so all are present in G' and therefore condition (i) is satisfied. Because M_0 is a maximal matching, there are no edges between vertices of $\{r_t, \ldots, r_0\}$, so condition (iv) holds. And because all the edges from $M_1 \cup \cdots \cup M_q$ are contained in the 1-factors $F_1, \ldots, F_q$, there are no edges present in G' of the form $l_i r_j$ where $i > j$, and no edges with both ends in $\{l_1, \ldots, l_m\}$. So conditions (ii) and (iii) hold, and by Lemma 3.4 G' is class 1. $\square$

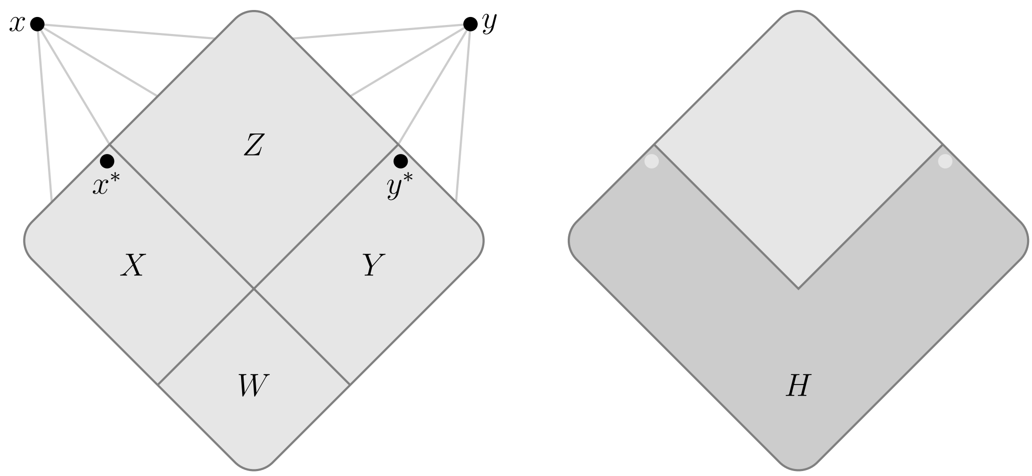


Figure 3.5: X, Y, Z, W and H.

To complete the proof of Theorem 3.8, we need to show that 1-factors $F_1, \dots, F_q$ satisfying the conditions in the statement of Lemma 3.9 exist when $\lambda \geq \frac{3}{4}$.

Suppose—for a contradiction—that they do not exist. Then for some t, where $1 \leq t \leq q$, there exist disjoint 1-factors $F_1, \dots, F_{t-1}$ such that:

(1) No edge of M_0 is contained in any of the 1-factors.

(2) Each edge in $M_1 \cup \cdots \cup M_{t-1}$ is contained in one of the 1-factors.

We may suppose that 1-factors $F_1, \dots, F_{t-1}$ have been chosen so that t is as large as possible. We will show that if $\lambda \geq \frac{3}{4}$ another 1-factor F_t can be found, which will contradict the assumption that the 1-factors have been chosen so as to maximize t.

Proof of Theorem 3.8. Let

$$M'_t = M_t - (F_1 \cup \cdots \cup F_{t-1})$$

and

$$G_t = G - x - y - V(M'_t) - (M_0 \cup F_1 \cup \cdots \cup F_{t-1}).$$

We will show that if $\lambda \geq \frac{3}{4}$ then G_t has a 1-factor F'_t, from which it follows that

$G - x - y$ has a 1-factor

$$F_t = F'_t \cup M'_t$$

such that $F_1, \ldots, F_t$ satisfy the above conditions (1) and (2). This will contradict our assumption that the 1-factors $F_1, \ldots, F_{t-1}$ were chosen to maximize t.

Let us calculate a lower bound on the minimum degree of G_t. By Remark 3.5, M_t—and therefore M'_t—covers at most $q-t$ vertices. All vertices in G are incident with an edge from each 1-factor $F_1, \ldots, F_{t-1}$, and some vertices are adjacent to x or y. Vertices that are not adjacent to x or y in G are covered by M_0. So we have

$$\begin{aligned}
\delta(G_t) &\geq 2n\lambda - 2 - (q - t) - (t - 1) \\
&= 2n\lambda - q - 1 \\
&\geq 2n\lambda - (2n(1 - \lambda) - 1) - 1 \qquad\qquad \text{(by (3.1))} \\
&= 2n(2\lambda - 1).
\end{aligned}$$

When $\lambda \geq \frac{3}{4}$, we have $2\lambda - 1 \geq \frac{1}{2}$ and so the minimum degree of G_t will be over half the order of G_t and so G_t will have a 1-factor F'_t by Corollary 2.8. □

We will now move on to the proof of Theorem 1.4. In light of Theorem 3.8, we only need to consider the case where we have chosen x and y so that $|X| = |Y| > 1$. We will need to choose two arbitrary vertices $x^* \in X$ and $y^* \in Y$. (See Figure 3.5.) Again, we will show that G is 1-factorizable by showing that $G - x - y$ is class 1. However, this will take a little bit more work.

Let H be the set of vertices $(X - x^*) \cup (Y - y^*) \cup W$. Let $q = |H|$ and let $k = |X| - 1 = |Y| - 1$ (by our assumption, $k \geq 1$). Let M_0 be a maximum matching in the subgraph of G induced by H. Again, we define q matchings $M_1, \ldots, M_q$. In the case that M_0 is empty, we let

$$M_1 = \cdots = M_q = \emptyset.$$

Otherwise, we can label the vertices of H as $l_1, \ldots, l_m, r_t, \ldots, r_m$ such that the edges of M_0 are $l_1 r_1, \ldots, l_m r_m$, and the vertices $r_t, \ldots, r_0$ are the vertices missed by M_0. As before, we consider the Chetwynd-Hilton decomposition of $\mathcal{H}_{m,t}$. For $i \in \{1, \ldots, q\}$ we let

$$M_i = M_i^+ \cap E(H)$$

where $E(H)$ denotes the set of edges of G with both ends in H.

Our next lemma is similar to Lemma 3.9, with the difference that it involves two kinds of matchings: 1-factors of G and 1-factors of $G - x - y$.

Lemma 3.10. *Suppose that there exist disjoint matchings $F_1, \ldots, F_q$ in G, such that*

(i) Exactly k of the matchings are 1-factors of G and the rest are 1-factors of $G - x - y$.

(ii) The matchings that are 1-factors of G each contain an edge joining x to $X - x^*$ and an edge joining y to $Y - y^*$.

(iii) Each edge in $M_1 \cup \cdots \cup M_q$ is contained in one of the matchings.

(iv) No edge of M_0 is contained in any of the matchings.

Then G is 1-factorizable.

Proof. Let

$$G' = G - x - y - (F_1 \cup \cdots \cup F_q).$$

The core of G' is the subgraph of G' induced by the vertices in H (the vertices in H have degree $2n\lambda - q$, x^* and y^* have degree $2n\lambda - q - 1$, and the vertices in Z have degree $2n\lambda - q - 2$).

If M_0 is empty, then the core of G' contains no edges, and so G' is class 1 by Theorem 2.6. Otherwise, we can use Lemma 3.4 to show that G' is class 1. We have already labelled the vertices of H as $l_1, \ldots, l_m, r_t, \ldots, r_m$ so we just need to check that the conditions from the statement of Lemma 3.4 hold. None of the edges of M_0 are contained in any of the matchings $F_1, \ldots, F_q$, so all are present in G' and therefore condition (i) is satisfied. Because M_0 is a maximal matching, there are no edges between vertices of $\{r_t, \ldots, r_0\}$, so condition (iv) holds. And because all the edges from $M_1 \cup \cdots \cup M_q$ are contained in the 1-factors $F_1, \ldots, F_q$, there are no edges present in G' of the form $l_i r_j$ where $i > j$, and no edges with both ends in $\{l_1, \ldots, l_m\}$. So conditions (ii) and (iii) hold, and by Lemma 3.4 G' is class 1.

Let G'' be the graph obtained from G' by adding all the edges in those matchings from $F_1, \ldots, F_q$ that are 1-factors of $G - x - y$. By Lemma 2.9 G'' is class 1. There are k remaining matchings in $F_1, \ldots, F_q$ that are all 1-factors of G containing an edge from x to X and an edge from y to Y. Suppose the matchings are $F_{i_1}, \ldots, F_{i_k}$. Let G''' be the graph obtained from G'' by adding in the vertices x and y together with all the edges joining them to their neighbours in G except those contained in $F_{i_1}, \ldots, F_{i_k}$. (So in G''' the vertex x is joined to the vertices in Z and the vertex x^*, and to no other vertices. Similarly, the vertex y is joined to the vertices in Z and the vertex y^*.) It follows from Lemma 3.3 that G''' is class 1. And since G is obtained from G''' by adding k 1-factors, it follows from Lemma 2.9 that G is class 1. $\square$

Before proving Theorem 1.4, we need the following lemma. We wish to choose the vertices x and y so that q is as small as possible. Since $q = |H| = 2n-4-|Z|$, we need to choose x and y so that Z is as large as possible. In other words, we need to choose two vertices x and y that have as many common neighbours as possible.

Lemma 3.11. *We can choose the vertices x and y so that*

$$q < 2n(1 - \lambda^2) - 3.$$

Proof. Let $d = 2n\lambda$, the degree of G. There are $2n\binom{d}{2}$ paths of length 2 in G. The average over all unordered pairs of vertices is

$$\frac{2n\binom{d}{2}}{\binom{2n}{2}} = \frac{d(d-1)}{2n-1}.$$

So there must exist two vertices with at least this many common neighbours. So we can choose x and y so that

$$|Z| \geq \frac{d(d-1)}{2n-1}.$$

We have

$$\begin{aligned} q &= |W| + |X| + |Y| - 2 \\ &= 2n - 4 - |Z| \\ &\leq 2n - 4 - \frac{(2n\lambda)^2 - 2n\lambda}{2n-1} \\ &< 2n - 4 - \frac{(2n\lambda)^2 - 2n\lambda}{2n} \\ &= 2n - 4 - 2n\lambda^2 + \lambda \\ &\leq 2n(1-\lambda^2) - 3. \end{aligned}$$ □

Recall from Section 3.3 the sequence

$$\mathcal{A} = r_m, \ldots, r_1, \ldots, r_t, l_1 \ldots, l_m.$$

Label the vertices in $X - x^*$ as $x_1, \ldots, x_k$ according to the order in which they appear in $\mathcal{A}$, and similarly label the vertices in $Y - y^*$ as $y_1, \ldots, y_k$ according to the order in which they appear in $\mathcal{A}$.

By Remark 3.5, M_i misses (at least) the first i vertices in $\mathcal{A}$. For $i \in \{1, \ldots, k\}$, x_i and y_i must both occur in the first $q - k + i$ vertices of $\mathcal{A}$. So it is certain that M_{q-k+i} misses the vertices x_i and y_i.

We will show that as long as λ is sufficiently large, there exist matchings $F_1, \ldots, F_q$ so that:

(1) $F_1, \ldots, F_{q-k}$ are 1-factors of $G - x - y$.

(2) $F_{q-k+1}, \ldots, F_q$ are 1-factors of G.

(3) For $i \in \{1, \ldots, k\}$, F_{q-k+i} contains the edges $x_i x$ and $y_i y$.

(4) No edge of M_0 is contained in any of the matchings.

(5) Each edge in $M_1 \cup \cdots \cup M_q$ is contained in one of the matchings.

Note that these conditions are stronger than those in the statement of Lemma 3.10. Suppose—for a contradiction—that such matchings $F_1, \ldots, F_q$ do not exist. Then for some t, where $1 \leq t \leq q$, there exist matchings $F_1, \ldots, F_{t-1}$ such that the conditions (1)–(4) above are satisfied along with the condition:

(5′) Each edge in $M_1 \cup \cdots \cup M_{t-1}$ is contained in one of the matchings.

As before, we will assume that the matchings $F_1, \ldots, F_{t-1}$ are chosen so as to maximize t. Let

$$M'_t = M_t - (F_1 \cup \cdots \cup F_{t-1})$$

and

$$G_t = \begin{cases} G - x - y - \{M_0 \cup F_1 \cup \cdots \cup F_{t-1}\} & \\ \qquad -V(M'_t) & \text{if } 1 \leq t \leq q-k, \\ G - x - y - \{M_0 \cup F_1 \cup \cdots \cup F_{t-1}\} & \\ \qquad -V(M'_t) - x_{t-q+k} - y_{t-q+k} & \text{if } q-k+1 \leq t \leq q. \end{cases} \tag{3.2}$$

We will show that if λ is sufficiently large, G_t has a 1-factor F'_t. We can then let

$$F_t = \begin{cases} F'_t \cup M'_t & \text{if } 1 \leq t \leq q-k, \\ F'_t \cup M'_t \cup \{x_{t-q+k}x, y_{t-q+k}y\} & \text{if } q-k+1 \leq t \leq q. \end{cases}$$

And then $F_1, \ldots, F_t$ will satisfy the above conditions. This will contradict our assumption that the 1-factors $F_1, \ldots, F_{t-1}$ were chosen to maximize t. As before, we will give a lower bound on the minimum degree of G_t.

The following lemma is similar to one from [5, 7], although we avoid an inconvenient constant term that their version contains.

Lemma 3.12. *If we choose x and y so that q is as small as possible then*

$$\delta(G_t) \geq 2n(\lambda^2 + \lambda - 1).$$

Proof. Note that a vertex cannot be adjacent to both x and y and be covered by M_0. The number of vertices covered by M'_t is at most $q - t$. It follows that the degree of a vertex in G_t is at least

$$\begin{aligned}
\delta(G_t) &\geq 2n\lambda - 4 - (q - t) - (t - 1) \\
&= 2n\lambda - q - 3 \\
&> 2n\lambda - (2n(1 - \lambda^2) - 3) - 3 && \text{(by Lemma 3.11)} \\
&= 2n(\lambda^2 + \lambda - 1).
\end{aligned}$$

□

Proof of Theorem 1.4. Let G be a regular graph of order $2n$ and degree $2n\lambda$ where

$$\lambda > \tfrac{1}{2}(\sqrt{7} - 1) \approx 0.82288.$$

By Theorem 3.2, if the degree of G is at least $2n - 3$ then G is 1-factorizable. So we will assume that $2n\lambda < 2n - 3$. (This means that $2n \geq 18$.)

We choose two vertices x and y so that q is as small as possible. If $|X| = |Y|$ is 0 or 1 then G is 1-factorizable by Theorem 3.8, so we may assume that this is not the case. We will follow the strategy outlined above, and assume that $F_1, \ldots, F_{t-1}$ have been chosen so as to maximize t. As observed above, if the graph G_t has a

1-factor, then there exists an F_t satisfying the appropriate conditions, which will give us a contradiction. By Corollary 2.8 we can be certain that G_t has a 1-factor F_t' if the minimum degree is at least half its order. But then by Lemma 3.12 this will be the case if

$$\lambda^2 + \lambda - 1 > \tfrac{1}{2},$$

which is equivalent to

$$2\lambda^2 + 2\lambda - 3 > 0.$$

But we have chosen λ to be larger than the positive root of this quadratic. $\square$

3.5 Critical quadrilaterals

We now turn to the task of improving Theorem 1.4. Suppose that G is a regular graph of order $2n$ and degree $2n\lambda$. If

$$\lambda > \tfrac{1}{6}(\sqrt{57} - 3) \approx 0.75831,$$

then we have

$$\lambda^2 + \lambda - 1 > \tfrac{1}{3},$$

and so by Lemma 3.12 we have $\delta(G_t) > \frac{1}{3}(2n)$. This means that $\delta(G_t)$ satisfies the assumptions for Sections 2.4 and 2.5, namely that G is a graph on at most $2n - 2$ vertices with minimum degree at least $\frac{1}{3}(2n)$.

