University of Louisville

ThinkIR: The University of Louisville's Institutional Repository

Electronic Theses and Dissertations

12-2012

Connected matchings in special families of graphs.

Christopher James Caragianis 1980-
University of Louisville

Follow this and additional works at: https://ir.library.louisville.edu/etd

CONNECTED MATCHINGS IN SPECIAL FAMILIES OF GRAPHS

By

Christopher James Caragianis
B.S., University of Louisville, 2007
M.A., University of Louisville, 2009

A Dissertation
Submitted to the Faculty of the
College of Arts and Sciences of the University of Louisville
in Partial Fulfillment of the Requirements
for the Degree of

Doctor of Philosophy

Department of Mathematics
University of Louisville
Louisville, Kentucky

December, 2012

CONNECTED MATCHINGS IN SPECIAL FAMILIES OF GRAPHS

By

Christopher James Caragianis
B.S., University of Louisville, 2007
M.A., University of Louisville, 2009

A Dissertation Approved on

November 15, 2012

by the following Dissertation Committee:

<u>André Kézdy</u>
Dissertation Director

<u>Grzegorz Kubicki</u>

<u>Thomas Riedel</u>

<u>David Wildstrom</u>

<u>John S. Usher</u>

DEDICATION

For my wife Sarah

who loved me through it all.

ACKNOWLEDGMENTS

ABSTRACT
Connected Matchings in Special Families of Graphs

Christopher J. Caragianis

November 15, 2012

A *connected matching* in a graph is a set of disjoint edges such that, for any pair of these edges, there is another edge of the graph incident to both of them. This dissertation investigates two problems related to finding large connected matchings in graphs.

The first problem is motivated by a famous and still open conjecture made by Hadwiger stating that every k-chromatic graph contains a minor of the complete graph K_k. If true, Hadwiger's conjecture would imply that every graph G has a minor of the complete graph $K_{n/\alpha(G)}$, where $\alpha(G)$ denotes the independence number of G. For a graph G with $\alpha(G) = 2$, Thomassé first noted the connection between connected matchings and large complete graph minors: there exists an $\epsilon > 0$ such that every graph G with $\alpha(G) = 2$ contains $K_{\frac{1}{3}+\epsilon}$ as a minor if and only if there exists a positive constant c such that every graph G with $\alpha(G) = 2$ contains a connected matching of size cn. In Chapter 3 we prove several structural properties of a vertex-minimal counterexample to these statements, extending work by Blasiak. We also prove the existence of large connected matchings in graphs with clique size close to

the Ramsey bound by proving: for any positive constants b and c with $c < \frac{1}{4}$, there exists a positive integer N such that, if G is a graph with $n \geq N$ vertices, $\alpha(G) = 2$, and clique size at most $b\sqrt{n \log(n)}$, then G contains a connected matching of size cn.

The second problem concerns the computational complexity of finding the size of a maximum connected matching in a graph. This problem has many applications including, when the underlying graph is chordal bipartite, applications to the bipartite margin shop problem. For general graphs, this problem is NP-complete. Cameron has shown the problem is polynomial-time solvable for chordal graphs. Inspired by this and applications to the margin shop problem, in Chapter 4 we focus on the class of chordal bipartite graphs and one of its subclasses, the convex bipartite graphs. We show that a polynomial-time algorithm to find the size of a maximum connected matching in a chordal bipartite graph reduces to finding a polynomial-time algorithm to recognize chordal bipartite graphs that have a perfect connected matching. We also prove that, in chordal bipartite graphs, a connected matching of size k is equivalent to several other statements about the graph and its biadjacency matrix, including for example, the statement that the complement of the latter contains a $k \times k$ submatrix that is permutation equivalent to strictly upper triangular matrix.

TABLE OF CONTENTS

LIST OF FIGURES

1. INTRODUCTION

In broadest terms, this dissertation investigates "pairing up" problems. Assigning workers to jobs, putting students into cooperative pairs, scheduling processes to machines, and choosing efficient locations for building bridges between roads are all examples of this sort of problem. We look at special kinds of pairing schemes that have certain interconnectedness properties, and how they can be accomplished under various restrictions in diverse circumstances.

When we model these scenarios with graphs, the type of pairing we are interested in will be called a connected matching. While gaining insight into practical issues such as the ones listed above is a worthy goal on its own, we will see that detecting connected matchings in graphs is a rich and complex computational task. Furthermore, in certain families of graphs, the existence of large connected matchings is equivalent to Hadwiger's conjecture, a famous open problem in graph theory.

The objective in this study is to carry forward the theory of connected matchings in both the computational and structural areas. We aim to further flesh out the complexity of detecting connected matchings, determining with greater specificity the classes of graphs for which connected matching problems are tractable. We also present progress on a conjecture related to the existence of large connected matchings

in graphs with independence number 2. We will also consider real world problems to which we apply our findings.

In the remainder of this chapter, we walk through the basic graph theoretic ideas employed in this dissertation, and introduce some new ideas that clarify later explanations. Next, we will introduce the extremal problem of connected matchings and place it in a theoretical context. In Chapter 3, we will present recent progress and original work on a version of that extremal problem. In Chapters 4 and 5, we will explore the computational issue of detecting connected matchings in some special families of graphs, and cite examples of optimization problems that can make use of these results. Finally, we will take a look at problems that are tangential to our main subject, but too appealing to be left behind.

1.1. Background

Formally speaking, a *finite simple graph* $G = (V(G), E(G))$ consists of a set $V(G)$ called the *vertex set* and a set $E(G)$ of two element subsets of $V(G)$ called the *edge set*. We think of a graph as a collection of vertices, some of which are joined by edges. Graphs are typically used to model systems with a connectedness relation, as in the following examples:

- A collection of people is modeled with a vertex set and an edge is placed between pairs of vertices corresponding to pairs of people who are friendly with each other (a *social network* graph).

- A collection of computers is modeled with a vertex set and an edge is placed

between pairs of vertices corresponding to pairs of computers with a physical network cable between them (a *computer network* graph).

- The courses in a mathematics department are modeled with vertices. An edge is placed between them if their meeting times overlap (a *scheduling* graph).

We use a shortened notation for edges. If edge e in a graph G is the set containing vertices u and v, we write $e = uv$. The *complement* of a graph G, denoted $\overline{G}$, is the graph on the same vertex set as G in which the edge $uv \in E(\overline{G})$ if and only if $uv \notin E(G)$.

1.1.1. Matchings and connected matchings

An edge e of a graph is said to be *incident* to each of its endpoints. A vertex v is *saturated* or *covered* by any collection of edges containing an edge incident to v. A *matching* in a graph is a collection of disjoint edges. A *perfect* matching is a matching that covers every vertex in the graph exactly once. Each vertex covered by a matching is "matched" unambiguously to one other vertex by the matching edge that touches it.

A powerful theoretical tool for proving the existence of matchings is *Hall's condition*. For a set S of vertices in a graph G, let $N(S)$ be the *open neighborhood* of S, that is, the set of vertices in $V(G) - S$ that are adjacent to a vertex from S. Hall's condition, which can be proven by induction, states that if S has the property that for every subgraph $T \subseteq S$, $|N(T) \cap V(G) - S| \geq |T|$, then there is a matching in G from S to $G - S$ that saturates S. We will apply Hall's condition several times in this

| neighborly | neighborly | separable |
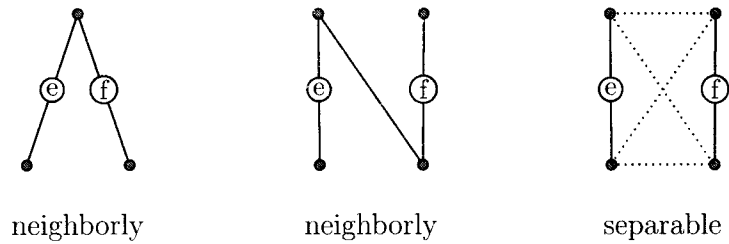
Figure 1.1: Neighborly and separable pairs of edges.

study.

Matchings in graphs are used to model pairings. Consider a graph with one subset of vertices representing jobs and another representing workers. Place an edge between a job and a worker if that worker is trained to perform that job. If each worker performs one job at a time, and each job needs only one worker, then a feasible assignment of jobs to workers is modeled by a matching.