Let us assume that

$$0.75831 \approx \tfrac{1}{6}(\sqrt{57} - 3) < \lambda < \tfrac{1}{2}(\sqrt{7} - 1) \approx 0.82288.$$

For λ in this range, we can use the machinery that we used to prove Theorem 1.4, except that we cannot use Corollary 2.8 to show that G_t has a 1-factor. In fact, for such λ we cannot be sure that G_t has a 1-factor, so we will adopt a slightly different strategy.

We will also change our assumptions slightly about how the matchings $F_1, \ldots,$ F_{t-1} are chosen. We will assume that:

(1) $F_1, \ldots, F_{q-k}$ are 1-factors of $G - x - y$.

(2) $F_{q-k+1}, \ldots, F_q$ are 1-factors of G.

(3) F_{q-k+i} contains the edges $x_i x$ and $y_i y$ for $i \in \{1, \ldots, k\}$.

(4) No edge of M_0 is contained in any of the matchings.

(5) Each edge in $M_1 \cup \cdots \cup M_{t-1}$ is contained in one of $F_1, \ldots, F_{t-1}, P$ where P is a 1-factor of $C(G_t)$.

We will assume that subject to these conditions, we have chosen the matchings $F_1, \ldots, F_{t-1}$ so that:

(i) t is as large as possible.

(ii) Subject to (i), G_t has as large a matching as possible.

We will show that for sufficiently large λ, either G_t has a 1-factor, or for some $i \in \{1, \ldots, t-1\}$ we can *change* F_i to obtain a new matching F_i' so that either:

(A) We can find a matching F_t so that $F_1, \ldots, F_i', \ldots, F_t$ satisfies conditions (1)-(5) above, or

(B) $F_1, \ldots, F_i', \ldots, F_{t-1}$ satisfies the conditions above, but the new G_t has a larger matching than the old one.

So we will obtain a contradiction if we assume that the matchings $F_1, \ldots, F_{t-1}$ are chosen subject to the conditions (i) and (ii) above. It will then follow that the required matchings $F_1, \ldots, F_q$ exist.

The method we will use to change the matchings is that of *exchanging a quadrilateral*, which was defined in Section 2.5. In fact, we will only exchange a certain kind of quadrilateral, which we will call a *critical quadrilateral*.

So from now on we will assume that G_t does not have a 1-factor. Consider the Gallai-Edmonds decomposition of G_t. Recall from Section 2.4 that $A(G_t)$ is either case 1 or case 2. Note that we will say that an edge is *marginal* if it belongs to $M_1 \cup \cdots \cup M_{t-1}$.

Suppose that $A(G_t)$ is case 1. Then a *critical type-1 quadrilateral* is a type-1 quadrilateral (a, b, c, d) where

(i) ab, cd are edges in F_i for some $i \in \{1, \ldots, t-1\}$.

(ii) cd is not marginal.

The following lemma is due to Cariolaro and Hilton [7, 5], although our presentation of it here is different. The proof is quite technical, and will be postponed until Section 3.7.

Lemma 3.13. *If $A(G_t)$ is case 1 and λ is greater than the second largest root of $\lambda^4 - \lambda^3 - 4\lambda^2 + 2\lambda + 1$ (≈ 0.78526) then G_t contains a critical type-1 quadrilateral.*

Suppose that $A(G_t)$ is case 2. A vertex in $A(G_t)$ is said to be *inflexible* if it is adjacent to at least $2n - 2\delta(G_t)$ vertices outside $A(G_t)$, otherwise it is *flexible*.

A *critical type-2 quadrilateral* is a type-2 quadrilateral (a, b, c, d) where

(i) ab, cd are edges in F_i for some $i \in \{1, \ldots, t-1\}$.

(ii) cd is not marginal.

(iii) If $c(a)$ and $c(b)$ denote the odd components of $D(G_t)$ containing a and b respectively, then in $B(G_t)$, $c(a)$ and $c(b)$ are not contained in the neighbourhood of a nearly tight set of flexible vertices.

The proof of this lemma is even more intricate than that of the previous lemma, and will be postponed until Section 3.8.

Lemma 3.14. *If $A(G_t)$ is case 2 and λ is greater than the second largest root of*

$$4x^6 - 28x^5 - 71x^4 + 54x^3 + 88x^2 - 62x + 3$$

(≈ 0.81112)*, then G_t contains a critical type-2 quadrilateral.*

3.6 Proof of Theorem 1.5

We will now turn to the proof of Theorem 1.5. The following lemma shows that exchanging a critical quadrilateral allows us to "do better" than before in our choice of matchings.

For this lemma we will need to assume that

$$\lambda > \tfrac{1}{4}(\sqrt{26} - 2) \approx 0.77475.$$

This ensures that

$$\lambda^2 + \lambda - 1 > \tfrac{3}{8},$$

and so by Lemma 3.12 we have $\delta(G_t) > \frac{3}{8}(2n)$ and so G_t satisfies the assumptions of Lemma 2.22. We will also assume that $2n \geq 10$, so that we can use Lemma 2.27.

Recall that we are supposing that we have chosen matchings $M_1, \ldots, M_{t-1}$ so that

(i) t is as large as possible.

(ii) Subject to (i), G_t has as large a matching as possible.

And we are requiring, among other things, that

$$M_1 \cup \cdots \cup M_{t-1} \subset F_1 \cup \cdots \cup F_{t-1} \cup P \tag{3.3}$$

where P is a 1-factor of $C(G_t)$.

A crucial step in the proof of the following lemma is where it is shown that exchanging a type-2 quadrilateral increases the size of a maximum matching in G_t. This method appears to be novel, and it may have other applications.

Lemma 3.15. *Suppose there is a critical quadrilateral (a, b, c, d) where the edges ab, cd are from the matching F_i for some $i \in \{1, \ldots, t-1\}$ that is type-1 or type-2 according to whether $A(G_t)$ is case 1 or case 2. Let*

$$F_i' = F_i \cup \{ac, bd\} - \{ab, cd\},$$

and let

$$G_t^{\sim} = G_t \cup \{ab, cd\} - \{ac, bd\}.$$

Then either

(A) there is a 1-factor of $G_t^{\sim}$ and we can find the desired F_t with

$$M_1 \cup \cdots \cup M_t \subset F_1 \cup \cdots \cup F_i' \cup \cdots \cup F_t,$$

or

(B) there is a matching in $G_t^{\sim}$ larger than any in G_t, and a 1-factor P' of $C(G_t^{\sim})$ such that

$$M_1 \cup \cdots \cup M_{t-1} \subset F_1 \cup \cdots \cup F_i' \cup \cdots \cup F_{t-1} \cup P'.$$

(Note that $G_t^{\sim}$ is the graph obtained if $F_1 \cup \cdots \cup F_i' \cup \cdots \cup F_t$ is used in (3.2) in place of $F_1 \cup \cdots \cup F_t$.)

Proof. We will consider the Gallai-Edmonds decomposition of G_t and its associated bipartite graph $B(G_t)$.

In the case that $A(G_t)$ is case 1, $C(G_t)$ is empty and so P is empty and can contain no marginal edges. By Lemma 2.24 we can find a 1-factor of $G_t^{\sim}$ that contains the edge ab. And so as before we can find F_t. So we have situation (A). From now on, we will assume that $A(G_t)$ is case 2.

Let ν be the size of a maximum matching in G_t. By part (1) of Lemma 2.25 the size of a maximum matching in $G_t^{\sim}$ is either ν or $\nu + 1$. We will show that in fact the first case cannot occur.

So suppose—for a contradiction—that the size of a maximum matching in $G_t^{\sim}$ is ν. By part (5) of Theorem 2.14 there is a matching M in $B(G_t)$ that covers $A(G_t)$ but misses $c(a)$. And by part (2) of Lemma 2.25, M must cover $c(b)$. By condition (iii) of the definition of a type-2 critical quadrilateral, it follows that there is an M-alternating path starting at $c(b)$ with an edge of M and ending at an inflexible vertex v of $A(G_t)$. Let M' be the matching obtained by exchanging on this alternating path. We can then form a matching F' in $G_t^{\sim}$ of size ν by taking a set of edges corresponding to the matching M', the edge ab, near 1-factors of all the odd components, and a 1-factor of $C(G_t)$.

If M misses a vertex of $\mathrm{Od}(G_t)$ that corresponds to a singleton vertex u, then let $M'' = M$. Otherwise, by Lemma 2.22, there must be an M-alternating path joining an unmatched vertex of $\mathrm{Od}(G_t)$ to a vertex of $\mathrm{Od}(G_t)$ that corresponds to a singleton u that is covered by M. Let M'' be the matching obtained by exchanging on this alternating path. We can form a matching F'' in $G_t^{\sim}$ that has size ν by taking a set of edges corresponding to M'' minus the edge that covers $c(b)$, the edge ab, near 1-factors of all the odd components, and a 1-factor of $C(G_t)$.

So we have two maximum matchings F' and F'' in $G_t^{\sim}$, with F' missing the inflexible vertex v and F'' missing the singleton u. Because they are missed by maximum matchings, both u and v must be in $D(G_t^{\sim})$ by the construction of the

Gallai-Edmonds decomposition. However, u is adjacent to at least $\delta(G_t)$ vertices of $A(G_t)$ and v is adjacent to at least $2n - 2\delta(G_t)$ vertices that are not in $A(G_t)$. So the vertex set $\{u, v\}$ has at least than $2n - \delta(G_t)$ neighbours, and so at most $\delta(G_t) - 2$ non-neighbours.

But if $A(G_t^{\sim})$ is case 2, any pair of vertices in $D(G_t^{\sim})$ must have at least $\delta(G_t)$ non-neighbours by Lemma 2.23. So $A(G_t^{\sim})$ cannot be case 2. But it cannot be case 1 either, because by Lemma 2.27 $G_t^{\sim}$ is connected. But if the size of a maximum matching in $G_t^{\sim}$ is ν there is no 1-factor, and so $A(G_t^{\sim})$ must be case 1 or case 2. So we have a contradiction.

We conclude that the size of a maximum matching in $G_t^{\sim}$ is $\nu + 1$. By Lemma 2.25 there is a 1-factor P' of $C(G_t^{\sim})$ that contains P. If $G_t^{\sim}$ has a 1-factor then $A(G_t^{\sim})$ and $D(G_t^{\sim})$ are empty and so P' is a 1-factor of $G_t^{\sim}$. And so in this case as before we can find F_t and we have situation (A). Otherwise we have situation (B). □

Proof of Theorem 1.5. Suppose λ is greater than the second largest root of

$$4x^6 - 28x^5 - 71x^4 + 54x^3 + 88x^2 - 62x + 3$$

(≈ 0.81112). By Theorem 3.2, if the degree of G is at least $2n - 3$ then G is 1-factorizable, so we will assume that $2n\lambda < 2n - 3$. (This means that $2n \geq 16$.) In addition, if $\lambda > \frac{1}{2}(\sqrt{7} - 1) \approx 0.82288$ then G is 1-factorizable by Theorem 1.4 so we may also assume this is not the case.

We choose two vertices x and y so that q is as small as possible. If $|X| = |Y|$ is 0 or 1 then G is 1-factorizable by Theorem 3.8, so we may also assume that this is not the case.

We claim that we can choose the required matchings $F_1, \ldots, F_q$ one at a time. Suppose we have chosen the first $t - 1$ matchings, but that we cannot choose F_t because G_t does not have a 1-factor. As mentioned already, we may assume that we have chosen the matchings subject to the condition that t is as large as possible, and subject to this, that G_t has as large a matching as possible.

If $A(G_t)$ is case 1 then by Lemma 3.13 there exists a critical type-1 quadrilateral, and if $A(G_t)$ is case 2 then by Lemma 3.14 there exists a critical type-2 quadrilateral. But then by Lemma 3.15 we can adjust our choice of matchings so that either we can choose F_t or if not, we can make G_t have a larger matching. In either case we contradict our assumptions about the choice of $F_1 \cup \cdots \cup F_{t-1}$. It follows that all the required matchings $F_1, \ldots, F_q$ can be found. $\square$

3.7 Proof of Lemma 3.13

In the following two sections we will use δ and Δ to denote the minimum and maximum degrees of G_t, $\delta(G_t)$ and $\Delta(G_t)$ respectively.

In this section we will assume that $A(G_t)$ is case 1. Recall from Lemma 2.17 that $D(G)$ consists of two large odd components, each with at least $\delta+1$ vertices. Suppose these components are called D_1 and D_2 and that they contain d_1 and d_2 vertices respectively. Let κ be the number of edges in the matchings $F_1, \ldots, F_{t-1}$ that have one end in D_1 and the other in D_2.

Lemma 3.16.

$$\kappa > (2n)^2(\lambda^2 + \lambda - 1)(\lambda^2 + 2\lambda - 2).$$

Proof. In G, any vertex in non-adjacent to $2n(1-\lambda) - 1$ other vertices. So in G a vertex $v \in D_1$ is adjacent to at least $d_2 - (2n(1-\lambda) - 1)$ vertices in D_2, and at most one of these is joined to v by an edge in M_0. So in G, v is joined to at least $d_2 - 2n(1-\lambda)$ vertices in D_2 by edges that are not present in G_t or M_0, so must be contained in the matchings $F_1, \ldots, F_{t-1}$. So

$$\begin{aligned}\kappa &\geq d_1(d_2 - 2n(1-\lambda)) \\ &\geq (\delta+1)(\delta+1-2n(1-\lambda)) \\ &> (2n)^2(\lambda^2+\lambda-1)(\lambda^2+2\lambda-2). \qquad \text{(by Lemma 3.12)}\end{aligned}$$ □

Suppose that F_i, where $i \in \{1, \ldots, t-1\}$, is the matching that contains the most edges that have one end in D_1 and the other end in D_2.

Lemma 3.17. *If $\kappa > \frac{1}{2}q^2$ then F_i contains at least one non-marginal edge that has one end in D_1 and the other end in D_2.*

Proof. It must be the case that F_i has at least

$$\frac{\kappa}{t-1}$$

edges with one end in D_1 and the other end in D_2. Since $t-1 < q$ this number is more than κ/q. No matching can contain more than $q/2$ marginal edges. So if $\kappa/q > q/2$ it is certain that F_i contains at least one non-marginal edge that has one end in D_1 and the other end in D_2. □

Suppose c and d are two vertices in G_t such that cd is a non-marginal edge in F_i that has one end in D_1 and the other end in D_2. Let F_i^* be those edges in F_i that have one end in D_1 and the other end in D_2, excluding the edge cd.