A pair of disjoint edges $e = uv$ and $f = xy$ in a graph G is *separable* if and only if none of the edges ux, uy, vx, vy are present in G. A pair of edges that are not separable is called *neighborly*. A graph is called separable if and only if it has a separable pair ofedges. Figure 1.1.1 exhibits the ways in which a pair of edges in a graph may be separable or neighborly.

A *connected matching* in a graph G is a matching M with the additional property that no two edges of M are separable in G. We use this property to model various "matching-type" problems with extra requirements.

4

1.1.2. Graph coloring

Suppose we have a graph $G = (V, E)$. If we take a subset V' of V and subset E' of E so that every e in E' is a subset of V', then the graph $G' = (V', E')$ is called a *subgraph* of G. Intuitively, we create a subgraph by selecting vertices from a graph, and then choosing to include or exclude edges from the original graph whose endpoints are among the vertices we have selected. An edge that has both endpoints in a set S of vertices is said to be *induced* by S. The subgraph created by choosing a set S of vertices and all edges induced by S is called the subgraph of G induced by S, denoted $G[S]$.

A set of vertices in a graph is called *independent* if it induces a graph with no edges. Let $\alpha(G)$ denote the size of the largest independent vertex set of G, called the *independence number* of a graph G.

Consider the scheduling graph we introduced earlier in this section. If two classes are in meetings at the same time at any point during the week, the corresponding vertices are joined by an edge. Each independent set therefore represents a collection of courses that do not overlap in time. Hence, a collection of courses we could feasibly assign to a single instructor corresponds to an independent set of vertices in the scheduling graph. We therefore model the task of assigning courses to instructors by partitioning the vertices of the scheduling graph into independent sets.

Another way of looking at a partition of this kind is as a *proper vertex coloring*. Suppose we give each instructor a color, and visualize the scheduling graph with each vertex colored according to the instructor assigned to the corresponding course. If the

assignment is feasible, then no pair of adjacent vertices has the same color assigned to each. Formally, we define a proper vertex coloring as a partition $\mathcal{C} = \{C_1, C_2, \ldots, C_k\}$ of the vertex set of a graph G with the property that each color class C_i is an independent set.

A natural question arises in the scheduling graph model. How many instructors are needed to teach a certain collection of courses? In general, given a graph G, how many colors are needed to properly color its vertices? The minimum number of colors needed to properly color a graph G is called the *chromatic number* of G, denoted $\chi(G)$. For instance, a collection of three mutually adjacent vertices clearly requires three different colors in a proper coloring. However, a sequence of vertices each of which is adjacent to the precedent and antecedent vertices (a *path* graph) requires only two colors no matter how many vertices are used. A particularly interesting class of graphs are those with chromatic number two, called *bipartite graphs*. We denote a bipartite graph G as $(A, B; E)$ where E is the edge set and A and B are the *partite sets*, i.e., the independent sets in a particular proper coloring of G.

1.1.3. Perfect graphs

The complementary notion of independence is *completeness*. If a subset S of vertices in a graph G induces all possible edges between vertices of S, we say S induces a *complete graph* or a *clique*. For the complete graph on n vertices, we use the notation K_n. The size of the largest clique contained in G is the *clique number*, denoted $\omega(G)$.

It's not hard to see that if a graph contains a clique on k vertices, then a proper coloring will require at least k colors. Each vertex of the clique is adjacent to all the
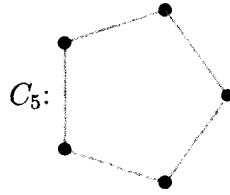
Figure 1.2: The graph C_5 requires three colors, but has no triangle as a subgraph.

others, and thus requires its own color. However, it is not the case that a graph must contain a clique on k vertices to require k colors for proper coloring. Consider a graph on five vertices v_1, v_2, v_3, v_4, and v_5. Add an edge between vertices whose indices differ by 1 modulo 5. This graph, as shown in Figure 1.1.3 is known as the cycle on five vertices (denoted C_5); it has no clique of three vertices, yet requires three colors for a proper coloring.

A graph G is *perfect* if and only if the chromatic number of each of its induced subgraphs is equal to the clique number of that subgraph. One of the great accomplishments of modern graph theory has been the characterization of perfect graphs given by the Strong Perfect Graph Theorem.

Theorem 1 (Strong Perfect Graph Theorem). *A graph G is perfect if and only if both G and $\overline{G}$ have no induced subgraph that is an odd cycle of length at least 5.*

This statement stood as a conjecture since its announcement by Berge in 1960 until finally proven by Chudnovsky, Robertson, Seymour, and Thomas [9] in 2006.
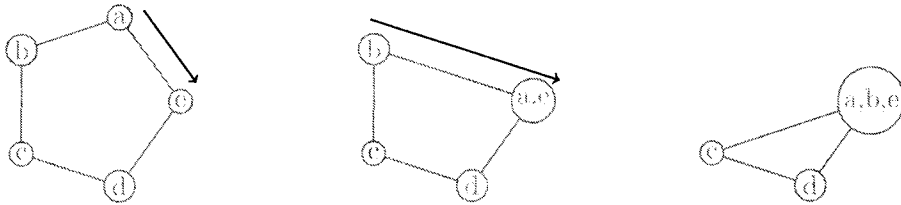
Figure 1.3: C_5 contains K_3 as a minor.

1.1.4. Hadwiger's conjecture

One of the key threads in the development of graph theory through the past 60 years has been the search for upper bounds on the chromatic numbers of graphs. One type of conjectured upper bound involves relaxing the notion of a clique to some broader class of graphs that includes cliques. A *subdivision* of a graph G is a graph obtained from G by successively *subdividing* edges of G. To subdivide an edge uv, we introduce a new vertex w, and replace the edge uv with edges uw and wv. Hajós conjectured that a graph with chromatic number at least k must contain a subdivision of a k-vertex clique as a subgraph. This was shown to be false for $k \geq 7$ by Catlin [8] in 1979, but the cases of $k = 5$ and $k = 6$ remain open. This (ultimately false) conjecture was offered as a strengthening of the classic conjecture of Hadwiger concerning graph minors.

We say that G contains H as a minor (denoted $H \leq_m G$) if a graph isomorphic to H can be obtained from a subgraph of G by a series of edge contractions and vertex deletions. An edge contraction, as we see in Figure 1.3, consists of identifying a pair of adjacent vertices u and v as a single vertex that is adjacent to any vertices adjacent to u or v. The sets of vertices contracted to exhibit a particular minor in a graph are

called the *branch sets* of that minor.

The *Hadwiger number* of a graph G, denoted $\eta(G)$, is the largest integer n for which G contains K_n as a minor. Hadwiger's conjecture [20] claims that this is an upper bound on the chromatic number of a graph.

Conjecture 1. *For any graph G,*

$$\eta(G) \geq \chi(G)$$

Some progress has been made toward resolving this conjecture. Erdős, Bollobas and Catlin [5] showed that almost all[1] graphs satisfy the conjecture. For chromatic numbers 1 and 2 the result is trivial. For chromatic number 3 it is a simple exercise[2]. Dirac [10] proved the case of chromatic number 4 in 1952. The case of chromatic number 5 can be shown to be equivalent to the famous four-color theorem, and Robertson, Seymour and Thomas proved the same in 1993 for the case of chromatic number 6 [32]. For larger chromatic numbers, the conjecture remains open. One motivation for this study is an attempt to develop other types of partial results, as we will see in the following chapter.

---

[1] A *random graph* is a probability space on all graphs with n vertices. The space used in this proof is such that the probability of any particular edge being present in the sample graph is a fixed value $0 < p < 1$. In this case, "almost all" means with probability approaching 1 as n increases without bound.

[2] To see this, consider that any graph without an odd cycle as an induced subgraph can be properly colored with two colors. Thus a graph with chromatic number 3 must have an odd cycle that can be contracted to a triangle.

1.2. Complexity

A large part of this dissertation is concerned with algorithms for finding connected matchings. We ask whether or not "good" algorithms for finding connected matchings can be found. The starting point, as we will see in Chapter 4, is that the general problem of finding connected matchings is *NP-complete*. The complexity class NP consists of computational problems that have a *polynomial-time verifier*, which is to say that a proposed solution can, if correct, be *shown* to be correct in a number of steps that scales polynomially with the size of the input. It is unknown whether or not the class NP coincides with the class P of problems that can be solved in a number of steps that scales polynomially with the size of the input. This is the famous P vs. NP question, far and away the outstanding open question in computational complexity theory.

If it can be shown that every other problem in NP can be reduced to a given problem in NP, then this problem is said to be NP-complete. In practice, this is typically accomplished by reducing one of a known collection of NP-complete problems to the problem in question. For an excellent overview of the theory of NP-completeness, see Garey and Johnson [16]

1.3. Proximity coloring

Two vertices u and v in a graph G are *connected* if there is a sequence of vertices, each adjacent to the previous, that begins with u and ends with v (Such a sequence is called a *path*). We define the *connectivity* of a pair of connected vertices in a graph

G as the minimum number of vertices that must be deleted from G so that u and v are not connected in the resulting graph. If all pairs of vertices in G are connected, then we call G a *connected graph*. If G has n vertices, we define the *connectivity* of G (denoted $\kappa(G)$) as $n-1$ if G is a complete graph and as the minimum connectivity of any pair of vertices from G otherwise. A collection of vertices whose removal renders a graph disconnected is called a *cut* and a cut in G of size $\kappa(G)$ is called a *minimum cut*.

When a connected graph is used as a "map" of some real-world network of interconnected nodes, we may be curious about how many "steps" it takes to get from one node to another. Suppose we look at a city road map and would like to determine the shortest route from the corner of 4th and Hill streets to the corner of 9th and Jefferson streets. What do we mean by shortest? An actual road map would be able to tell us the distance "as the crow flies" or the distance along roads. We could estimate our travel time based on speed limit and traffic considerations. Or we could ask for the route that minimizes the number of stop lights we encounter. Let us consider this last notion of distance. We model the intersections as vertices in a graph and place an edge between vertices corresponding to intersections joined by an uninterrupted stretch of roadway. The "minimum stop light path" is the same as the shortest (in the sense of fewest edges) path in the graph from one vertex to another.

The distance in a graph G between vertices u and v (denoted $d(u, v)$ or $d_G(u, v)$ if we wish to emphasize the choice of graph) is the number of edges in the shortest path between u and v using edges of G. The maximum distance between vertices of G is called the *diameter* of G, denoted $diam(G)$. In the above example, we may not

11

be interested in the number of miles we drive from one intersection to the other, but only in the number of intersections through which we travel. Graph distance tells us about the proximity of vertices in the topology of the graph.

Suppose G is a connected graph on n vertices. Consider also the edges of the complete graph on n vertices (denoted K_n). We construct the *proximity partition* of $E(K_n)$ by collecting these edges into sets according to the minimum distance in G between the endpoints. Define a relation $\sim$ on the edges of K_n so that $uv \sim u'v'$ if and only if $d_G(u,v) = d_G(u',v')$. Clearly, this is an equivalence relation. The proximity partition of $E(K_n)$ induced by G consists of the equivalence classes of this relation. Sometimes we may only be interested in distances that fall below a certain threshold. Perhaps in our application certain distances are simply considered "too far" to be of any note. In this case, we consider the *proximity k-partition* $\mathcal{P}_G^k$ induced by G, where for $1 \leq i < k$, vertices are related as above, and all pairs u,v with $d_G(u,v) > k$ are collected into a single class. For small values of k, we may think of this partition as an edge coloring of K_n and refer to the *proximity k-coloring* of K_n.

The proximity 3-coloring is helpful in studying connected matchings, so we introduce a further shorthand. We color the distance 1 edges (i.e. the underlying graph) blue, the distance 2 edges green, and all other edges red, as we see in Figure 1.4. We then refer to the resulting edge colored K_n as the *RGB* graph induced by G (see, e.g., Figure 1.4).
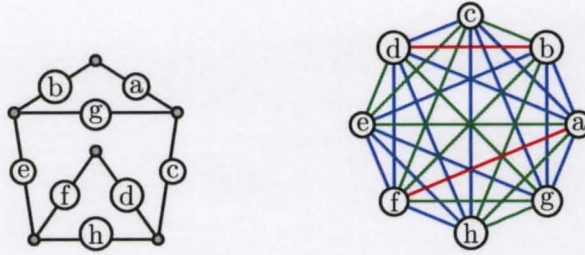
Figure 1.4: A graph and the induced RGB graph.
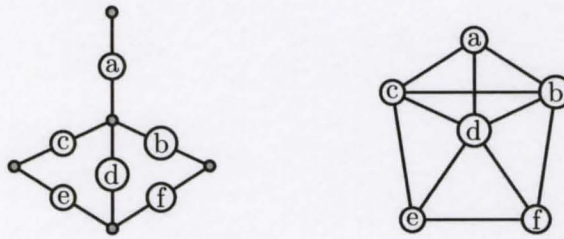
Figure 1.5: A graph and its line graph.

1.3.1. Line Graphs

Proximity colorings and RGB graphs become useful to connected matching problems when we consider *line graphs*. The line graph of a graph G (denoted $L(G)$) is a graph whose vertex set is the edge set of G and vertices in $L(G)$ are adjacent if and only if the corresponding edges in G share an endpoint. A simple example of a line graph is exhibited in Figure 1.5.

Proposition 1. *Connected matchings in a graph G correspond to green cliques in the RGB graph induced by $L(G)$.*

Proof. Incident edges of G correspond to adjacent vertices in $L(G)$. If we have a pair of edges $e, f \in E(G)$ that are disjoint and neighborly, then there is some edge $g \in E(G)$ between an endpoint of e and an endpoint of f. Hence, $eg, fg \in E(L(G))$

13
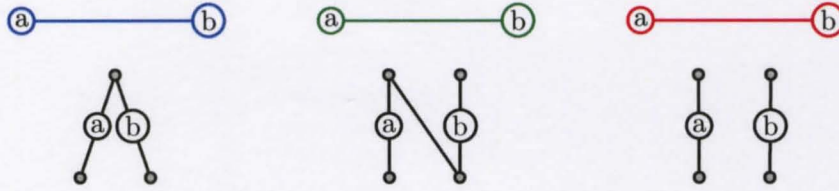
Figure 1.6: Correspondence between pairs of edges in G and the RGB edges of $L(G)$.

and $ef \notin E(L(G))$. This means that $d_{L(G)}(e, f) = 2$, and ef is green in the RGB graph induced by $L(G)$, as shown in Figure 1.3.1. Since connected matchings are collections of pairwise non-incident and neighborly edges, the green cliques in the RGB graph induced by $L(G)$ correspond to connected matchings in G. $\qquad\square$

Predictably, some complexity results on connected matching problems essentially rest on clique problems in RGB graphs of line graphs. While clique problems are in general NP hard, there are special classes of graphs, such as perfect graphs, for which they can be solved efficiently. We will see that some graphs retain this quality in the green graph of their line graph.

2. THE EXTREMAL PROBLEM

Most of the work that has been done on connected matchings concerns the minimum size of a largest connected matching in a graph with certain properties. This is the *extremal* problem of connected matchings. In particular, how many vertices must a member of a certain class of graphs have before the existence of a connected matching of a certain size is guaranteed? In this chapter we discuss a connection between a certain special case of Hadwiger's conjecture and the extremal problem of connected matchings.

2.1. Extremal problems

Extremal graph theory is concerned with finding the maximum or minimum (by a variety of measures) graphs that have a certain property. Typically, an extremal problem asks "How many edges must be present in a graph with n vertices to ensure X", or "How many vertices must a graph with property P have to ensure X". The iconic example of an extremal problem, one which we use in Section 3.2.1, is the problem of edges and cliques. How many edges must a graph on n vertices have to ensure the existence of a clique of size k? The answer is due to Túran in 1941. Túran's answer [34] explicitly constructed the extremal graph, that is, the graph on n vertices

with no k-clique and the maximum number of edges. Let $T(n, r)$ be the graph on n vertices constructed by dividing the vertices as evenly as possible into r parts and adding all edges among the parts. This is the *complete balanced r-partite graph* on n vertices, also called a *Túran graph*.