Lemma 3.18. *If $|F_i^*| > 2n - 2\delta$ then there are vertices a, b such that (a, b, c, d) is a critical type-1 quadrilateral.*

Proof. We will find vertices a, b such that ac, bd are edges in G_t and ab is an edge in F_i^*. (Note that we have already found cd, a non-marginal edge in F_i.)

The vertex c is adjacent to at least $|F_i^*| - (d_1 - 1 - \delta)$ vertices that are ends of edges in F_i^*. If we look at the other ends of these edges, d can be non-adjacent to at most $(d_2 - 1 - \delta)$ of them. So there will be suitable vertices a, b if

$$|F_i^*| - (d_1 - 1 - \delta) - (d_2 - 1 - \delta) > 0. \tag{3.4}$$

But we have assumed that

$$|F_i^*| - 2n + 2\delta > 0,$$

which, as $d_1 + d_2 < 2n$, implies

$$|F_i^*| - (d_1 + d_2) + 2\delta > 0,$$

which implies (3.4). $\square$

Proof of Lemma 3.13. We wish to show that for λ sufficiently large $\kappa > \frac{1}{2}q^2$ and $|F_i^*| > 2n - 2\delta$. By Lemmas 3.11 and 3.16 the first will certainly hold if

$$(2n)^2(\lambda^2 + \lambda - 1)(\lambda^2 + 2\lambda - 2) > \tfrac{1}{2}(2n)^2(1 - \lambda^2)^2. \tag{3.5}$$

Since $|F_i^*| > \kappa/q$ the second condition will hold if

$$\kappa/q > 2n - 2\delta,$$

which by Lemmas 3.11 and 3.16 will hold if

$$(2n)\frac{(\lambda^2 + \lambda - 1)(\lambda^2 + 2\lambda - 2)}{1 - \lambda} > 2n(3 - 2\lambda - 2\lambda^2). \tag{3.6}$$

Both (3.5) and (3.6) are satisfied if λ is greater than the second largest root of

$$\lambda^4 - \lambda^3 - 4\lambda^2 + 2\lambda + 1$$

(≈ 0.78526). $\square$

3.8 Proof of Lemma 3.14

In this section we will assume that $A(G_t)$ is case 2. Let w be the number of vertices in $G - x - y$ that are not present in G_t (so w will count the vertices in $V(M'_t)$ and x_{t-q+k} and y_{t-q+k} in the case that $t \geq q - k + 1$). As before let $s = |A(G_t)|$. Clearly we have

$$w \leq 2n - 2s - 4.$$

Note that in the proof of Theorem 1.4 we established that if $\lambda > \frac{1}{2}(\sqrt{7} - 1) \approx 0.82288$ then G_t has a 1-factor. So it must be the case that λ is less than this bound. (Incidentally, the following lemma also holds if $A(G_t)$ is case 1.)

Lemma 3.19.

$$\Delta - \delta \leq w + 2.$$

Proof. In $G - \{F_1 \cup \cdots \cup F_{t-1}\}$ all vertices except x and y have degree $2n\lambda - t + 1$. Removing x, y and the edges of M_0 reduces the degrees of the remaining vertices by at most 2, and removing the vertices of $V(M'_t)$ (and x_{t-q+k} and y_{t-q+k} in the case that $t \geq q - k + 1$) reduces the degrees further by at most w. □

Lemma 3.20.

$$t - 1 \geq 2n(\lambda^2 + 2\lambda - 2) + 1.$$

Proof. In G a vertex is non-adjacent to $2n(1 - \lambda) - 1$ other vertices. Yet in G_t a vertex in $D(G_t)$ is non-adjacent to all the vertices in the other odd components of $D(G_t)$, of which there are at least

$$\begin{aligned} s + 1 &\geq \delta + 1 \\ &\geq 2n(\lambda^2 + \lambda - 1) + 1. \end{aligned}$$

It follows that for a vertex $v \in D(G_t)$ at least

$$2n(\lambda^2 + \lambda - 1) + 1 - (2n(1 - \lambda) - 1) = 2n(\lambda^2 + 2\lambda - 2) + 2$$

edges incident with v in G must be present in the matchings $M_0, F_1, \ldots, F_{t-1}$. So we have

$$\begin{aligned} t - 1 &\geq 2n(\lambda^2 + 2\lambda - 2) + 1 \\ &> 2n(\lambda^2 + 2\lambda - 2). \end{aligned}$$

□

Lemma 3.21. *A vertex in $A(G_t)$ must have fewer than n neighbours in $D(G_t)$.*

Proof. First note that for any $X, Y \in \mathbb{R}$ we have

$$\min\{X, Y - X\} \leq \tfrac{1}{2}Y. \tag{3.7}$$

Let v be a vertex in $A(G_t)$. The maximum number of neighbours that v can have in $D(G_t)$ is clearly bounded above by both Δ and the size of $D(G_t)$ which is at most $2n - w - s - 2$. So the maximum number of neighbours is at most

$$\begin{aligned} \min\{\Delta, 2n - w - s - 2\} &\leq \min\{\delta + w + 2, 2n - w - \delta - 2\} \\ &\leq \min\{\delta + w + 2, 2n - (\delta + w + 2)\} \\ &\leq n. \end{aligned}$$

The first inequality comes from Lemma 3.19 and the last inequality follows from (3.7). □

We will now define three quantities, β, θ and ψ. Let β be the number of inflexible vertices in $A(G_t)$. Let θ be the number of edges in the matchings $F_1, \ldots, F_{t-1}$ that have both ends in $A(G_t)$. And let ψ be the number of non-marginal edges in the matchings $F_1, \ldots, F_{t-1}$ that have both ends in $A(G_t)$. We will now obtain bounds for these quantities.

Lemma 3.22.

$$\beta > 2n\frac{2(\lambda^2+\lambda-1)(3\lambda^2+3\lambda-4)}{4\lambda^2+4\lambda-5}.$$

Proof. Let β be the number of inflexible vertices in $A(G_t)$. By Lemma 3.21 the maximum number of edges joining such vertices to their neighbours in $D(G_t)$ is at most $n\beta$. So we have

$$n\beta + (s-\beta)(2n-2\delta) \geq (s+2)\delta.$$

So we have

$$\begin{aligned}\beta &\geq \frac{s(3\delta-2n)+2\delta}{2\delta-n} \\ &> \frac{\delta(3\delta-2n)}{2\delta-n}.\end{aligned}$$

The result then follows by Lemma 3.12. □

Remark 3.23. *There are at most* $2(s-\beta)$ *vertices in* $D(G_t)$ *that are contained in a nearly-tight neighbourhood of a set of flexible vertices in* $A(G_t)$.

Proof. There are at most $s-\beta$ flexible vertices in $A(G_t)$. The number of vertices in $A(G_t)$ that are contained in nearly-tight neighbourhoods of these flexible vertices is clearly bounded above by $2(s-\beta)$. (This bound is attained when each flexible vertex is adjacent to exactly two vertices in $D(G_t)$.) □

Lemma 3.24.

$$\theta > (2n)^2\frac{(\lambda^2+\lambda-1)(3\lambda^2+4\lambda-5)}{2}.$$

Proof. Let $v \in V(G) - x - y$. The number of edges in $F_1, \dots, F_{t-1}$ that are incident with v is $t-1$. The number of these edges that go to $C(G_t) \cup D(G_t)$

is bounded above by the number of vertices in $C(G_t) \cup D(G_t)$ which is at most $2n - \delta - 2$. And at most one of these edges goes to $\{x, y\}$. However, by Lemma 2.18 there are at least $\delta(s+2)$ edges in G_t that join $A(G_t)$ to $D(G_t)$. So we have

$$\begin{aligned}
\theta &\geq \tfrac{1}{2}(s(t-1) - (s(2n-\delta-1) - \delta(s+2))) \\
&> \tfrac{1}{2}(s(t-1) - s(2n-2\delta-1)) \\
&\geq \tfrac{1}{2}(s(2n(\lambda^2+2\lambda-2)) - s(2n-2n(2\lambda^2+2\lambda-2)-1)) \\
&\geq \tfrac{1}{2}s(2n(3\lambda^2+4\lambda-5)+1) \\
&> \tfrac{1}{2}(2n)^2(\lambda^2+\lambda-1)(3\lambda^2+4\lambda-5).
\end{aligned}$$

□

Lemma 3.25.

$$\psi > (2n)^2 \frac{2(\lambda^2+\lambda-1)(3\lambda^2+4\lambda-5) - (1-\lambda^2)^2}{4}.$$

Proof. By Lemma 3.11 we have $q < 2n(1-\lambda^2) - 3$. By Remark 3.3 the number of marginal edges is at most

$$\tfrac{1}{4}q^2 < \tfrac{1}{4}(2n)^2(1-\lambda^2)^2.$$

The result then follows by Lemma 3.24. □

Suppose that F_i, where $i \in \{1, \ldots, t-1\}$, is the matching that contains the most non-marginal edges that have both ends in $A(G_t)$, and let ϕ be the number of such edges that F_i contains.

Lemma 3.26.

$$\phi > 2n \frac{2(\lambda^2+\lambda-1)(3\lambda^2+4\lambda-5) - (1-\lambda^2)^2}{4(1-\lambda^2)}.$$

Proof. There are at least ψ non-marginal edges in the matchings $F_1, \ldots, F_{t-1}$ that have both ends in $A(G_t)$. So there must exist a matching F_i, where $i \in \{1, \ldots, t-1\}$, that contributes at least $\psi/(t-1)$ edges to this total. The result follows from Lemma 3.25 and the fact that $t - 1 < 2n(1 - \lambda^2)$. $\square$

Lemma 3.27. *Suppose that*

$$2\phi - 2n + 2\delta + 4 > 0. \tag{3.8}$$

Then for any pair of vertices a, b in different odd components of $D(G_t)$ there is an edge $cd \in F_i$ such that $c, d \in A(G_t)$, and ac, bd are edges in G_t.

Proof. Suppose that a, b are two vertices in different odd components of $D(G_t)$, and that the sizes of the odd components they are contained in are σ_a and σ_b. Since there are at least $s + 2$ odd components, we have

$$\sigma_a + \sigma_b - 2 \leq 2n - 2s - 4. \tag{3.9}$$

The vertex a is adjacent to at least $\delta - \sigma_a + 1$ vertices in $A(G_t)$, and b is adjacent to at least $\delta - \sigma_b + 1$ such vertices. There are more than 2ϕ vertices in $A(G_t)$ incident with a non-marginal edge of F_i. So a must be adjacent to at least

$$2\phi - s + (\delta - \sigma_a + 1)$$

of them. There is an equinumerous set of vertices that are joined to these vertices by an edge of F_i. Now b must be adjacent to at least

$$2\phi - s + (\delta - \sigma_a + 1) - s + (\delta - \sigma_a + 1) = 2\phi - 2s + 2\delta - (\sigma_a + \sigma_b - 2).$$

And by (3.9) this is at least

$$2\phi - 2n + 2\delta + 4.$$

It follows that if this number is positive then the required vertices c, d exist. $\square$

Lemma 3.28. *Suppose that*

$$2\phi - 2n + 2\beta + 4 > 0. \tag{3.10}$$

Then F_i contains an edge ab, such that the vertices a, b are in different odd components of $D(G_t)$ and are not contained in a nearly-tight neighbourhood of a set of flexible vertices.

Proof. By Remark 3.23, there are at most $2(s - \beta)$ vertices in $D(G_t)$ that are contained in a nearly-tight neighbourhood of a set of flexible vertices in $A(G_t)$. So there can be at most $s - \beta$ edges in F_i that have both ends in a nearly-tight neighbourhood of a set of flexible vertices.

Since each component of $D(G_t)$ is of odd order, each component must contain at least one vertex that is incident with an edge of F_i, the other end of which is not in the component. Two of these edges may go to x and y; at most $s - 2\phi$ go to vertices in $A(G_t)$; and at most $2n - 2s - 4$ go to vertices in $C(G_t)$. So the number of such edges that go to a different component of $D(G_t)$ is at least

$$\tfrac{1}{2}((s+2) - 2 - (s - 2\phi) - (2n - 2s - 4)) = \tfrac{1}{2}(2\phi - 2n + 2s + 4).$$

So at least

$$\tfrac{1}{2}(2\phi - 2n + 2s + 4) - (s - \beta)$$

of these have at least one end not in a nearly-tight neighbourhood of a set of flexible vertices. It follows that if this number is positive then the required edge ab exists. $\square$

Proof of Lemma 3.14. We wish to show that for λ sufficiently large, conditions (3.8) and (3.10) are satisfied. By using the bounds obtained in Lemmas 3.12, 3.26 and 3.22, we obtain the inequalities

$$\frac{\lambda^4 + 10\lambda^3 + 4\lambda^2 - 14\lambda + 3}{2(1-\lambda^2)} > 0,$$

and

$$\frac{4\lambda^6 - 28\lambda^5 - 71\lambda^4 + 54\lambda^3 + 88\lambda^2 - 62\lambda + 3}{2(\lambda^2-1)(4\lambda^2+4\lambda-5)} > 0.$$

Both of these are satisfied if λ is greater than the second largest root of

$$4x^6 - 28x^5 - 71x^4 + 54x^3 + 88x^2 - 62x + 3$$

(≈ 0.81112). So by Lemma 3.28 we can find two vertices a, b in different odd components of $D(G_t)$ that are not contained in a nearly-tight neighbourhood of a set of flexible vertices. And by Lemma 3.27 F_i contains a non-marginal edge cd such that $c, d \in A(G_t)$ and ac, bd are edges in G_t. Hence (a, b, c, d) is a critical type-2 quadrilateral. $\square$

3.9 The Overfull Conjecture

In this section our graphs will be of both odd and even order. Accordingly, they will be of order n not $2n$. A graph G is said to be *overfull* if it contains a subgraph H such that H is of odd order at least 3, $\Delta(G) = \Delta(H)$ and

$$|E(H)| > \frac{|V(H)| - 1}{2}\Delta(H). \tag{3.11}$$

An overfull graph is necessarily class 2, as more than $\Delta(G)$ colours are needed to colour the edges of H. Regular graphs of odd order are always overfull, since (3.11) is satisfied by taking $H = G$. (Note that a graph G is overfull if and only if its fractional chromatic index is $\Delta(G) + 1$. See e.g. [35].)

In 1986 Chetwynd and Hilton made the following conjecture which came to be known as the *Overfull Conjecture*.

Conjecture 3.29. (Chetwynd and Hilton [9]) *Let G be a graph of order n where*

$$\Delta(G) > \tfrac{1}{3}n.$$

Then G is class 2 if and only if G is overfull.

In the case of regular graphs, overfullness has a simple interpretation.

Lemma 3.30. *Let G be a regular graph of degree d. Then G is overfull if and only if there is a set $S \subseteq V(G)$ of odd size such that there are fewer than d edges between S and $V(G) - S$.*

Proof. Suppose G is overfull. Then there is a set of vertices S, of odd size k, such that the subgraph of G induced by S has over $d(k-1)/2$ edges. Since the number of ends incident with vertices in S is dk, there are fewer than $dk - d(k-1) = d$ edges between S and $V(G) - S$.

Conversely, suppose there is a set $S \subseteq V(G)$, of odd size k, such that there are fewer than d edges between S and $V(G) - S$. Then there are more than $(kd - d)/2 = d(k-1)/2$ edges in the subgraph of G induced by S and so G is overfull. □

The following lemma shows that Conjecture 3.29 is implied by Conjecture 1.1.

Lemma 3.31. *Let G be a regular graph of even order n and degree at least $n/2$. Then G is not overfull.*

Proof. Suppose G is overfull. Then by Lemma 3.30 there is a set S, of odd size k, such that there are fewer than d edges between S and $V(G) - S$. Note that if a set S satisfies this property, so does its complement. So we may assume that $k \leq n/2$.

The maximum number of edges that the subgraph induced by S can contain is $k(k-1)/2$. So the number of edges between S and $V(G) - S$ is at least

$kd - k(k-1) = k(d-k+1)$. This is a polynomial in k which takes the value d when $k = 1$ and $k = d$ and exceeds this value in between. So we must have $k > d$. But we have assumed that $k \leq n/2 \leq d$, so we have a contradiction. Hence G is not overfull. $\square$

In what may be the only partial result on the Overfull Conjecture, Plantholt proved the following.

Lemma 3.32. (Plantholt [34]) *Suppose $\lambda \geq \frac{3}{4}$ is a real number such that all regular graphs of order n and degree at least $n\lambda$ are 1-factorizable. Let G be a graph of order n. If*

$$\frac{3\delta(G) - \Delta(G)}{2} \geq n\lambda,$$

then G is class 2 if and only if G is overfull.

Combining Lemma 3.32 with Theorem 1.4, we have the following.

Corollary 3.33. *Let G be a graph of order $2n$. If*

$$3\delta(G) - \Delta(G) \geq n(\sqrt{7} - 1),$$

then G is class 2 if and only if G is overfull.

And using the trivial bound $\Delta(G) < n$ we have:

Theorem 3.34. (Plantholt [34]) *Let G be a graph of order n and minimum degree $n\mu$ where*

$$\mu > \frac{\sqrt{7}}{3} \approx 0.88192,$$

then G is class 2 if and only if G is overfull.

If instead of using Theorem 1.4 and use our new Theorem 1.5, we obtain the following.

Theorem 3.35. *Let λ^* be the second largest root of*

$$4x^6 - 28x^5 - 71x^4 + 54x^3 + 88x^2 - 62x + 3.$$

Let G be a graph of order n and minimum degree $n\mu$ where

$$\mu > \frac{2\lambda^* + 1}{3} \approx 0.87408,$$

then G is class 2 if and only if G is overfull.

Chapter 4

Partial latin squares and Hall's Condition

In this chapter we present our investigation into partial latin squares that satisfy Hall's Condition. Sections 4.1 and 4.2 provide an introduction to Hall's Condition. In Section 4.3 we prove Theorem 1.6, and in Section 4.4 we prove the more general Theorem 1.7. Section 4.5 contains some results that will be needed in Section 4.6, where we show that the problem of determining the completability of partial latin squares that satisfy Hall's Condition is NP-hard. Finally in Section 4.7 we give some variations on this result.

Although Hall's Condition was first investigated in connection with the problem of determining completability of partial latin squares [27], most of the recent work is concerned with how it applies to list-colourings of graphs, so we introduce it here via list-colourings. For a recent survey on Hall's Condition see [28].

4.1 List-colourings

Let $\mathcal{A} = \{A_1, \ldots, A_n\}$ be a finite collection of finite sets. A *system of distinct representatives* for $\mathcal{A}$ is a set $\{a_1, \ldots, a_n\}$ of n distinct elements such that $a_i \in A_i$ for all $i \in \{1, \ldots, n\}$. Theorem 2.10 (Hall's Theorem) can be restated in the following way. (In fact, this is how Hall originally stated his theorem.)

Theorem 4.1. (Hall, 1936 [26, 1]) *$\mathcal{A}$ has a system of distinct representatives if and only if for all subsets $I \subseteq \{1, \ldots, n\}$ we have*

$$\left| \bigcup_{i \in I} A_i \right| \geq |I| \,. \tag{4.1}$$

Let G be a graph, and let L be a map that assigns to each vertex $v \in V(G)$ a set $L(v)$ of positive integers. We refer to $L(v)$ as the *list* of v. We say that G is L-colourable if there is a map c from $V(G)$ to the positive integers such that $c(v) \in L(v)$ for all $v \in V(G)$ and for all pairs of adjacent vertices u, v we have $c(u) \neq c(v)$. If $\sigma \in L(v)$ then we say that v *supports* σ.

Let (G, L) be a graph with lists. We say that (G, L) satisfies *Hall's Condition* if for all induced subgraphs H of G we have

$$\sum_{\sigma \in \mathbb{N}} \alpha(L, \sigma, H) \geq |V(H)| \tag{4.2}$$

where $\alpha(L, \sigma, H)$ is the size of a maximum independent set in the subgraph of H induced by those vertices that support σ. Despite the name, Hall's Condition is actually a set of $2^{|V(G)|}$ inequalities, one for each induced subgraph of G. If H is an induced subgraph of G then we refer to the inequality involving H as *Hall's Inequality on H*, which we will abbreviate to $\mathcal{H}(H)$.

Is is not hard to see that Hall's Condition is a necessary condition for a graph to have a list-colouring:

Lemma 4.2. *Let (G, L) be a graph with lists. Then (G, L) is list-colourable only if (G, L) satisfies Hall's Condition.*

Proof. Suppose we have a list-colouring of (G, L). Let H be an induced subgraph of G. For $\sigma \in \mathbb{N}$, let V_σ be the set of vertices in H coloured σ. We must have

$\sum_{\sigma \in \mathbb{N}} |V_\sigma| = |H|$. As each set V_σ is an independent set of vertices in H, we have $\alpha(L, \sigma, H) \geq |V_\sigma|$. Therefore

$$\sum_{\sigma \in \mathbb{N}} \alpha(L, \sigma, H) \geq \sum_{\sigma \in \mathbb{N}} |V_\sigma| = |V(H)|$$

and so $\mathcal{H}(H)$ holds. As this holds for all induced subgraphs H, (G, L) satisfies Hall's Condition. $\square$

For certain sorts of graphs with lists, Hall's Condition is both a necessary and a sufficient condition for list-colourability. For example, we have the following.

Lemma 4.3. *A complete graph K_n with lists has a list-colouring if and only if Hall's Condition is satisfied.*

Proof. Suppose the vertices are labelled $v_1, \dots, v_n$. Since any two vertices are adjacent, finding a list-colouring is equivalent to finding a system of distinct representatives for $L(v_1), \dots, L(v_n)$. And as every induced subgraph of K_n is a complete graph, condition (4.2) implies that for every set of vertices J, $\left|\bigcup_{v \in J} L(v)\right| \geq |J|$. So in this case condition (4.2) is equivalent to condition (4.1). $\square$

However, it is not difficult to find examples of graphs with lists that satisfy Hall's Condition but are not list-colourable—see Figure 4.1 for one example. (In this graph every set of vertices of size 3 is list-colourable, so to verify Hall's Condition we just need to check Hall's Inequality for the whole graph, and indeed it holds. However the graph is clearly not list-colourable.)

4.2 Partial latin squares

A partial latin square P gives rise to a graph with lists (G, L) in the following natural way: create a vertex for each cell, join two vertices by an edge if and only if their corresponding cells are in the same row or column, and for each vertex v in G let $L(v)$ be the set of symbols supported by the cell corresponding to v.

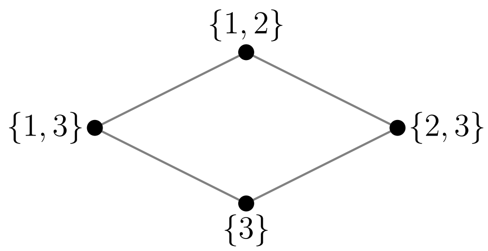


Figure 4.1: A graph with lists that satisfies Hall's Condition but is not list-colourable.

(So if the cell contains the symbol σ, $L(v) = \{\sigma\}$, and if the cell is empty, $L(v)$ is the set of symbols that do not occur in the same row or column.) We refer to the graph with lists (G, L) as $L(P)$.

Note that the graph of $L(P)$ is sometimes called the *cartesian product* $K_n \square K_n$. (For two graphs G_1 and G_2, the cartesian product $G_1 \square G_2$ is the graph with vertex set $V(G_1) \times V(G_2)$, where (u_1, v_1) is adjacent to (u_2, v_2) if either $u_1u_2 \in E(G_1)$ and $v_1 = v_2$, or $v_1v_2 \in E(G_2)$ and $u_1 = u_2$. See e.g. [4].)

It should be clear that a vertex of $L(P)$ supports a symbol σ if and only if the corresponding cell of P supports σ, and that a set of vertices of $L(P)$ is independent if and only if the corresponding cells of P are independent. So we can see that not only is a list-colouring of $L(P)$ entirely equivalent to a completion of P, but that Hall's Condition for P—as defined in Chapter 1—is the same as Hall's Condition for $L(P)$. So we can use both definitions of Hall's Condition interchangeably.

If we have a set of cells T of a partial latin square P, by $\mathcal{H}(T)$ we mean the Hall's Inequality on the set of vertices of $L(P)$ that correspond to the cells of T.

Lemma 4.4. *Let P be a partial latin square of order n, T a set of cells and $f \in T$ a filled cell. Then $\mathcal{H}(T)$ holds if and only if $\mathcal{H}(T - f)$ holds.*

Proof. Assume that $\mathcal{H}(T)$ holds. As the cell f only supports one symbol, it can only contribute 1 to the quantity

$$\sum_{\sigma \in \{1,\ldots,n\}} \alpha(\sigma, T),$$

so

$$\sum_{\sigma \in \{1,\ldots,n\}} \alpha(\sigma, T - f) \geq \left(\sum_{\sigma \in \{1,\ldots,n\}} \alpha(\sigma, T) \right) - 1 \geq |T| - 1 = |T - f|\,,$$

and so $\mathcal{H}(T - f)$ holds. Conversely, suppose that $\mathcal{H}(T - f)$ holds, and that f contains a symbol $\sigma \in \{1, \ldots, n\}$. As no other cells of T in the same row or column as f support σ, the size of a maximum independent set of cells of T that support σ is exactly one greater than the size of such a set in $T - f$. So

$$\sum_{\sigma \in \{1,\ldots,n\}} \alpha(\sigma, T) = \left(\sum_{\sigma \in \{1,\ldots,n\}} \alpha(\sigma, T - f) \right) + 1 \geq |T - f| + 1 = |T|\,,$$

and so $\mathcal{H}(T)$ holds. □

A consequence of Lemma 4.4 is that when determining if a partial latin square satisfies Hall's Condition it is sufficient to verify that Hall's Inequality is satisfied for each set of empty cells.

Recall Goldwasser's square from Chapter 1 (see Figure 1.1). It can easily be verified that Goldwasser's square is incompletable. To see that it satisfies Hall's Condition, we will use the following lemma, which will also be needed later. Note that we say that a symbol is *missing* from a row or column if it does not appear in the filled cells of that row or column.

A *vertex cover* of a graph G is a set of vertices C, such that each edge of G has at least one end in C. We shall need the following classical theorem:

Theorem 4.5. (König–Egevary, 1931 [1, 4, 14]) *Let G be a bipartite graph. The size of a maximum matching in G is equal to the size of a minimum vertex cover.*

It can be easily verified that Goldwasser's square (see Figure 1.1) is incompletable. To see that it satisfies Hall's Condition, we shall use the following lemma, which will also be needed in Section 4.6.

Lemma 4.6. *Let P be a partial latin square. Suppose that a symbol σ is missing from k columns and k rows of P but is present in all the other rows and columns of P. Suppose that we have a set of empty cells T which contains t cells that support σ. Then $\alpha(\sigma, T) \geq \lceil t/k \rceil$.*

Proof. Let G be the bipartite graph on $2k$ vertices, defined as follows. There are k vertices $r_1, \ldots, r_k$, representing the rows from which σ is missing, and there are k vertices $c_1, \ldots, c_k$, representing the columns from which σ is missing. For each cell of T that supports σ, we place an edge between the vertex that represents the cell's row, and the vertex that represents the cell's column. Matchings in G correspond to independent sets of cells in P. By Theorem 4.5, the size of a maximum matching in G is equal to the size of the smallest vertex cover. But G has t edges and maximum degree at most k, so any vertex cover must contain at least $\lceil t/k \rceil$ vertices. Hence there is a matching in G of size $\lceil t/k \rceil$, and therefore $\alpha(\sigma, T) \geq \lceil t/k \rceil$. □

In the preceding proof we found that determining the size of a maximum independent set in a set of cells is equivalent to finding a maximum matching in a certain bipartite graph. Since the size of a maximum matching in a bipartite graph on n vertices can be determined in $O(n^3)$ time (see e.g. [38, Chapter 20]), we have the following result:

Lemma 4.7. *Let P be a partial latin square, and T a set of cells. Then $\mathcal{H}(T)$ can be determined in $O(n^4)$ time.*

In Goldwasser's square, all symbols are missing from 2 rows and 2 columns, and each empty cell supports 2 symbols. By Lemma 4.4, Hall's Condition holds if Hall's Inequality holds for each set of empty cells. So let T be a set of empty cells of Goldwasser's square. For each $i \in \{1, \ldots, 6\}$, let a_i be the number of cells of T that support i. As each cell of T supports 2 symbols we have

$$\sum_{i=1}^{6} a_i = 2\,|T|\,. \tag{4.3}$$

But then

$$\begin{aligned}
\sum_{i=1}^{6} \alpha(i,T) &\geq \sum_{i=1}^{6} \lceil a_i/2 \rceil && \text{(by Lemma 4.6)} \\
&\geq \tfrac{1}{2} \sum_{i=1}^{6} a_i \\
&= |T|\,. && \text{(by (4.3))}
\end{aligned}$$

And so $\mathcal{H}(T)$ holds. So Hall's Inequality holds for each set of empty cells T, and so Goldwasser's square satisfies Hall's Condition.

4.3 Hall's Condition and Ryser's condition

The following is a classical theorem of Ryser, which gives a necessary and sufficient condition for a latin rectangle to be extendable to a latin square.

Theorem 4.8. (Ryser, 1951 [37, 1]) *Let P be a partial latin square of order n whose filled cells are those in the upper-left $r \times s$ rectangle R, for some $r, s \in \{1, \ldots, n\}$. P is completable if and only if $\nu(\sigma) \geq r + s - n$ for all $\sigma \in \{1, \ldots, n\}$, where $\nu(\sigma)$ is the number of times that σ occurs in R.*

Hilton and Johnson [27] proved that Ryser's condition is equivalent to Hall's Condition—in fact, it is equivalent to a single Hall's Inequality. This theorem provoked much of the current interest in Hall's Condition. We say that a symbol σ has an *r-independent set*, in a set of cells H, if there is an independent set of size r in the cells of H that support σ. To prove Theorem 1.6 we need the following lemma.