Theorem 2. *The n vertex graph with no complete subgraph on r vertices and the maximum number of edges is $T(n, r - 1)$.*

Many proofs of this theorem can be found in various graph theory texts, see [1] for several interesting proofs.

2.2. A special case of Hadwiger's conjecture

In Chapter 1 we discussed some partial results on Hadwiger's conjecture, primarily for the special cases arising from restricting the chromatic number. Now we turn our attention to a different sort of special case. A proper vertex coloring can be thought of as a partitioning of the vertex set into independent sets. This gives us an easy lower bound on the chromatic number in terms of the size of a largest independent set. The least number of colors needed would be realized when all color classes are the same size, so $\chi(G) \geq n/\alpha(G)$. This leads to the following weakening of Hadwiger's conjecture.

Conjecture 2 (Weaker version of Hadwiger's conjecture). *For all graphs G,*

$$\eta(G) \geq \frac{n}{\alpha(G)}.$$

16

At the present time, this conjecture is open for any particular value of α. However, it was recently shown by Fradkin [14] to hold for claw-free graphs with $\alpha \geq 3$. An examination of this problem by Duchet and Meyniel [11] yielded the following bound.

Theorem 3. *For any graph* G,

$$\eta(G) \geq \frac{n}{2\alpha(G) - 1}$$

This is turn was improved by Kawarabayashi et. al [21] for almost all values of α

Theorem 4. *For any graph* G *on* n *vertices with* $\alpha(G) \geq 3$

$$\eta(G) \geq \frac{n(4\alpha(G) - 2)}{(4\alpha(G) - 3)(2\alpha(G) - 1)}$$

The first improvement by an absolute constant factor comes from Fox [13] who shows that

Theorem 5. *Let* $c = \frac{29 - \sqrt{813}}{28}$. *Then for any graph* G,

$$\eta(G) \geq \frac{n}{(2 - c)\alpha}$$

The specific case of $\alpha(G) = 2$ has attracted much attention. Plummer, Stiebitz and Toft gave this case a thorough treatment in [30]. Despite many partial and tangential

results on this case, the bound in Theorem 3 is still the best known for $\alpha = 2$ (Note that for small values of α, the bound in Theorem 3 is better than the bound in Theorem 5). In addition to their work on Hadwiger's conjecture, Plummer et. al introduce the idea of a connected matching. This led to the following extended conjecture.

Conjecture 3 (PST extension of Hadwiger's conjecture). *Every graph G with $\alpha(G) = 2$ and n vertices has a connected matching M such that the contractions of the edges in M to $|M|$ single vertices result in a graph containing $K_{\lceil n/c \rceil}$*

This was also conjectured by Seymour and is sometimes referred to as *Seymour's strengthening of Hadwiger's conjecture.* Plummer, Stiebitz and Toft prove this conjecture for all inflations[1] of graphs with independence number 2 and fewer than 12 vertices, as well as inflations of an infinite family of $\alpha = 2$ graphs.

2.3. Connected matchings in graphs with $\alpha = 2$

The strengthened version of Hadwiger's conjecture for graphs with independence number 2 placed connected matchings front and center. Seymour is credited with presenting the problem of improving the bound of Duchet and Meyniel in the case of independence number 2.

Conjecture 4. *There exists $\epsilon > 0$ so that every graph G with n vertices and $\alpha(G) = 2$ contains $K_{\lceil (\frac{1}{3} + \epsilon) \rceil n}$ as a minor.*

---

[1]An *inflation* of a graph is obtained by replacing some vertices with complete graphs of any size all of whose vertices are adjacent to the neighbors of the replaced vertex.

One of the results of Kawarabayashi et. al in [21] effectively reduces this problem to an extremal problems on connected matchings.

Theorem 6. *If a graph G on n vertices with $\alpha(G) \leq 2$ contains a connected matching of size greater than or equal to $kn > 0$, then G has $K_{\lceil (n/3)(1+k/3) \rceil}$ as a minor.*

Conversely, if G has $K_{\lceil cn \rceil}$ as a minor for $c > \frac{1}{3}$, then G contains a connected matching of size at least $(3c - 1)n/4 - \frac{1}{2}$.

Thus, if there is some k for which *every* graph with independence number two on n vertices has a connected matching of size kn, then Conjecture 4 must be true. Gyárfás, Füredi, and Simonyi presented this extremal conjecture explicitly in [15]

Conjecture 5. *There exists some constant c such that every graph G with n vertices and $\alpha(G) = 2$ has a connected matching of size cn.*

Furthermore, they conjecture on the value of the constant c

Conjecture 6. *Every graph G with $4t - 1$ vertices and $\alpha(G) = 2$ has a connected matching of size t.*

They prove this for values of t up to 17, and show that it is sharp by exhibiting the example of G consisting of two disjoint and disconnected cliques.

Another result found in [30] is that if H is a 4-vertex graph with $\alpha(H) \leq 2$, and G is an n vertex graph with $\alpha(G) = 2$ and no copy of H as an induced subgraph, then G has $K_{\lceil n/2 \rceil}$ as a minor. Kriesell has recently [24] improved this result by adding the 5 vertex graphs to this list.
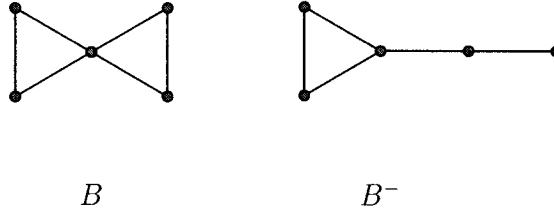
Figure 2.1: The graphs B and B^- referred to in Theorem 8.

Theorem 7. *Let H be any graph with $\alpha(H) \leq 2$ on at most 5 vertices. Then every $\{\overline{K_3}, H\}$-free graph on n vertices has a collection of $\lceil n/2 \rceil$ pairwise adjacent edges and vertices.*

From this and the second part of theorem 6 we conclude that every $\{\overline{K_3}, H\}$-free graph on n vertices has a connected matching of size $\lfloor \frac{n}{8} \rfloor$. However when $H = \overline{K_{2,3}}$, Kriesell has found that we can say even more.

Theorem 8. *Every connected, $\{\overline{K_3}, \overline{K_{2,3}}\}$-free graph on n vertices nonisomorphic to B or B^- in Figure 2.1 has a connected dominating matching of size $\lfloor \frac{n}{2} \rfloor$.*

3. THE EXTREMAL CONJECTURE

Now we focus directly on the extremal conjecture of Gyárfás, Füredi and Simonyi (Conjecture 5). We show that among certain families of graphs with independence number 2, the size of a maximum connected matching is *linear* in the number of vertices. We "piggyback" in some cases on results related to Seymour's strengthening of Hadwiger's conjecture, optimizing them for connected matchings. Furthermore, we prove the conjecture for a class of graphs that appear with probability one in a natural random process generating graphs with independence number 2.

The progress on Conjecture 5 presented in this chapter splits into *large clique cases* and *small clique cases*. For graphs with $\alpha = 2$ that have large cliques (i.e., cliques whose size is linear in the number of vertices in the graph) it is relatively simple to construct the largest possible connected matching arising from a large clique as we show in Section 3.1. The structure of graphs with independence number 2 allows us to draw conclusions on the possible minimum degree and connectivity properties of a counterexample to Conjecture 5.

3.1. Large clique cases

As noted earlier, Gyárfás, Füredi and Simonyi prove that for $t \leq 16$, any graph G with $\alpha(G) = 2$ has a connected matching of size t. In so doing, they implicitly introduce an important lemma concerning the relationship between cliques and connected matchings.