Lemma 4.9. *Let P be a partial latin square of order n whose filled cells are those in the upper-left $r \times s$ rectangle R, for some $r, s \in \{1, \ldots, n\}$. Let H be the set of cells in the top r rows, and for each $\sigma \in \{1, \ldots, n\}$ let $\nu(\sigma)$ denote the number of times that σ appears in R. Then*

$$\alpha(\sigma, H) = \min\{r,\ \nu(\sigma) + n - s\}.$$

Proof. Let S_1 be the set of cells in R that contain σ. S_1 is an independent set of size $\nu(\sigma)$. In H there are $n - s$ columns with empty cells, and $r - \nu(\sigma)$ rows that do not contain σ. From the cells in these $n - s$ columns and $r - \nu(\sigma)$ rows we can select an independent set S_2 of size $\min\{r - \nu(\sigma),\ n - s\}$. Then $S = S_1 \cup S_2$ is an independent set of size $\min\{r,\ \nu(\sigma) + n - s\}$. This shows that $\alpha(\sigma, H) \geq \min\{r,\ \nu(\sigma) + n - s\}$. Moreover, $\alpha(\sigma, H) \leq r$ as there are r rows in H; also $\alpha(\sigma, H) \leq \nu(\sigma) + n - s$ as no independent set for σ can use more than $\nu(\sigma) + n - s$ columns. So in fact we have $\alpha(\sigma, H) = \min\{r,\ \nu(\sigma) + n - s\}$. $\square$

Proof of Theorem 1.6. If P is completable then it satisfies Hall's Condition by Lemma 4.2. Conversely, suppose P satisfies Hall's Condition. Then $\mathcal{H}(H)$ holds, which means that

$$\sum_{\sigma \in \{1,\ldots,n\}} \alpha(\sigma, H) \geq rn,$$

which in turn implies that $\alpha(\sigma, H) = r$ for all $\sigma \in \{1, \ldots, n\}$. By Lemma 4.9 we have $\nu(\sigma) + n - s \geq r$ for all $\sigma \in \{1, \ldots, n\}$. Hence P is completable by Theorem 4.8 (Ryser's Theorem). $\square$

4.4 A more general result

In this section we will present the proof of Theorem 1.7, the most general "positive" result on Hall's Condition that we have been able to obtain. The proof makes

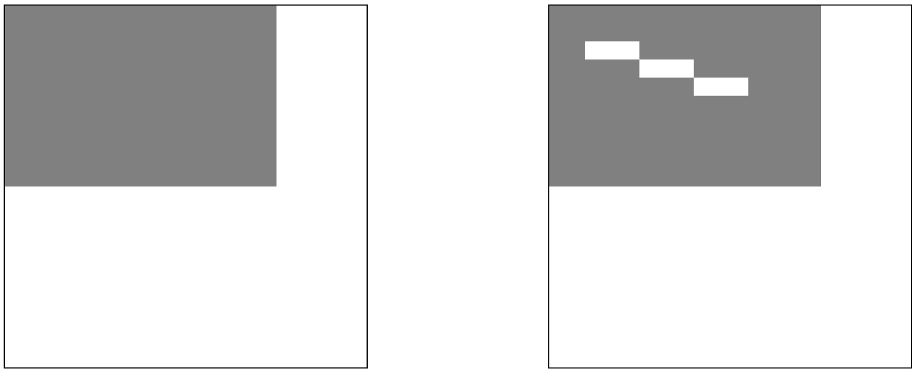


Figure 4.2: The shapes of partial latin square considered in Theorems 1.6 and 1.7.

use of the following classical result of Ford and Fulkerson, commonly known as the "Max-flow Min-cut" Theorem. For basic definitions concerning flows and cuts see [4]. Note that an *integral flow* is a flow in which the flow along each edge is an integer.

Theorem 4.10. (Ford and Fulkerson, 1956 [19, 4, 14]) *Let G be a directed graph with integral edge capacities and two distinguished vertices α (the "source") and ω (the "sink"). Then the size of a maximum flow between α and ω is equal to the minimum size of a cut that separates these two vertices. Moreover, there is a maximum flow that is integral.*

Theorem 4.10 is in a sense more fundamental than Theorem 2.10 (Hall's Theorem). In fact, it is a common exercise to deduce Hall's Theorem from it. In the proof of Theorem 1.7 we will use a technique inspired by this exercise. The following lemma is a generalization of Lemma 4.9.

Lemma 4.11. *Let P be a partial latin square of order n whose filled cells are all in the upper-left $r \times s$ rectangle R, for some $r, s \in \{1, \ldots, n\}$, although at most one cell in each column inside the rectangle may be empty. Let J be a subset of these empty cells. Let H be the set of cells in the top r rows, and for each $\sigma \in \{1, \ldots, n\}$ let $\nu(\sigma)$ denote the number of times that σ appears in R, and let $\rho(\sigma)$ be the number of rows in which there is an empty cell in $R - J$ that supports*

σ. *Then*

$$\alpha(\sigma, H - J) = \min\{r,\ \nu(\sigma) + \rho(\sigma) + n - s\}.$$

Proof. Let S_1 be the set of cells in H that contain σ, plus one empty cell from $R-J$ that supports σ from each row that contains such a cell. S_1 is an independent set of size $\nu(\sigma) + \rho(\sigma)$. We can find an independent set S_2 of size $\min\{r - \nu(\sigma) - \rho(\sigma),\ n - s\}$ from the cells in the rightmost $n - s$ columns that are in the $r - \nu(\sigma) - \rho(\sigma)$ rows that have no cells in S_1. Then $S = S_1 \cup S_2$ is an independent set for σ of size $\min\{r,\ \nu(\sigma) + \rho(\sigma) + n - s\}$. This shows that $\alpha(\sigma, H - J) \geq \min\{r,\ \nu(\sigma) + \rho(\sigma) + n - s\}$. Moreover, $\alpha(\sigma, H - J) \leq r$ as there are r rows in H; also $\alpha(\sigma, H - J) \leq \nu(\sigma) + \rho(\sigma) + n - s$ as no independent set for σ can use more than $\nu(\sigma) + \rho(\sigma) + n - s$ columns. So in fact we have $\alpha(\sigma, H - J) = \min\{r,\ \nu(\sigma) + \rho(\sigma) + n - s\}$. □

Proof of Theorem 1.7. If P is completable then it satisfies Hall's Condition by Lemma 4.2. Conversely, suppose P satisfies Hall's Condition. We will show that P can be completed to a latin square of order n. We will do this by showing that we can fill all the empty cells in R, in such a way that Ryser's condition given in Theorem 4.8 is satisfied.

Let H be the set of rn cells in the first r rows of P, and let $B = \{b_1, \ldots, b_t\}$ be the set of empty cells in R. By Lemma 4.11, first with $J = B$, and second with $J = \emptyset$, we have

$$\alpha(\sigma, H - B) = \min\{r,\ \nu(\sigma) + n - s\}$$

and

$$\alpha(\sigma, H) = \min\{r,\ \nu(\sigma) + \rho(\sigma) + n - s\},$$

where $\rho(\sigma)$ is the number of rows in which σ is supported by an empty cell in R.

$\mathcal{H}(H)$ implies that for all $\sigma \in \{1, \ldots, n\}$ we have $\alpha(\sigma, H) = r$ and therefore

$$\nu(\sigma) + \rho(\sigma) + n - s \geq r.$$

For each symbol $\sigma \in \{1, \ldots, n\}$ we let $\mu(\sigma)$ be the least non-negative integer such that

$$\nu(\sigma) + \mu(\sigma) \geq r + s - n. \tag{4.4}$$

So we have

$$\mu(\sigma) \leq \rho(\sigma). \tag{4.5}$$

Note that if σ already appears at least $r+s-n$ times in R we have $\mu(\sigma) = 0$; otherwise we have $\mu(\sigma) > 0$ and

$$\nu(\sigma) + \mu(\sigma) = r + s - n. \tag{4.6}$$

The proof will be in four stages. Let

$$u = \sum_{\sigma \in \{1,\ldots,n\}} \mu(\sigma).$$

In Stage 1, we show that we can create a partial latin square Q_1 containing P by filling in u cells of R in such a way that each symbol $\sigma \in \{1, \ldots, n\}$ appears $\mu(\sigma)$ times in the empty cells of R. It will follow from (4.4) that Q_1 satisfies Ryser's condition of Theorem 4.8, namely that each symbol appears at least $r+s-n$ times in R. However, Q_1 may still have some empty cells in the upper-left $r \times s$ rectangle R. In Stage 1 we will use the Hall's Inequality $\mathcal{H}(H - B')$ for each subset $B' \subseteq B$, and our main tool will be the Max-flow Min-cut Theorem (Theorem 4.10).

Stage 2 is independent of Stage 1. We create a second partial latin square Q_2 containing P. Q_2 will not depend in any way on Q_1. In the upper-left $r \times s$ rectangle R of Q_2 there will be no empty cells. Thus R will be completely filled, but Q_2 will not necessarily satisfy Ryser's condition of Theorem 4.8. In Stage 2 we will use Hall's Inequality $\mathcal{H}(B')$ for each subset $B' \subseteq B$ whose cells all lie in just one row.

In Stage 3 we use the Dulmage-Mendelsohn Theorem (Theorem 2.12) to produce from Q_1 and Q_2 a partial latin square Q_3 that also contains P, and where the upper-left $r \times s$ rectangle R contains no empty cells, and which satisfies Ryser's condition of Theorem 4.8.

In Stage 4 we complete Q_3 using Theorem 4.8, and, since Q_3 contains P, in so doing we finish the completion of P.

Stage 1.

We will show that we can fill some of the cells of B to create a partial latin square Q_1. In the partial filling of B, each symbol $\sigma \in \{1, \ldots, n\}$ will be used $\mu(\sigma)$ times. If a symbol σ has $\mu(\sigma) = 0$ then there is nothing to do. We just need to consider those symbols for which $\mu(\sigma) > 0$. If this is the case, then σ satisfies (4.6).

In order to show that we can fill some of the cells of B using each symbol at least $\mu(\sigma)$ times, we will consider an edge-capacitated directed graph G. The graph G has vertex set $U \cup X \cup B \cup \{\alpha, \omega\}$ where

$$U = \{\sigma : \sigma \in \{1, \ldots, n\} \text{ and } \mu(\sigma) > 0\},$$

$$X = \{(\sigma, w) : w \in \{1, \ldots, r\} \text{ and at least one empty cell in row } w \text{ supports } \sigma\},$$

and α and ω are two additional vertices (the "source" and the "sink").

In G, α is joined to each vertex $\sigma \in U$ by an edge of capacity $\mu(\sigma)$, the direction being from α to σ. Each vertex $\sigma \in U$ is joined to all the vertices in X of the form (σ, w) for some $w \in \{1, \ldots, r\}$ by edges of capacity 1, the direction being from U to X. Each vertex $(\sigma, w) \in X$ is joined to all $b \in B$ where b is a cell in row w that supports σ, with edges of capacity u, the direction being from X to B. Each vertex in B is joined to ω by an edge of capacity 1, the direction

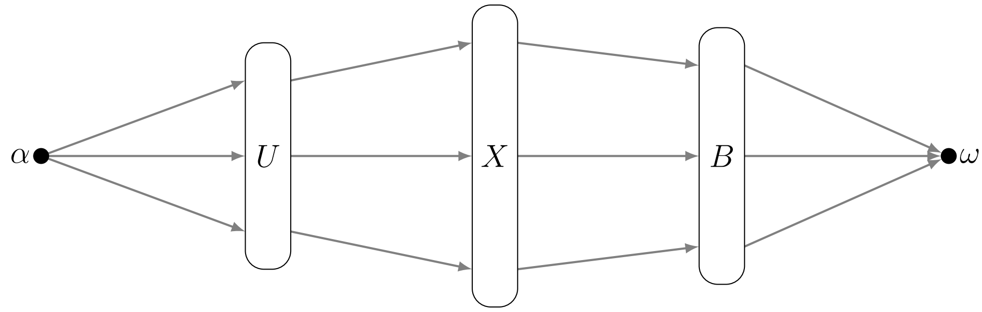


Figure 4.3: The directed graph G from Theorem 1.7.

being from B to ω. (See Figure 4.3.)

Claim 1. We can fill some of the cells of B using each symbol $\sigma \in U$ $\mu(\sigma)$ times if and only if there is a u-flow in G from α to ω.

Proof. Suppose there is such a partial filling of B. For each instance that a symbol $\sigma \in U$ is placed in cell $b \in B$ of row w, we can create a 1-flow in G from α to ω by sending a flow of 1 from α to $\sigma \in U$ to $(\sigma, w) \in X$ to $b \in B$ to ω. The sum of all these 1-flows gives a u-flow from α to ω.

Conversely, suppose that there is a u-flow from α to ω. By the Max-flow Min-cut Theorem (Theorem 4.10) there is such a flow that is integral. All the edges from α to U must carry a flow equal to their capacity, and there must be u paths from U to ω each carrying 1-flows. These flows indicate how the cells of B can be filled. The fact that the edges between U and X have capacity 1 ensures that each symbol is placed at most once in each row. This proves Claim 1.

Suppose, for a contradiction, that it is not possible to fill some of the cells of B using each symbol $\sigma \in U$ $\mu(\sigma)$ times. It follows that there does not exist a u-flow in G from α to ω. Hence by the Max-flow Min-cut Theorem (Theorem 4.10), there must be a cut T in G of size less than u that separates α from ω.

T cannot contain any of the edges between X and B as each of these edges has capacity u. Also T cannot contain *all* the edges between α and U (or it would have size u).

Let $U' \subseteq U$ be the set of vertices that α is joined to with an edge that is not in T. (See Figure 4.4.) For each $\sigma \in U'$ let $c(\sigma)$ be the number of edges joining σ to X that are in T. We may suppose that for all $\sigma \in U'$ we have $c(\sigma) < \mu(\sigma)$ as otherwise we can create a new cut T' from T, where $|T'| \leq |T|$, by adding the edge joining α to σ and removing the edges joining σ to X.