Lemma 1 (GFS). *Let $0 < c < 1/4$. If G is a graph with $\alpha(G) \leq 2$ with $\omega(G) \geq cn$, then G has a $\lfloor cn \rfloor$-connected matching.*

Proof. Let S be a set of cn vertices inducing a clique. For any subset of $S' \subseteq S$, the intersection $I = \bigcap_{s \in S'} \{v \in V(G) : sv \notin E(G)\}$ induces a clique. If for any $S' \subseteq S$, $|I| > n/2$, then there is a $n/4$-connected matching in the clique induced by I. Otherwise,

$$|N(S')| = n - |S'| - |I|$$
$$\geq n - |S'| - \frac{n}{2}$$
$$\geq \frac{n}{4}$$
$$> |S'|$$

for all $S' \subseteq S$. Hence, by Hall's condition (see *e.g.*, [35]), a matching from the vertices of S to the vertices of $V(G) - S$ exists. This matching must be connected, since S induces a clique. $\square$

We might suggestively call this the "spider lemma" as it exhibits a connected matching with a "head" (the clique) and many "legs" (the matching with one side

inducing the clique). Using the spider lemma, we deduce some structural qualities that a counterexample to Conjecture 5 must possess. Chief among these is high connectivity, because, in graphs with independence number at most two, disjoint sets of vertices with no edges between them must induce cliques. First, we make explicit the relationship between Conjecture 5 and Seymour's strengthening of Hadwiger's conjecture (As described in Chapter 2. Hereafter, SSH.). The following proposition shows that SSH implies the GFS conjecture.

Proposition 2. *If SSH holds for a graph G on n vertices with independence number 2, then $\nu_c(G) \geq n/4$.*

Proof. Let $\mathcal{M}$ be the collection of branch sets of a $K_{n/2}$ minor of G that satisfies the hypothesis of SSH (all branch sets of size 2 or 1). Let M_1 be the collection of elements of $\mathcal{M}$ consisting of single vertices and M_2 the collection of elements of $\mathcal{M}$ consisting of edges. Obviously, M_2 is a connected matching. Furthermore, any matching of the clique induced by M_1 forms a connected matching that extends the connected matching formed by M_2, because

$$\nu_c(G) \geq \lfloor \frac{|M1|}{2} \rfloor + |M2| \geq \lfloor \frac{|M1| + |M2|}{2} \rfloor = \lfloor n/4 \rfloor$$

$\square$

In Lemma 2.1 of Blasiak's paper [3], the author shows that any $\alpha = 2$ graph with connectivity less than $n/2$ satisfies SSH. We show the following for higher connectivity.

Lemma 2. *If G is a graph on n vertices with $\alpha(G) \leq 2$, then $\nu_c(G) \geq \frac{n - \kappa(G)}{4}$. If*

$\kappa(G) \geq n/2$, *then* $\nu_c(G) \geq n - \kappa(G)$.

Proof. Let G be a graph on n vertices with $\alpha(G) \leq 2$. Because cuts in $\alpha = 2$ graphs separate components that must be cliques, at least one component has at least $\frac{n-\kappa(G)}{2}$ vertices. Any matching in this clique component is connected, so G has a connected matching of size $\frac{n-\kappa(G)}{4}$.

The proof of the second claim follows the strategy of Lemma 2.1 of [3]. Assume that $\kappa(G) \geq n/2$. Let S be a minimum cut set of G. Let L, R be a partition of $V(G) - S$ so that there are no edges joining vertices from L to vertices from R. Since $\alpha(G) = 2$, L and R are cliques. Every vertex of S is adjacent to every vertex of L or adjacent to every vertex of R. We say that a vertex is *complete to* a set of vertices if it is adjancet to every vertex in the set. Let S_L be the set of vertices complete to L and S_R be the set of vertices complete to R. We claim that between any $A \subseteq S_L$ with $|A| \leq |R|$ and R (S_R and L resp.) there is a matching that saturates A. Suppose there is no matching from R that saturates A. Halls condition then implies that there is a subset T of A such that $|N(T) \cap R| < |T|$. But then $(S - T) \cup (N(T) \cap R)$ is a cut set separating $L \cup T$ and $R - N(T)$. This set is smaller than S, yielding a contradiction.

Let M be the largest possible matching obtained with edges between S_L and R (temporarily named *type 1 edges*) and edges between S_R and L (*type 2 edges*). This matching is connected. To see this, note that the collections of edges of each type form "spiders" because R and L are cliques. Furthermore, without loss of generality, the S_L ends of the type 1 edges are complete to L, and hence adjacent to an endpoint of every type 2 edge.

If both $|R| \leq |S_L|$ and $|L| \leq |S_R|$ (and $\kappa(G) \geq n/2$) , then we can find $T_L \subseteq S_L$ and $T_R \subseteq S_R$ so that T_L and T_R are disjoint, $|T_R| = |L|$, and $|T_L| = |R|$. Thus, using the above claim, we construct a connected matching saturating $V(G) - S$.

Without loss of generality, $|R| > S_L$. This means that $|L| \leq S_R$, as $S \geq n/2$ and $S \subseteq S_L \cup S_R$. Let R^u denote the set of vertices of R unmatched by M. Since $|S| \geq n/2$, we assume that we have matched all the vertices of S_L to vertices in R. Consequently, there are at least $|R^u|$ unmatched vertices of S_R (denoted S_R^u). Augment M with any matching from the biclique between R^u and S_R^u saturating R^u to yield M' . These new edges are mutually connected, and connected to any type 1 or type 2 edges via edges of R in the case of type 1, or edges from S_R to R in the case of type 2. Now M' is a connected matching saturating $R \cup L$, and $|R \cup L| = n - \kappa(G)$. $\qquad\square$

This lemma, together with the spider lemma, allows us to collect the properties of a "large-clique" counterexample to Conjecture 5. Say that an edge e *dominates* a vertex $v \notin e$ if v is adjacent to an endpoint of e. The following proposition collects these properties.

Proposition 3. *If G_c is a vertex-minimal counterexample to Conjecture 5 with n vertices, then the following conditions must be satisfied*

1. *$\omega(G_c) < cn$.*

2. *$\delta(G_c) \geq (1 - c)n$.*

3. *G_c has diameter 2.*

4. *G_c is $(1 - c)n$-connected.*

5. G_c has no edge dominating $c(n-1)$ vertices.

Proof. Items 1 and 4 follow directly from the two lemmas. The collection of non-neighbors of a vertex in an $\alpha = 2$ graph form a clique, which leads to item 2. Applying the pigeonhole principle together with the assumption in item 2, every pair of vertices share a neighbor, so G_c has diameter 2.

Consider a vertex-minimal counterexample G_c. Following from the work of Gyárfás et al., we know that small graphs (fewer than 67 vertices) satisfy Conjecture 5 for any $c \le 4$. Assume that the conjecture holds up to $n-1$ vertices. Any edge dominating a connected matching clearly extends that connected matching. Thus our minimal counterexample must not dominate $c(n-1)$ vertices, lest it dominate a connected matching of size $n-1$ by the induction hypothesis. $\square$

Taken altogether, the results of this section show that a counterexample to Conjecture 5 must be highly connected, yet avoid large cliques. As Blasiak remarks in [3], this means that the most "mysterious" cases arise when minimum degree and connectivity are close to the number of vertices in the graph. These are what we call the small clique cases. We show next that one important class of these graphs, the *Ramsey graphs* with independence number 2, satisfy Conjecture 5.

3.2. The small clique cases

To work on graphs with independence number 2 and only the smallest possible cliques, we employ results from an area known as *Ramsey theory*. Ramsey theory is concerned with the study of highly organized, unavoidable substructures in large structures.

This unavoidability is sometimes summarized by the statement "Complete disorder is impossible".

3.2.1. Ramsey's theorem

Ramsey theory tells us that a large enough graph must either have a clique or independent set of a certain size. This makes intuitive sense, if there are few enough edges in a graph to avoid a large clique, then there ought to be a large independent set. Ramsey formalized and proved this intuition in the following theorem [31].

Theorem 9. *For any pair of positive integers (r, s), there exists a least positive integer $R(r, s)$ such that for any graph on $R(r, s)$ vertices, there exists either a complete subgraph on r vertices in G or an independent set of s vertices in G.*

This number $R(r, s)$ is known as a *Ramsey number.*

We study graphs that are, in a sense, extremal with respect to Ramsey numbers. In particular, we study $\alpha = 2$ graphs. This means we are most interested in the *triangle* Ramsey numbers $R(3, k)$. In [23], Kim famously proved that the magnitude of $R(3, k)$ is on the order of $k^2/\log k$. For any positive constant B, we call the class of $\alpha = 2$ graphs with n vertices and no clique of size $B\sqrt{n \log n}$ the *triangle Ramsey graphs.* These graphs are very highly connected, so the results of the previous section are of little help. In this section, we show the Conjecture 5 holds for sufficiently large Ramsey graphs, with a value of c arbitrarily close to $1/4$. We also discuss a natural random $\alpha = 2$ graph model which almost certainly satisfies Conjecture 5.

Theorem 10. *Let $c < 1/4$ be a constant. For any constant b and sufficently large*
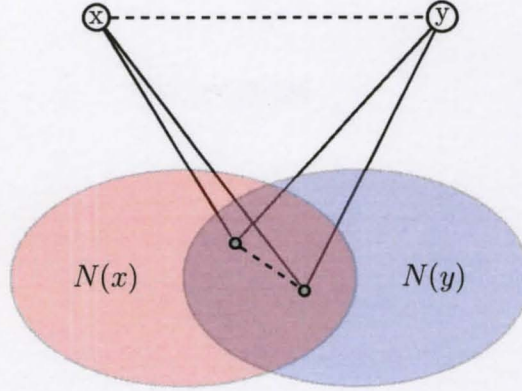
Figure 3.1: Illustration of the count from Eq. 3.1

n, every $\alpha = 2$ graph G on n vertices with $\omega(G) < b\sqrt{n \log n}$ has a cn-connected matching.

First, we prove a lemma that will place a bound on the number of pairs of separable edges in an $\alpha = 2$ graph with a given clique number. In order to simplify the notation in the proof, we work on the complementary notion of cycles of four vertices in triangle-free graphs with a given independence number.

Lemma 3. *For every pair of positive constants ϵ, d there is $n_{\epsilon,d}$ such that every triangle-free graph G with $n > n_{\epsilon,d}$ vertices and $\alpha(G) < d\sqrt{n \log n}$ has fewer than ϵn^3 copies of C_4.*

Proof. Fix $\epsilon, d > 0$ and let G be a triangle free graph on n vertices with $\alpha(G) < d\sqrt{n \log n}$. Let X_{C_4} be the number of copies of C_4 in G. Then

$$X_{C_4} = \frac{1}{2} \sum_{\{u,v\} \notin E(G)} \binom{|N(u) \cap N(v)|}{2} \tag{3.1}$$

For each nonadjacent vertex pair $\{u, v\}$, we count the number of distinct pairs of

28

vertices in the intersection of the neighborhoods of u and v. This counts each C_4 twice, so we divide by two. Fix $\epsilon_1 < \sqrt{8\epsilon}$.

Claim. For sufficiently large n, fewer than $n^2(\log n)^{-2}$ pairs of vertices u, v have neighborhood intersection larger than $\epsilon_1\sqrt{n}$.

Suppose the contrary is true, and there are more than $n^2(\log n)^{-2}$ pairs u, v so that $|N(u) \cap N(v)| \geq \epsilon_1\sqrt{n}$. We count the total number of vertices in these intersections. The count is at least $\epsilon_1 n^{5/2}(\log n)^{-2}$, meaning some vertex is counted at least $\epsilon_1 n^{3/2}(\log n)^{-2}$ times. However, $\Delta(G) \leq \alpha(G) < d\sqrt{n\log n}$, so each vertex is in at most

$$\binom{d\sqrt{n\log n}}{2} < \frac{d^2}{2}n\log n$$

neighborhood intersections. Thus, for sufficiently large n, the claim holds.

Now we bound X_{C_4}. We overestimate by supposing that there are precisely $n^2(\log n)^{-2}$ pairs of vertices with the largest possible vertex intersection, and the remainder have neighborhood intersection of size $\epsilon_1\sqrt{n}$.

$$X_{C_4} < \frac{1}{2}\left[\frac{n^2}{(\log n)^2}\binom{d\sqrt{n\log n}}{2} + \left(|E(\overline{G})| - \frac{n^2}{(\log n)^2}\right)\binom{\epsilon_1\sqrt{n}}{2}\right] \qquad (3.2)$$

The right hand side asymptotically equals

$$\sim \frac{\epsilon_1^2 - 2\epsilon_1}{8}n^3 \qquad (3.3)$$

This is strictly greater than cn^3, so for sufficiently large n, the desired bound on X_{C_4} holds. $\qquad \square$

Now we prove Theorem 10. We use the language of RGB proximity colorings introduced in Chapter 1.

Proof (of Theorem 10). Fix constants d and $c < 1/4$, and let G be a $\alpha = 2$ graph with n vertices, m edges, and $\omega(G) < b\sqrt{n \log n}$. Consider the RGB graph $\mathcal{G}$ induced by $L(G)$ (recalling that green k-cliques correspond to k-connected matchings in G and red edges correspond to induced $\overline{C_4}$s in G). If R, G, and B denote the number of red, green and blue edges respectively, we would like to show that

$$G = \binom{m}{2} - R - B \geq \binom{m}{2} - cn\binom{m/cn}{2} \tag{3.4}$$

equivalently

$$R + B \leq cn\binom{m/cn}{2} \tag{3.5}$$

guaranteeing by Túran's theorem (see, *e.g.*, [35]) a green clique on cn vertices in $\mathcal{G}$, and a cn-connected matching in G.

We obtain a crude upper bound on B by taking the number of edges in the line graph of K_n,

$$B < \frac{n^3}{2} - \frac{3n^2}{2} + n. \tag{3.6}$$

We bound R using Lemma 1. For any $\epsilon > 0$ and sufficiently large n,

$$R < \epsilon n^3. \tag{3.7}$$

Thus for sufficiently large n,

$$B + R < \frac{n^3}{2} + \epsilon n^3. \qquad (3.8)$$

We compare this with the right hand side of (3.5)

$$cn \binom{m/cn}{2} = \frac{cn}{2} \left(\frac{m^2}{c^2 n^2} - \frac{m}{cn} \right) \qquad (3.9)$$

$$= \frac{1}{2c} m^2 n^{-1} - \frac{m}{2} \qquad (3.10)$$

$$\sim \frac{n^3}{8c}. \qquad (3.11)$$

Since $c < 1/4$, and we can take $\epsilon < \frac{1-4c}{8c}$, for sufficiently large n, the inequality (3.5) holds and G has a cn-connected matching. $\qquad \square$

3.2.2. The triangle-free process

The *triangle-free process* is a method of stochastically constructing maximal triangle-free graphs. Let G_0 be the empty graph on n vertices and let O_i be the set of edges of $K_n - G_i$ that will not create a triangle when added to G_i. Then for each G_i, construct G_{i+1} by adding an edge chosen uniformly at random from O_i until some step k at which O_k is empty. The complementary version of this process is a natural source of $\alpha = 2$ graphs.

Bohman has shown in [4] that the triangle free process asymptotically almost surely produces graphs which satisfy the hypotheses of Theorem 10. This unfortunately falls short of a proof that Conjecture 5 holds for almost all $\alpha = 2$ graphs

31

because the triangle-free process does not produce a uniform distribution. Nonetheless, the triangle-free process indicates that the $\alpha = 2$ Ramsey graphs are in a sense "typical" among $\alpha = 2$ graphs and well worth studying.

In conclusion, we see that in $\alpha = 2$ graphs that are highly "spread-out" (Theorem 10) and in ones that are "bunched-up" (spider lemma) Conjecture 5 succeeds. It remains to be seen if further work in tuning and sharpening these techniques can close the gap, if new approaches are needed, or if indeed there lurks a counterexample somewhere in the middle ground

4. COMPUTING CONNECTED MATCHINGS

While the extremal problem of connected matchings has driven most of the theoretical development to date, there is another aspect of connected matchings that demands our attention. How do we *find* large connected matchings in graphs? In particular, how do we compute the size of the largest connected matching in a particular graph? Plummer et. al show in [30] that this problem is NP-hard in general.