Let $B' \subseteq B$ be the set of vertices in B that are joined to vertices in U' by paths containing no edges of T. Because the cut T separates α from ω all the edges that join B' to ω must be in T. Since we have assumed that T has size less than u we must have

$$\sum_{\sigma \in U - U'} \mu(\sigma) + \sum_{\sigma \in U'} c(\sigma) + |B'| < u = \sum_{\sigma \in U} \mu(\sigma),$$

so that

$$\sum_{\sigma \in U'} (\mu(\sigma) - c(\sigma)) > |B'| \,. \tag{4.7}$$

By Lemma 4.11 we have for all $\sigma \in U$,

$$\begin{aligned} \alpha(\sigma, H - B') &= \min\{r,\ \nu(\sigma) + \rho'(\sigma) + n - s\} \\ &\leq \nu(\sigma) + \rho'(\sigma) + n - s, \end{aligned}$$

where $\rho'(\sigma)$ is the number of rows in which σ is supported by a cell in $B - B'$. Then by (4.6) we have

$$\alpha(\sigma, H - B') \leq r - \mu(\sigma) + \rho'(\sigma). \tag{4.8}$$

Consider a row w containing an empty cell $b \in B - B'$ supporting a symbol $\sigma \in U'$. This occurrence of σ in a cell of row w contributes 1 to $\rho'(\sigma)$. The occurrence of cell $b \in B - B'$ in row w supporting σ corresponds to a unique path joining $\sigma \in U'$ to $b \in B - B'$ and passing through (σ, w) in X. Such a path must contain one of the $c(\sigma)$ edges of T joining $\sigma \in U'$ to X. Therefore for $\sigma \in U'$ we

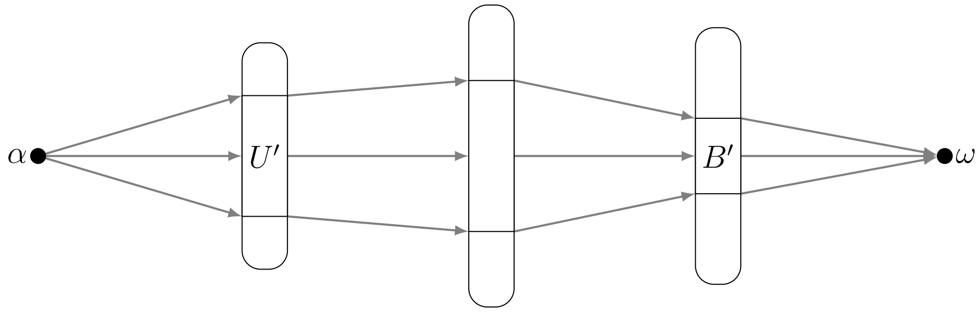


Figure 4.4: The sets U' and B' from Theorem 1.7.

have

$$\rho'(\sigma) \leq c(\sigma),$$

and therefore by (4.8),

$$\alpha(\sigma, H - B') \leq r - \mu(\sigma) + c(\sigma).$$

So

$$\begin{aligned}
\sum_{\sigma \in \{1,\ldots,n\}} \alpha(\sigma, H - B') &\leq \sum_{\sigma \in U - U'} r + \sum_{\sigma \in U'} (r - \mu(\sigma) + c(\sigma)) \\
&\leq rn - \sum_{\sigma \in U'} (\mu(\sigma) - c(\sigma)) \\
&< rn - |B'|, && \text{(by (4.7))}
\end{aligned}$$

which contradicts $\mathcal{H}(H - B')$.

Hence it is not possible that G has a cut of size less than u. It follows that the required partial latin square Q_1, which contains P, and in which each symbol appears at least $r + s - n$ times in R, can be found. This completes Stage 1.

Stage 2.

We must now show that we can find a partial latin square Q_2, that contains P, and has no empty cells within the upper-left $r \times s$ rectangle R. We will construct Q_2 by taking P and filling the cells in B one row at a time.

For each row $w \in \{1, \ldots, r\}$ that contains cells in B, consider the bipartite graph G_w on the vertex set $S \cup B_w$ where $S = \{1, \ldots, n\}$ and B_w is the subset of B consisting of the members of B that are in row w. In G_w each cell $b \in B_w$ is joined to the symbols in S that it supports. So a matching in G_w that covers B_w corresponds to a filling of the cells B_w with distinct symbols.

Claim 2. There is a matching in G_w that covers B_w.

Proof. Suppose not. Then by Hall's Theorem (Theorem 2.10) there is a subset $B' \subseteq B_w$ with neighbour set $Y \in S$ such that $|Y| < |B'|$. But for any symbol $\sigma \in \{1, \ldots, n\}$ we have $\alpha(\sigma, B') \leq 1$, so $\mathcal{H}(B')$ implies that the cells of B' support at least $|B'|$ symbols, which contradicts B' having fewer than $|B'|$ neighbours in G_w. This proves Claim 2.

It follows that the required partial latin square Q_2, which contains P, can be found, by filling the empty cells of R row by row.

Stage 3.

In Stage 3 we use Q_1 and Q_2 to construct a partial latin square Q_3 that contains P and whose filled cells are those in the upper-left $r \times s$ rectangle R, which has no empty cells and which satisfies Ryser's condition of Theorem 4.8.

For each row $w \in \{1, \ldots, r\}$ that contains cells in B, the way that the cells of B_w are filled in Q_1 and Q_2 give two matchings in G_w, M_1 and M_2. By the Dulmage-Mendelsohn Theorem (Theorem 2.12), there is a matching $M_3 \subseteq M_1 \cup M_2$ that covers all the vertices in B_w that are covered by M_2 (i.e. in fact, all the vertices of B_w), and all the vertices in S that are covered by M_1. In Q_3 we fill the cells of B in row w according to the matching M_3.

Once all the rows have been filled, there are no empty cells in R. And since the symbols in each row of Q_1 are a subset of those in the same row of Q_3, it follows that Q_3 satisfies the conditions of Ryser's Theorem (Theorem 4.8).

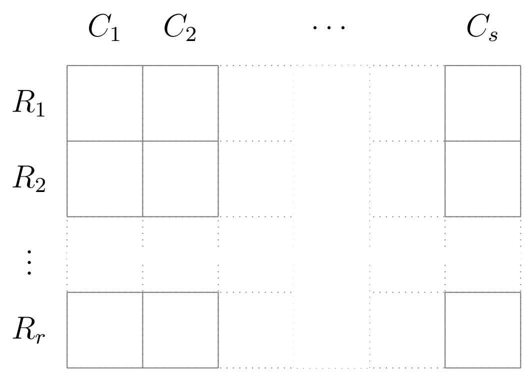


Figure 4.5: An empty framework.

Stage 4.

We apply Ryser's Theorem (Theorem 4.8) to complete Q_3 to a latin square of order n. As Q_3 contains P, we have completed P. This proves Theorem 1.7. $\square$

4.5 Frameworks

In what follows we will need the concept of a *framework*. Informally, a framework is an array equipped with row and column lists (or sets). They play an important role in the proofs by Colbourn [11] and Easton and Parker [17] that the problem of deciding if a partial latin square is completable is NP-complete.

Frameworks have also been studied independently of partial latin squares. Frieze [21] and Fon-Der-Flaass [18] showed that the problem of deciding if a framework is completable is NP-complete. A related concept is the *patterned hole*, used by Lindner and Rodger [30] to prove embedding theorems for cycle systems and other graph designs.

Formally, a *framework* $R = (r, s, t, P, R_1, \ldots, R_r, C_1, \ldots, C_s)$ consists of an $r \times s$ partial latin rectangle P on the symbols $1, \ldots, t$, together with row and column lists (i.e. row and column sets) $R_1, \ldots, R_r$ and $C_1, \ldots, C_s$. We require each list to be a subset of $\{1, \ldots, t\}$, and for all i, $1 \leq i \leq r$, R_i must not contain any symbol that occurs in row i of P, and for all j, $1 \leq j \leq s$, C_j must not contain any symbol that occurs in column j of P.

A framework R is said to be *balanced* if the following two conditions hold:

(i) for all i, $1 \le i \le r$, $|R_i|$ is equal to the number of empty cells in row i of P, and for all j, $1 \le j \le s$, $|C_j|$ is equal to the number of empty cells in column j of P; and

(ii) each of the symbols $\{1, \ldots, t\}$ occurs in the same number of row lists as column lists.

The *admissible symbol array* of R is an $r \times s$ array $A(R)$ in which $A(R)_{ij} = R_i \cap C_j$ if cell (i, j) of P is empty and $A(R)_{ij}$ is the singleton set containing the symbol in the cell (i, j) of P otherwise. Figure 4.5 is an illustration of a framework in which all the cells are empty.

A *completion* for R is an $r \times s$ latin rectangle P^* on the symbols $\{1, \ldots, t\}$ with the conditions that for all cells (i, j) $(1 \le i \le r,\ 1 \le j \le s)$, the following holds: if the cell (i, j) of P is filled then the cell (i, j) of P^* must contain the same symbol, and if the cell (i, j) of P is empty then the cell (i, j) of P^* must contain a symbol from $R_i \cap C_j$. Thus a completion for R is an $r \times s$ latin rectangle P^* such that the cell (i, j) of P^* contains a symbol from the set $A(R)_{ij}$ for all $1 \le i \le r$ and $1 \le j \le s$.

If a symbol σ belongs to $A(R)_{ij}$ we say that the cell (i, j) *supports* σ, and the set of cells that support σ is called the *support* of σ. Note that in a balanced framework, if a symbol appears in k row lists (and therefore k column lists) it is supported by k^2 empty cells. Indeed, the empty cells that support each symbol can be made into a square by a suitable permutation of the rows and columns.

A partial latin square Q of order n is said to *realize* the framework R if

(i) it contains P in its top-left corner,

(ii) the only empty cells of Q are those of P,

(iii) for all i, $1 \leq i \leq r$, the symbols missing from row i of Q are those in R_i, and (iv) for all j, $1 \leq j \leq s$, the symbols missing from column j of Q are those in C_j.

We now give the following theorem, which shows that any balanced framework can be realized by a partial latin square. It is similar to some proofs of Ryser's Theorem (see e.g. [1]).

Lemma 4.12. *Let $R = (r, s, t, P, R_1, \ldots, R_r, C_1, \ldots, C_s)$ be a balanced framework. If $n \geq \max\{t, r+s\}$, then R can be realized by a partial latin square Q of order n.*

Proof. Suppose $n \geq \max\{t, r+s\}$. Let Q' be a partial latin square of order n where all cells are empty except for an $r \times s$ array of cells in the top-left corner which is filled by P. We will describe a procedure for obtaining the required partial latin square Q by filling in all the cells of Q' that are not in the top-left $r \times s$ rectangle.

We do this in two stages. In the first stage, we will describe how the cells in the last $n - s$ columns of rows 1 to r can be filled. In the second stage, we describe how the cells in the bottom $n - r$ rows can be filled.

Consider a bipartite graph G_1 with bipartition (A, B), where the vertices in A are labelled $a_1, \ldots, a_n$ and correspond to the symbols $1, \ldots, n$, and the vertices in B are labelled $b_1, \ldots, b_r$ and correspond to the top r rows of Q'. For all i, j ($1 \leq i \leq n$, $1 \leq j \leq r$), G_1 has an edge between a_i and b_j if and only if the symbol i is not in R_j and does not occur in the jth row of P.

The vertex a_i has degree $r - \nu(i)$ where $\nu(i)$ is the number of times the symbol i occurs in the row lists plus the number of times it appears in P. The vertex b_j has degree $n - s$, because there are s symbols in the jth row of P and the row list R_j. The maximum degree $\Delta(G)$ is $n - s$ because otherwise we would have some i for which $r - \nu(i) > n - s$, which contradicts the assumption $n \geq r + s$.

By Theorem 2.11 (König's Theorem), G_1 has an edge-colouring using the colours $\{1, \ldots, n-s\}$. So the cells in the last $n-s$ columns of rows $1, \ldots, r$ of Q' can be filled according to this edge-colouring: symbol i is placed in cell $(j, s+k)$ if a_i is joined to b_j with an edge of colour k.

It remains to fill in the bottom $n-r$ rows of Q'. Consider a bipartite graph G_2 with bipartition (C, D), where the vertices in C are labelled $c_1, \ldots, c_n$ and correspond to the n columns of Q', and the vertices in D are labelled $d_1, \ldots, d_n$ and correspond to the symbols $1, \ldots, n$. For all i, j $(1 \leq i \leq n,\ 1 \leq j \leq n)$, G_2 has an edge between d_i and c_j if and only if i is not in C_j and does not occur in the jth column of P (for $1 \leq j \leq s$) or i does not appear in the top r cells of column j of Q' (for $s+1 \leq j \leq n$). Each vertex has degree $n-r$ so by Theorem 2.11 (König's Theorem), we can give G_2 an edge-colouring using the colours $\{1, \ldots, n-r\}$. We can now obtain Q from Q' by putting the symbol i in cell $(r+k, j)$ if d_i is joined to c_j with an edge of colour k.

This procedure creates a partial latin square Q, where the symbols missing from rows 1 to r are those in the lists $R_1, \ldots, R_r$ and the symbols missing from columns 1 to s are those in the lists $C_1, \ldots, C_s$. In other words, R is realized by Q. □

A framework R is said to be *symmetric* if P is a symmetric partial latin square, $r = s$, $R_i = C_i$ for all $1 \leq i \leq r$, and each symbol occurs an even number of times in the lists $R_1, \ldots, R_r$. Note that a symmetric framework necessarily satisfies condition (ii) of the definition of a balanced framework. A framework R is said to be *idempotent* if $r = s$, and the leading diagonal of P is filled with distinct symbols. The following lemma is based on a theorem of Cruse [13].

Lemma 4.13. *Let $R = (r, r, t, P, R_1, \ldots, R_r, R_1, \ldots, R_r)$ be an idempotent symmetric balanced framework. If n is odd, and $n \geq \max\{t, 2r+1\}$, then R can be realized by an idempotent symmetric partial latin square Q of order n.*

Proof. Suppose n is odd and $n \geq \max\{t, 2r+1\}$. Let Q' be a partial latin square of order n where all cells are empty except for an $r \times r$ array of cells in the top-left corner which is filled by P. We will describe a procedure for obtaining the required partial latin square Q by filling in all the cells of Q' that are not in the top-left $r \times r$ subsquare. The procedure will consist of $n - r$ stages. During the lth stage $(1 \leq l \leq n - r)$ a “border” of cells will be filled, consisting of the first $r + l - 1$ cells of column $r + l$, the first $r + l - 1$ cells of row $r + l$ and cell $(r + l, r + l)$.

We will assume that before stage l is performed, each symbol appears at least $2(r + l - 1) - n$ times in the row lists and the upper-left $(r + l - 1) \times (r + l - 1)$ subsquare. Since $n \geq 2r + 1$, this will certainly be the case when $l = 1$.

To perform stage l we construct a bipartite graph G with bipartition (A, B), where the vertices in A are labelled $a_1, \ldots, a_n$ and correspond to the symbols $1, \ldots, n$ and the vertices in B are labelled $b_1, \ldots, b_{r+l}$ and correspond to the first $r + l - 1$ cells in row (or column) $r + l$, together with a further vertex b_{r+l}. For all i, j, $(1 \leq i \leq n$ and $1 \leq j \leq r + l - 1)$, G has an edge between a_i and b_j if and only if the symbol i is not in R_j (if $1 \leq j \leq r$) and does not appear in the jth row of Q' (all $1 \leq j \leq r + l - 1$). For all i $(1 \leq i \leq n)$, G has an edge between a_i and b_{r+l} if and only if the symbol i does not appear on the diagonal of Q'.

We will say that a symbol is *critical* if it appears fewer than $2(r + l) - n$ times in the row lists and the upper-left $(r + l - 1) \times (r + l - 1)$ subsquare. We claim that the maximum degree $\Delta(G) = n - (r + l - 1)$, and that all vertices in B and each vertex in A that corresponds to a critical symbol has degree $n - (r + l - 1)$.

First, each vertex in B has degree $n - (r + l - 1)$, since in each row there are $r + l - 1$ symbols that are either in the row list or already appear in the row. Consider a symbol σ that is not critical. Then a_σ has degree at most $(r + l) - (2(r + l) - n) = n - (r + l)$.

Now consider a symbol τ that is critical. Let $\nu(\tau)$ be the number of times τ appears in the row lists and the upper-left $(r + l - 1) \times (r + l - 1)$ subsquare. So $\nu(\tau)$ is either $2(r + l - 1) - n$ or $2(r + l - 1) - n + 1$. We consider two cases, depending on whether τ appears on the diagonal or not. First, suppose τ

appears on the diagonal. In this case $\nu(\tau)$ is odd, as any symbol must occur an even number of times off the diagonal and an even number of times in the row lists. Since n is also odd, we must have $\nu(\tau) = 2(r + l - 1) - n$. In G, a_τ is not adjacent to b_{r+l} and nor to $2(r + l - 1) - n$ vertices of $\{b_1, \ldots, b_{r+l-1}\}$, so the degree of a_τ is $(r + l) - (2(r + l - 1) - n) - 1 = n - (r + l - 1)$.