A connected matching of size k has a clique minor of size k, so it may not be surprising that the general problem of finding connected matchings is difficult. There is a "clique-like" element to the problem, and so we might expect that the computational difficulty is similar to clique problems. However, maximum matchings in graphs can be found efficiently. Edmonds in [12] showed an algorithm that computes the maximum matching in a graph in time polynomial to the number of vertices in the graph. Furthermore, there are many elegant matching algorithms that we may adapt to the problem of constructing a maximum connected matching.

In this chapter, we investigate computational problems concerning connected matchings in special families of graphs. We attempt to utilize the special structure of these families to develop algorithms that find maximum connected matchings.

4.1. Known results

In [6], Cameron shows that the maximum connected matching problem is solvable in polynomial time for *chordal* graphs, by considering connected matchings as matchings contained in *neighborly* sets of edges. Neighborly sets of edges from a graph G correspond to cliques in the square of the line graph of G, $L(G)^2$. It then follows that if $L(G)^2$ has M maximal cliques, then the maximum connected matching in G can be found by solving M maximum matching problems. Hence, we state the following general result from which the result on chordal graphs follows.

Theorem 11. *If $\mathcal{P}$ is a class of graphs such that for any $G \in \mathcal{P}$ with n vertices, the number of maximal cliques in $L(G)^2$ is less than $f(n)$ where f is a polynomial, then the maximum connected matching problem can be solved in polynomial time for graphs from $\mathcal{P}$.*

The square of the line graph of a chordal graph is itself a chordal graph[1], and hence has a polynomial number of maximal cliques.

Conversely, Cameron also shows that the maximum weighted connected matching problem remains NP-complete on (0,1)-weighted bipartite graphs. It is not difficult to reduce this problem to the maximum clique problem on general graphs, by taking any graph H and replacing each vertex v with a new edge joining two new vertices v_1 and v_2, assigning the edge v_1v_2 a weight of 1. If two vertices u and v are adjacent in H, we produce edges u_1v_2 and u_2v_1 of weight zero. In the resulting graph, the edges

---

[1]In fact, Cameron showed in [7] that a larger class of graphs, the *weakly chordal* graphs also have this property. That is, the square of the line graph of a weakly chordal graph is also weakly chordal.
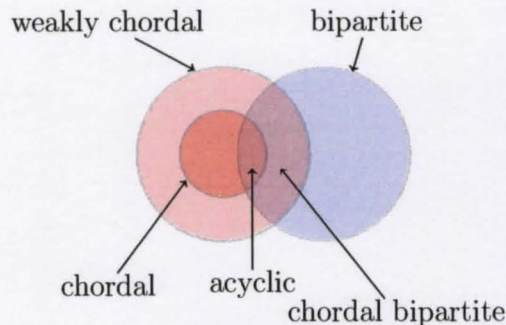
Figure 4.1: Inclusion diagram

of positive weight in any connected matching correspond to a clique in H and vice versa. It is important to note that this reasoning does not extend to the unweighted bipartite graphs.

4.2. Chordal bipartite graphs

To extend the results in the previous section, we turn to a relaxation of chordal graphs, the *weakly* chordal graphs. A weakly chordal graph is a graph in which every cycle of length five or greater has a chord. Also, we closely examine those weakly chordal graphs that are also bipartite; these graphs are known as *chordal bipartite* graphs (see Figure 4.2).

An important characterization of chordal bipartite graphs is that any non-empty induced subgraph of a chordal bipartite graph contains a *bisimplicial edge*. An edge uv is bisimplicial if the neighborhoods of the endpoints induce a complete bipartite graph. That is, for any $a \in N(u)$ and $b \in N(v)$, ab is an edge of the graph. Bisimplicial edges equip the chordal bipartite graphs with a perfect edge without vertex elimination ordering, i.e., upon removing a bisimplicial edge from a chordal bipartite graph, the

resulting graph is still chordal bipartite. These results, among others developing the theory of chordal bipartite graphs, can be found in [18].

4.2.1. Inert bisimplicial edges

If a graph is non-separable, then we quickly compute the maximum connected matching using any maximum matching algorithm we choose. The following result of Golumbic [17] is valuable when dealing with non-separable chordal bipartite graphs.

Theorem 12. *Let H be a chordal bipartite graph. If H is separable, then it has at least two separable bisimplicial edges.*

We will see shortly that once we have identified a pair of bisimplicial edges, we can always remove one of them from the graph without reducing the size of a maximum connected matching in the graph.

Let us say that an edge e in a graph G is *inert* if $\nu_c(G) = \nu_c(G - e)$. Bisimplicial edges in chordal bipartite graphs are inert unless the conditions in the following lemma are satisfied.

Lemma 4. *Let $G = (A, B; E)$ be chordal bipartite. A bisimplicial edge $e = uv$ is inert unless at least one of the following is true:*

1. *e is contained in every maximum connected matching.*

2. *For any maximum connected matching M, every edge of M has exactly one endpoint adjacent to an endpoint of e.*

3. *For any maximum connected matching M, $N(e)$ is covered by M.*

36
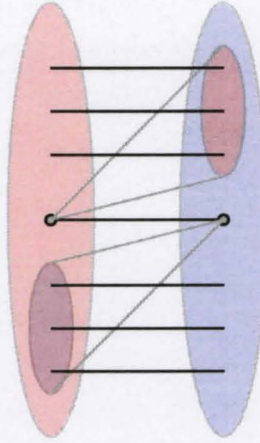
Figure 4.2: Picture of a non-inert bisimplicial edge. Every maximum connected matching saturates the neighborhood of the edge.

Proof. Suppose we have a chordal bipartite graph $G = (A, B; E)$ that has a bisimplicial edge $e = uv$ with $u \in A$ and $v \in B$. To prove that condition 1 is necessary, we must show that removing e does not reduce the size of a maximum connected matching that does not include e.

In fact we show more, to wit, that removing e does not reduce the size of *any* connected matching that does not include e. Suppose we have a connected matching M. If M does not cover the endpoints of e, then obviously removing e has no effect on the size of M. Thus we have $f = uu'$ and $g = v'v$ included in M. The only way removing e could reduce the size of M is by disconnecting f and g. However, since e is bisimplicial, $u' \in N(u)$, and $v' \in N(v)$; the edge $u'v'$ must be present. Hence removing e does not disconnect f and g (see Figure 4.2.1). In proving the necessity of condition 2, we choose a maximum connected matching M. By condition 1, we assume that $e \in M$. Suppose now that $f = u''v''$ is another edge of M and edges $g = uv''$ and $h = u''v$ are both present. We claim that we can remove e and f from

M
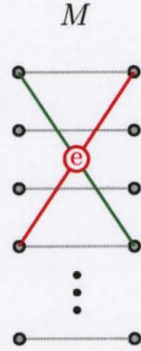
Figure 4.3: Illustration of proof of condition 1.

M
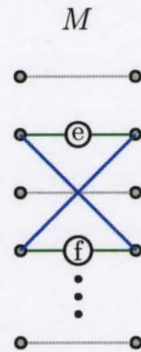
Figure 4.4: Illustration of proof of condition 2.

M and replace them with g and h. Let $d = u'''v'''$ be an edge of M. This edge d must

be connected to one of either g or h via an endpoint of e. Without loss of generality,

suppose d is connected to g by the edge uv'''. Now $v''' \in N(u)$ and $u'' \in N(v)$. Since

e is bisimplicial, $u''v'''$ is present and d is connected to h as well (see Figure 4.2.1).

To prove that condition 3 is necessary, once again choose a maximum connected

matching M. Suppose now that condition 3 does not hold, and there is (WLOG) an

edge $f = ux$ such that x is not covered by M. We claim that we can remove e and

replace it with f in M. We need only consider some edge g contained in M that is

not connected to e via u. Then there is an endpoint w of g so that $w \in N(v)$. Clearly

$x \in N(u)$, so wx is present and g is connected to f (see Figure 4.2.1). □
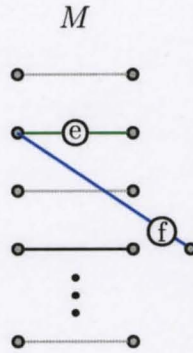
Figure 4.5: Illustration of proof of condition 3.