Now suppose τ does not appear on the diagonal. In this case $\nu(\tau)$ is even, and so we must have $\nu(\tau) = 2(r + l - 1) - n + 1$. In G, a_τ is adjacent to b_{r+l} but not adjacent to $2(r + l - 1) - n + 1$ vertices of $\{b_1, \ldots, b_{r+l-1}\}$, so the degree of a_τ is $(r + l) - (2(r + l - 1) - n + 1) = n - (r + l - 1)$.

Thus $\Delta(G) = n - (r + l - 1)$ and the critical vertices of A, and all vertices in B, have degree $n - (r + l - 1)$. By Theorem 2.11 (König's Theorem), G has an edge-colouring using the colours $\{1, \ldots, n - (r + l - 1)\}$. By taking one of the colour classes, we get a matching M, that covers all the vertices of B and each vertex of A that corresponds to a critical symbol. The matching M indicates how to fill in the first $r + l - 1$ cells of column $r + l$, and the first $r + l - 1$ cells of row $r + l$, for we place symbol i in cells $(r + l, j)$ and $(j, r + l)$ if a_i is joined to b_j by an edge of M, if $1 \leq j \leq r + l - 1$, and we place the symbol i in cell $(r + l, r + l)$ if a_i is joined to b_{r+l} by an edge of M.

We claim that after we have done this, each symbol appears at least $2(r+l)-n$ times in the row lists and the upper-left $(r + l) \times (r + l)$ subsquare. For symbols that were not critical, this is clearly the case. If a critical symbol was placed in cell $(r + l, r + l)$ then it appears in one more cell than before, but as it appeared an even number of times previously, it now appears an odd number of times, and so appears $2(r + l) - n$ times. All other critical symbols appear two more times than before, and so they each appear at least $2(r + l) - n$ times.

By performing stages $1, \ldots, n - r$ we create a partial latin square Q where the symbols missing from rows 1 to r are those in the lists $R_1, \ldots, R_r$ and the symbols missing from columns 1 to r are those in the lists $R_1, \ldots, R_r$. In other words, R is realized by Q. $\square$

4.6 Complexity Questions

In this section, we use the terminology of Garey and Johnson [22]. We will need the following theorem of Kratochvíl [29]. A (k-in-m)-colouring of an m-uniform hypergraph is a colouring of the vertices with red and blue such that each edge contains exactly k red vertices and $m-k$ blue vertices.

Theorem 4.14. *For every $q \geq 3$, $m \geq 3$ and $1 \leq k \leq m-1$, the problem of deciding $(k$-in-$m)$-colourability of q-regular m-uniform hypergraphs is NP-complete.*

In particular, the following problem is NP-complete.

Problem 4.15. Let H be a 4-uniform 4-regular hypergraph. Decide if H is 2-in-4 colourable.

A partial latin square of order n is said to be *L-shaped* if the cells in the upper-left $r \times s$ rectangle are empty, for some $r, s \in \{1, \ldots, n\}$, and the remaining cells are filled.

Problem 4.16. Let Q be an L-shaped partial latin square. Decide if Q is completable.

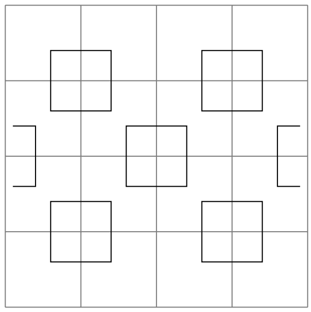
We shall show that Problem 4.16 is NP-complete, by giving a reduction from Problem 4.15. This reduction is a Karp reduction, also called a *polynomial transformation* [22]. Our reduction will have the feature that it maps 4-uniform 4-regular hypergraphs to L-shaped partial latin squares that satisfy Hall's Condition.

Lemma 4.17. *Problem 4.16 is NP-complete. Moreover, there is a reduction from Problem 4.15 to Problem 4.16 that maps 4-uniform 4-regular hypergraphs to L-shaped partial latin squares that satisfy Hall's Condition.*

We will need the following lemma.

(a)

$a_{i,j,0}$ $a_{i,j,1}$ $b_{i,k,0}$	$a_{i,j,0}$ $a_{i,j,1}$ $b_{i,k,0}$	$a_{i,j,0}$ $a_{i,j,1}$ $b_{i,k,1}$	$a_{i,j,0}$ $a_{i,j,1}$ $b_{i,k,1}$
$b_{i,k,0}$ $b_{i,k-1,5}$	$b_{i,k,0}$ $b_{i,k,4}$	$b_{i,k,1}$ $b_{i,k,4}$	$b_{i,k,1}$ $b_{i,k,5}$
$b_{i,k-1,5}$ $b_{i,k,2}$	$b_{i,k,4}$ $b_{i,k,2}$	$b_{i,k,4}$ $b_{i,k,3}$	$b_{i,k,5}$ $b_{i,k,3}$
$b_{i,k,2}$ $a_{i,j,3}$ $a_{i,j,2}$	$b_{i,k,2}$ $a_{i,j,3}$ $a_{i,j,2}$	$b_{i,k,3}$ $a_{i,j,3}$ $a_{i,j,2}$	$b_{i,k,3}$ $a_{i,j,3}$ $a_{i,j,2}$



(b)

Figure 4.6: (a) The symbols supported in a 4×4 subsquare. (b) The supports indicated by rectangles.

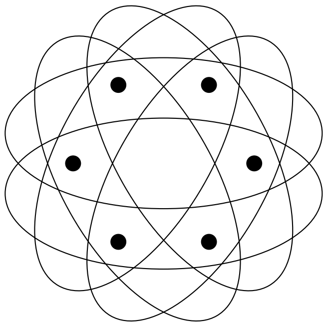
Lemma 4.18. *Let P be a partial latin square of order n. For all $\sigma \in \{1, \ldots, n\}$, let $\nu(\sigma)$ denote the number of times that σ appears in P. For each empty cell b of P we let $S(b)$ denote the set of symbols supported by b. Suppose that for each empty cell b of P,*

$$\sum_{\sigma \in S(b)} \frac{1}{n - \nu(\sigma)} \geq 1. \tag{4.9}$$

Then P satisfies Hall's Condition.

Proof. By Lemma 4.4, Hall's Condition holds if Hall's Inequality holds for each set of empty cells. Let T be a set of empty cells of P. For each symbol $\sigma \in \{1, \ldots, n\}$, let T_σ denote the subset of T consisting of the cells that support σ. Then

$$\begin{aligned}
\sum_{\sigma=1}^{n} \alpha(\sigma, T) &\geq \sum_{\sigma=1}^{n} \frac{|T_\sigma|}{n - \nu(\sigma)} && \text{(by Lemma 4.6)} \\
&= \sum_{b \in T} \sum_{\sigma \in S(b)} \frac{1}{n - \nu(\sigma)} \\
&\geq \sum_{b \in T} 1 && \text{(by (4.18))} \\
&= |T|,
\end{aligned}$$



$$\begin{pmatrix} 1 & 0 & 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 \end{pmatrix}$$

Figure 4.7: An example of a 4-uniform 4-regular hypergraph H, where the edges are drawn as ellipses (left), and its incidence matrix D (right).

and so $\mathcal{H}(T)$ is satisfied. □

Proof of Lemma 4.17. We will show that Problem 4.16 is NP-complete, by giving a reduction from Problem 4.15.

Let H be a 4-uniform 4-regular hypergraph on n vertices. Because H is 4-uniform and 4-regular, it has n edges. We will describe the construction of a framework $R = R(H)$, derived from H. We will suppose the vertices and edges of H are each labelled $0, \ldots, n-1$. Let $D = D(H)$ be the $n \times n$ incidence matrix of H defined by the rule that D_{ij} is 1 if vertex i and edge j are incident, and 0 otherwise.

We will construct a balanced framework $R = (r, s, t, P,\ R_1, \ldots, R_r,\ C_1, \ldots, C_s)$ where $r = s = 4n$, P is an empty partial latin square of order $4n$, and $t = 4n^2 + 12n$. Instead of using the positive integers $\{1, \ldots, t\}$ as symbols, we will use symbols from three sets A, B, and C. A consists of the symbols $a_{j,k}$ for all j, k, where $0 \le j \le n-1$ and $0 \le k \le 3$. B consists of the symbols $b_{i,j,k}$ for all i, j, k where $0 \le i \le n-1$, $0 \le j \le 3$ and $0 \le k \le 5$. C consists of the symbols $c_{i,j,k}$ for all i, j, k where $0 \le i \le n-1$, $0 \le j \le n-1$, $0 \le k \le 3$ and vertex i is not incident with edge j.

We will describe R by giving the support of each symbol, or in other words, by giving the $4n \times 4n$ admissible symbol array of R. In a balanced framework,

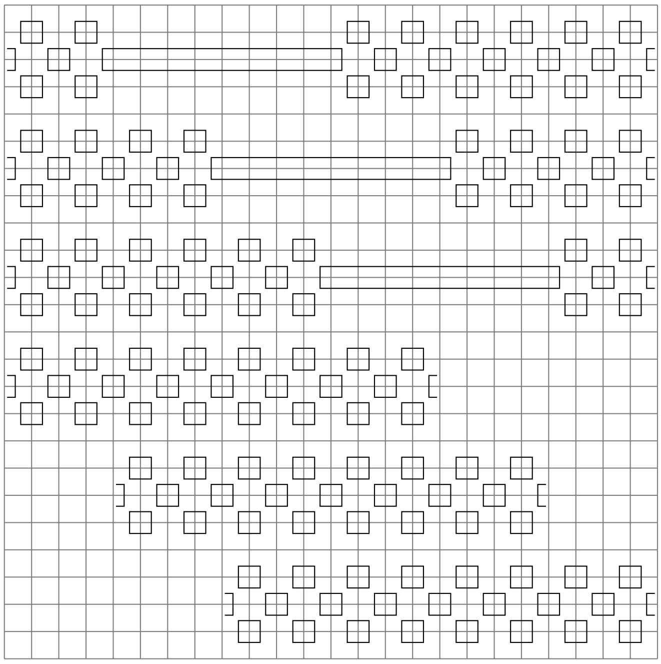


Figure 4.8: The framework $R = R(H)$.

each symbol must be supported by a set of cells that can be made into a square by permuting the rows and columns, and we will ensure that this is the case. Symbols in A and C will each appear in four row and column lists, and symbols in B will each appear in two row and column lists. Given the supports, it is an easy task to construct the row and column lists.

R is constructed in the following way. Each entry (i, j) of the incidence matrix D corresponds to a 4×4 subsquare of cells in R whose top-left corner is cell $(4i, 4j)$.

If $D_{ij} = 1$, the cells in the 4×4 subsquare at cell $(4i, 4j)$ support some symbols from A and some from B. Suppose that the (i, j) entry in D is the $(k + 1)$th occurrence of 1 on the ith row. (In other words, edge j is the $(k + 1)$th edge that is incident with vertex i in our ordering.) Each of the 4 cells in the top row of the subsquare support symbols $a_{j,0}$ and $a_{j,1}$, and each of the 4 cells in the bottom row of the subsquare support symbols $a_{j,2}$ and $a_{j,3}$. The cells of the subsquare also support symbols $b_{i,k,l}$ for all $0 \leq l \leq 5$, and $b_{i,k-1,5}$ (where subtraction is taken modulo 4), in the manner shown in Figure 4.6(a). Figure 4.6(b) gives a simplified picture, where the supports are indicated by rectangles.

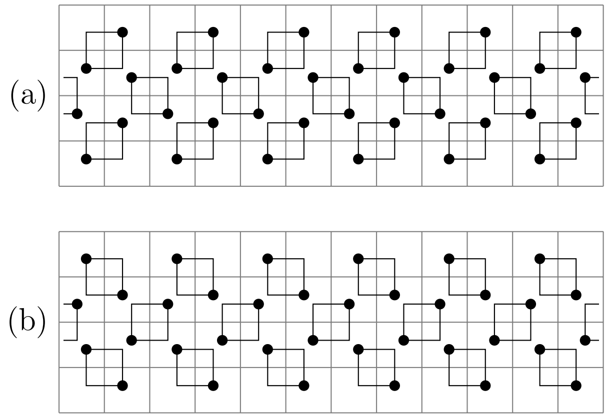


Figure 4.9: (a) "sloping left" and (b) "sloping right".

If $D_{ij} = 0$, the cells in the 4×4 subsquare at cell $(4i, 4j)$ support some symbols from C. Each of the 16 cells supports the 4 symbols $c_{i,j,0}$, $c_{i,j,1}$, $c_{i,j,2}$ and $c_{i,j,3}$. These symbols are merely "placeholders", and play no interesting role in the reduction.

Figure 4.7 gives an example of a 4-uniform 4-regular hypergraph H and its incidence matrix D. Figure 4.8 illustrates the framework $R = R(H)$, where the supports of the symbols in B are indicated by rectangles. It is a simple task to verify that the support of each symbol can be made into a square by permuting rows and columns, and that each row and column list contains $4n$ symbols.

We claim that R that can be completed if and only if H is 2-in-4 colourable.

First, suppose R is completable. Notice that for a fixed vertex i of H $(0 \leq i \leq n-1)$ the position of just one symbol $b_{i,k,l}$ (for some $0 \leq k \leq 3$ and $0 \leq l \leq 5$) determines where all the others are placed. Each of these symbols appears twice in the row and column lists, and so is supported by four cells, which we can call top-left, top-right, bottom-left and bottom-right. In a completion of R, each of these symbols appears twice, in either the bottom-left and top-right cells, or the top-left and bottom-right cells. However, due to the way the supports intersect, there are only two possible ways that these symbols can be placed. These are indicated in Figure 4.9, where we have used rectangles to indicate the supports, and circles to indicate the placement of the symbols. The first case we call

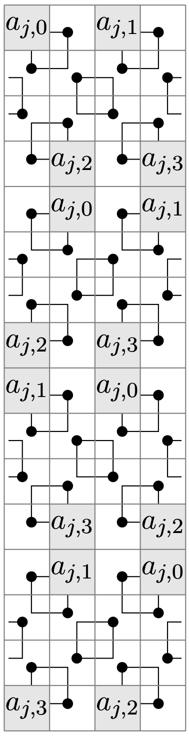


Figure 4.10: Placing the symbols $a_{j,k}$ for $0 \leq k \leq 3$.

"sloping left" and the other case "sloping right".

Next observe that for any edge j of H, two of its incident vertices must "slope left" and two must "slope right". To see this, consider the four vertices incident with edge j, and the four 4×4 subsquares corresponding to these incidences. In these four subsquares, we need to place four copies of each of $a_{j,0}$, $a_{j,1}$, $a_{j,2}$ and $a_{j,3}$. If more than two vertices "sloped left", at most one of these symbols could be placed in column $4j$, and so it would not be possible to place each of these symbols four times in the columns $4j, \ldots, 4j+3$. A similar argument applies in the case that more than two vertices "slope right". Figure 4.10 illustrates how the symbols $a_{j,0}$, $a_{j,1}$, $a_{j,2}$ and $a_{j,3}$ might be placed, in the case where two vertices "slope left" and two "slope right".

So if we colour each vertex in H red or blue according to whether its associated symbols in B "slope" left or right, we obtain a colouring of H, where each edge has 2 blue and 2 red vertices. Hence H is 2-in-4 colourable.

Conversely, suppose H is 2-in-4 colourable. Then we can choose a 2-in-4 colouring of H, and place the symbols of B according to whether their associated vertices in H are coloured red or blue (e.g. red vertices "slope left", blue vertices "slope right"). This leaves empty cells in which the symbols $a_{j,k}$ for all j, k, $0 \le j \le n-1$ and $0 \le k \le 3$, can be placed. The symbols of C can be placed without any difficulty, as the symbols $c_{i,j,0}$, $c_{i,j,1}$, $c_{i,j,2}$ and $c_{i,j,3}$ for a fixed i and j ($0 \le i \le n-1$ and $0 \le j \le n-1$) are supported by only the cells in the 4×4 subsquare whose top-left corner is $(4i, 4j)$. So we just need to fill this subsquare with a latin square on these 4 symbols. Hence R is completable.

By Theorem 4.12, R can be realized by a partial latin square P of order $n = r + s$. The proof of Theorem 4.12 given in Section 4.5 is of a constructive nature, and provides a procedure for computing P. This procedure consists of edge-colouring two bipartite graphs, each with $O(t^2)$ edges. This can be done in time polynomial in t, as an edge-colouring of a bipartite graph $G = (V, E)$ with $\Delta(G)$ colours can be found in $O(|V|\,|E|)$ time [38, Chapter 20].

Finally, we claim that P satisfies Hall's Condition. Because of the manner in which we constructed R, and subsequently P, each empty cell supports either: (i) two symbols from B, (ii) one symbol from B and two symbols from A, or (iii) four symbols from C. Each symbol from B is missing from two rows and columns, and each symbol from A and C is missing from four rows and columns. Hence by Lemma 4.18, P satisfies Hall's Condition. $\square$

We can now prove Theorem 1.11.

Proof of Theorem 1.11. A Turing machine equipped with an oracle for solving Problem 1.10 could solve Problem 4.15 in polynomial time by transforming instances of Problem 4.15 into instances of Problem 1.10, using the reduction given in Lemma 4.17, and then calling the oracle. Since Problem 4.15 is NP-complete, it follows that Problem 1.10 is NP-hard. $\square$

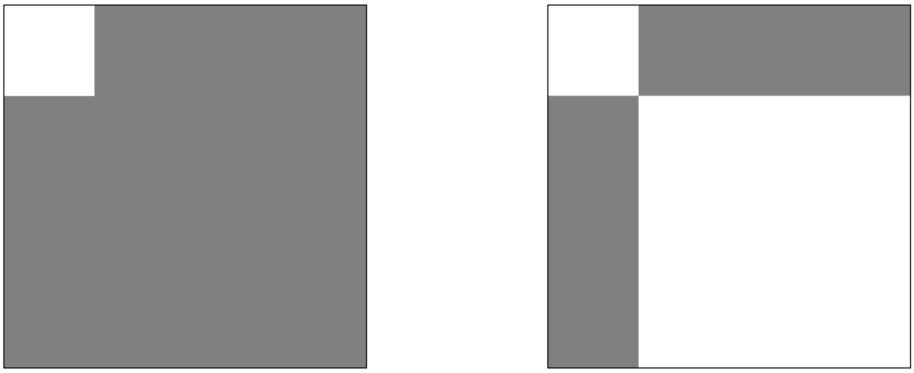


Figure 4.11: From the proof of Theorem 4.21.

4.7 Variants of Theorem 1.11

We can also consider the following problems, where $\epsilon > 0$ is fixed.

Problem 4.19. Let P be a partial latin square that satisfies Hall's Condition where the proportion of empty cells is less than ϵ. Decide if P is completable.

Problem 4.20. Let P be a partial latin square that satisfies Hall's Condition where the proportion of filled cells is less than ϵ. Decide if P is completable.

Theorem 4.21. *Problems 4.19 and 4.20 are NP-hard.*

Proof. In Lemma 4.17, we gave a reduction from Problem 4.15 to Problem 1.10. Given a 4-regular 4-uniform hypergraph H on u vertices, we constructed a balanced framework $R = R(H)$, and then argued that by Theorem 4.12 it could be realized by a partial latin square of order $n = 4u^2 + 12u$ where the upper left $4u \times 4u$ subsquare is empty. So as n tends to infinity, the proportion of empty cells tends to zero. In fact, this actually shows that Problem 4.19 is NP-hard. However, just to be sure we can modify the reduction of Lemma 4.17 to create partial latin squares where the proportion of empty cells is less than ϵ. This is easy to do, because Theorem 4.12 says that we can find an L-shaped partial latin square realizing R of any order at least n. So in particular, we can find an

L-shaped partial latin square P of order $n' = \lceil 2/\epsilon \rceil n$ that realizes R. In this way, for any $\epsilon > 0$, we can obtain a reduction from Problem 4.15 to Problem 4.19.

To show that Problem 4.20 is also NP-hard, we can use the same reduction, but with the extra step that we delete the symbols in the bottom-right $(n'-r)\times(n'-s)$ rectangle of P to create a partial latin square P^* (see Figure 4.11). The proportion of filled cells in P^* is less than ϵ and it is clear that P^* can be completed if and only if P can. It remains to show that P^* satisfies Hall's Condition. Suppose we have a partial latin square that satisfies Hall's Condition and we delete a symbol. The support of a symbol σ in the old square is a subset of the support of σ in the new square, and so if the old square satisfies Hall's Condition, the new one must too. So as P satisfies Hall's Condition, so does P^*. □

Finally, we have the following problem. A *symmetric* partial latin square is one that is equal to its transpose. An *idempotent* partial latin square is a partial latin square where each symbol appears once on the diagonal. Note that a symmetric idempotent latin square is necessarily of odd order, as each symbol must appear an even number of times off the diagonal. Idempotent symmetric latin squares exist for any odd order n, as the addition table for $\mathbb{Z}/n\mathbb{Z}$ is symmetric and contains distinct entries on the diagonal.

Problem 4.22. Let P be a symmetric idempotent partial latin square that satisfies Hall's Condition. Decide if P is completable.

Note that by saying that P is *completable*, we mean P is completable to a latin square that is symmetric and idempotent.

Theorem 4.23. *Problem 4.22 is NP-hard.*

Proof. We will give a reduction from Problem 4.15 to Problem 4.22. Let H be a 4-uniform 4-regular hypergraph on n vertices. By the process described in the proof of Theorem 4.17, we can construct a balanced framework $R = (r, s, t, P, R_1, \ldots, R_r, C_1, \ldots, C_s)$ that is completable if and only if H is 2-in-4 colourable.

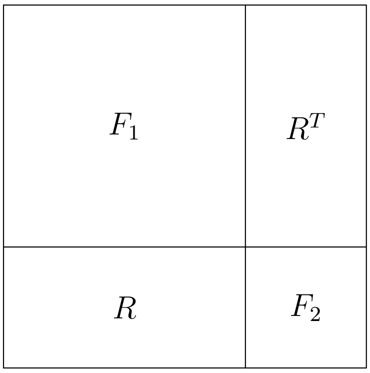


Figure 4.12: From the proof of Theorem 4.23.

We will assume that both r and s are odd. If there are not, we can add an extra row and/or column to R, and some new symbols to the row and column lists.

We will describe the construction of a symmetric idempotent framework $R^S = (r+s, r+s, t+r+s, P, L_1, \ldots, L_{r+s}, L_1, \ldots, L_{r+s})$, which will be derived from R. R^S will use the symbols $\{1, \ldots, t\}$ and the additional symbols $I = \{i_1, \ldots, i_s\}$ and $J = \{j_1, \ldots, j_r\}$. R^S will have row lists $L_1, \ldots, L_{r+s}$ as follows: $L_k = C_k \cup (I - i_k)$ for $1 \leq k \leq s$, and $L_k = R_{k-s} \cup (J - j_{k-s})$ for $s+1 \leq k \leq r+s$. Because R^S is symmetric, the column lists are also $L_1, \ldots, L_{r+s}$. P will have the symbols $i_1, \ldots, i_s, j_1, \ldots, j_r$ on the diagonal, in this order, and all other cells empty. R is a valid symmetric framework, as each symbol appears an even number of times in the lists.

We claim that R^S can be completed if and only if R can be completed. If R can be completed, then we can complete R^S in the manner shown in Figure 4.12. This involves placing a completion of R and the transpose of such a completion as shown, together with idempotent symmetric latin squares F_1 on the symbols I and F_2 on the symbols J. Note that we required r and s to be odd so that suitable squares F_1 and F_2 exist.

Conversely, suppose that R^S can be completed. Because the symbols I only appear in the first s rows and columns, the top-left $s \times s$ subsquare must be a latin square on these symbols. Similarly, the bottom-right $r \times r$ subsquare must be a latin square on the symbols J. Hence the situation must be as in Figure 4.12, with two copies of a completion of R (one transposed) filling the remaining cells.

By Theorem 4.13, R^S can be realized by an idempotent symmetric partial latin square P of any odd order $n \geq \max\{t, 2r+1\}$. The procedure for finding P is given in the proof of Theorem 4.13 and requires edge-colouring bipartite graphs. As before, it is not hard to verify that this task can be done in polynomial time. Finally, it is a simple application of Lemma 4.18 to verify that P satisfies Hall's Condition. $\square$

Bibliography

[1] A. S. Asratian, T. M. J. Denley and R. Häggkvist, *Bipartite graphs and their applications*, Cambridge University Press, 1998.

[2] B. B. Bobga, *Some necessary conditions for list colorability of graphs and a conjecture on completing partial latin squares*, Ph.D. thesis, Auburn University, 2008.

[3] B. B. Bobga and P. D. Johnson, Jr., Completing Partial Latin Squares: Cropper's Problem. *Congr. Numer.*, **188** (2007), 211–216.

[4] J. A. Bondy and U. S. R. Murty, *Graph Theory*, Springer, 2008.

[5] D. Cariolaro, *The 1-factorization problem and some related conjectures*, Ph.D. thesis, University of Reading, 2004.

[6] D. Cariolaro and A. J. W. Hilton, Class 1 graphs associated with regular graphs of high degree, *Congr. Numer.*, **159** (2002), 133–142.

[7] D. Cariolaro and A. J. W. Hilton, An application of Tutte's Theorem to 1-factorization of regular graphs of high degree, *Discrete Math.*, **309** (2009), 4736–4745.

[8] A. G. Chetwynd and A. J. W. Hilton, Regular graphs of high degree are 1-factorizable, *Proc. Lond. Math. Soc.*, **50** (1985), 193–206.

[9] A. G. Chetwynd and A. J. W. Hilton, Star multigraphs with three vertices of maximum degree, *Math. Proc. Cambridge Philos. Soc.*, **100** (1986), 303–317.

[10] A. G. Chetwynd and A. J. W. Hilton, 1-factorizing regular graphs of high degree—an improved bound, *Disc. Math.*, **75** (1989), 103–112.

[11] C. J. Colbourn. Complexity of completing partial Latin squares. *Disc. Appl. Math.*, **8** (1984), 25–30.

[12] M. Cropper, personal communication.

[13] A. B. Cruse, On embedding incomplete symmetric Latin squares, *J. Combin. Theory Ser. A*, **16** (1974), 18–22.

[14] R. Diestel. *Graph Theory*, 3rd edition, Springer, 2005.

[15] G. A. Dirac, Some theorems on abstract graphs. *Proc. Lond. Math. Soc.*, **2** (1952), 69–81.

[16] A. L. Dulmage and N. S. Mendelsohn, Coverings of bipartite graphs, *Canad. J. Math.*, **10** (1958), 517–534.

[17] T. Easton and R. G. Parker. On completing latin squares. *Disc. Appl. Math.*, **113** (2001), 167–181.

[18] D. G. Fon-Der-Flaass, Arrays of distinct representatives—a very simple NP-complete problem. *Discrete Math.*, **171** (1997), 295–298.

[19] L. R. Ford and D. R. Fulkerson, Maximal flow through a network, *Canad. J. Math.*, **8** (1956), 399–404.

[20] J.-C. Fournier, Colorations des arêtes d'un graphe, *Cahiers Centre Études Recherche Opér.*, **15** (1973), 311–314.

[21] A. M. Frieze, Complexity of a 3-dimensional assignment problem. *European J. Oper. Res.*, **13** (1983), 161–164.

[22] M. R. Garey and D. S. Johnson. *Computers and Intractability: A Guide to the Theory of NP-Completeness*, W. H. Freeman, 1979.

[23] O. Goldreich, On promise problems: a survey, *Lecture Notes in Comput. Sci.*, vol. 3895, Springer, 2006, 254–290.

[24] J. Goldwasser, personal communication.

[25] R. Häggkvist, unpublished manuscript.

[26] P. Hall, On representatives of subsets, *J. Lond. Math. Soc.*, **10** (1935), 26–30.

[27] A. J. W. Hilton and P. D. Johnson, Jr., A variation of Ryser's theorem and a necessary condition for the list-colouring problem, *Graph colourings*, Longman, 1990.

[28] D. G. Hoffman and P. D. Johnson, Jr., Extending Hall's theorem into list colorings: a partial history, *Int. J. Math. Math. Sci.*, 2007.

[29] J. Kratochvíl. Complexity of Hypergraph Coloring and Seidel's Switching. *Graph Theoretic Concepts in Computer Science*, Springer, 2003, 297–308.

[30] C. C. Lindner and C. A. Rodger, Generalized Embedding Theorems for Partial Latin Squares, *Bull. Inst. Combin. Appl.*, **5** (1992), 81–99.

[31] L. Lovász and M. D. Plummer, *Matching theory*, North-Holland, 1986.

[32] T. Niessen and L. Volkmann, Class 1 conditions depending on the minimum degree and the number of vertices of maximum degree, *J. Graph Theory*, **14** (1990), 225–246.

[33] L. Perković and B. Reed, Edge coloring regular graphs of high degree, *Discrete Math.*, **165–166** (1997), 567–578.

[34] M. Plantholt, Overfull conjecture for graphs with high minimum degree, *J. Graph Theory*, **47** (2004), 73–80.

[35] B. Reed, Edge coloring nearly bipartite graphs, *Oper. Res. Lett.*, **24** (1999), 11–14.

[36] A. Rosa and W. D. Wallis, Premature sets of 1-factors or how not to schedule round robin tournaments, *Discrete Appl. Math.*, **4** (1982), 291–297.

[37] H. J. Ryser, A combinatorial theorem with an application to latin rectangles, *Proc. Amer. Math. Soc.*, **2** (1951), 550–552.

[38] A. Schrijver. *Combinatorial Optimization*, Springer, 2003.

[39] Z.-X. Song and H. P. Yap, Chromatic index critical graphs of even order with five major vertices, *Graphs Combin.*, **21** (2005), 239–246.