Taken altogether, these observations about bisimplicial edges allow us to reduce the maximum connected matching problem to a *saturating* connected matching problem. Presented with a separable chordal bipartite graph, we find two separable bisimplicial edges. If the neighborhood of either endpoint of one of these edges is not saturated by a connected matching, then the edge is inert and is removed. If both edges have connected matchings saturating their neighborhoods, we simply remove the edge of smaller degree. If they additionally have the same degree, either one may be removed. In this manner, we remove bisimplicial edges until the graph is rendered nonseparable. Having done so, any maximum matching algorithm will produce a maximum connected matching.

In short, we reduce Maximum Connected Matching to the following problem.

Saturating Connected Matching

Input: Bipartite graph $G = (A, B; E)$

Question: Is there a connected matching in G that saturates A?
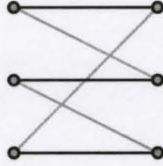
Figure 4.6: A connected matching that is not an ordered connected matching.

4.2.2. Ordered connected matchings

Now we explore ideas equivalent to connected matchings in chordal bipartite graphs. The first is a connected matching with extra requirements. Next, we define a family of graphs with perfect connected matchings. After that, we define a property of hypergraphs. Finally, we describe a different sort of matching problem that is in fact equivalent to connected matchings on this restricted class of graphs.

We define an ordered connected matching as follows. Let $M = \{e_1, e_2, \ldots, e_k\}$ be a connected matching in a bipartite graph $G = (A, B; E)$ with the endpoints of each e_i labeled with $a_i \in A$ and $b_i \in B$. If there is an ordering $\sigma : M \to [k]$ so that

1. $deg(a_i) \geq \sigma(e_i)$ for each $i \in [k]$ and

2. $N(a_l) \subseteq N(a_j)$ whenever $j \geq l$ for all $j, l \in [k]$

then we say M is an *ordered connected matching*. Note that if we find the appropriate sequence of vertices in A, there is neccessarily an appropriate sequence of vertices in B. Not every bipartite connected matching is an ordered connected matching, as we see in Figure 4.6. However, it is worth noting that this example is a cycle on six vertices.

Secondly, we define a family of graphs with perfect connected matchings. Let

40
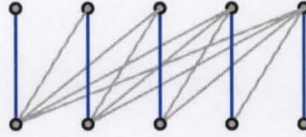
Figure 4.7: The graph L_5.

L_n be a balanced bipartite graph with $2n$ vertices in sets $A = \{a_1, a_2, \ldots, a_3\}$ and $B = \{b_1, b_2, \ldots, b_n\}$. Each vertex a_i in A is adjacent to precisely those vertices in B whose indices are less than or equal to i. Then then edges between vertices with equal indices form an ordered connected matching. In Figure 4.7 we see an illustration of L_5.

Now we consider a hypergraph $H = \{V(H), E(H)\}$. Let $S = \{S_i\}_{i=1}^{k}$ be a sequence of subsets of $V(H)$. If for each $i \leq k$ there is a distinct $E_i \in E(H)$ such that $S_i \subseteq E_i$, then we say that S is *dominated* by H. A *chain* is a sequence of sets $C_1, C_2, \ldots, C_k$ with the property that $C_i \subsetneq C_{i+1}$ for $1 \leq i < k$. In this section, we are interested in dominated chains in hypergraphs.

Finally, consider the following computational question:

Maximum Red Matching Free of Blue-Red Alternating Cycles (MR)

Given a complete bipartite graph $H = (X, Y; E)$ whose edges are partitioned into a set B of blue edges and a set R of red edges and whose vertices are covered by blue edges, what is a maximum matching M in the red component $(X \cup Y, R)$ such that the subgraph $(X, Y; B \cup M)$ has no blue-red alternatiing cycles?

The main result of this section is an equivalence theorem that provides many

41

different perspectives on connected matchings in chordal bipartite graphs. Before we state the theorem, we need to introduce two ideas. The first is the *biadjacency matrix* M_B of a bipartite graph $G = (A, B; E)$. This is the (0,1)-matrix formed by indexing rows with the vertices in A and indexing columns with the vertices in B. The element a_{ij} of M_B is 1 if the *ith* vertex of A is adjacent to the *jth* vertex of B, and zero otherwise.

The second notion we need to introduce is the *neigborhood hypergraph* generated by a set S of vertices from a graph G. This is a hypergraph defined on the vertices of G with an edge defined by the open neighborhood of each vertex from S.

Theorem 13. *Let $G = (A, B; E)$ be a chordal bipartite graph. The following statements are equivalent:*

1. *G has a connected matching of size k.*

2. *G has an ordered connected matching of size k.*

3. *G has a copy of L_k as a subgraph.*

4. *Tthe neighborhood hypergraph of A has a dominated chain of size k.*

5. *The biadjacency matrix M_B of the complement $\overline{G}$ of G has a submatrix that is permutation-equivalent to a strictly upper triangular matrix.*

6. *For a red/blue edge-colored bigraph $H = (X, Y; B \cup R)$, if G is the red graph H, then there is a matching M of size k in G with no alternating blue/red cycle in $(X, Y; B \cup M)$.*

Proof. First we show that statement 1 is equivalent to statement 2. We wish to prove that any connected matching M in a chordal bipartite graph $G = (A, B; E)$ has a vertex $a \in A$ so that a is covered by M and a is adjacent to the B-endpoints of every edge of M. We say that such a vertex *dominates M*. If this is true, we can find the proper ordering of M by sequentially removing the M-edge covering a dominating edge and working on the smaller connected matching that remains.

We proceed by induction on the size n of the connected matching. Small cases of $n = 1, 2$ are trivial, and the case of $n = 3$ is easily checked. Now suppose that every chordal bipartite connected matching of up to $n-1$ edges has a dominating vertex and let M be a connected matching of size n in a chordal bipartite graph $G = (A, B; E)$. Let H be the subgraph of G induced by vertices covered by M. Applying lemma 4, we may remove any bisimplicial edges of H that are not contained in M without reducing the size of M; neither does this introduce any new dominating vertices. Remove these edges until all remaining bisimplicial edges are contained in M.

The resulting graph is chordal bipartite, so there exists a bisimplicial edge $ab \in M$, with $a \in A$ and $b \in B$. If a is not a dominating vertex, then consider M_b, the connected matching contained in M that touches the neighbors of b excluding a. Some $a' \in A$ dominates M_b, which means that a' is adjacent to all non-neighbors of a. Furthermore, because the edge ab is bisimplicial, a' is adjacent to all of the *neighbors* of a as well. Hence, a' is a vertex that dominates M.

We have already noted that L_k contains a connected matching of size k. Hence we show now that the subgraph induced by the vertices covered by an ordered connected matching has L_k as a subgraph. This is enough to prove that statement 2 is equivalent

to statement 3. Given an ordered connected matching of size k in a bipartite graph G, with a labeling and edge ordering σ consistent with the definition of an ordered connected matching, relabel the vertices so that the endpoints of e_i are $a_{\sigma(e_i)}$ and $b_{\sigma(e_i)}$. Applying the neighborhood inclusion property of ordered connected matchings, it is easy to see that the vertices as labeled induce all the edges of L_k.

To show that statement 3 is equivalent to statement 4, let us consider the vertices of A included in the copy of L_k and their neighborhoods within L_k. These form the chain that is dominated by the neighborhoods of these vertices in G.

The complement of our copy of L_k gives us the submatrix of M_B that must be permutation-equivalent to a strictly upper triangular matrix. By the definition of L_k, there is a dominating vertex, and a sequence of vertices wherein each is nonadjacent to at most one more vertex than the last. By permuting the columns so that the row corresponding to the high degree (in G) vertex is at the bottom, and the rows going up correspond to the sequence we just described, we end up with a strictly upper triangular matrix.

We prove the equivalence of the MR problem to the connected matching problem in chordal bipartite graphs as follows. First, if there are no alternating blue/red four cycles in $(X, Y; B \cup M)$, then M is certainly a connected matching (in the red graph). Conversely, consider the copy of L_k contained in a k-connected matching. Any alternating cycle must avoid the "outer" edges, and the endpoints of those edges. Working inductively after removing those edges and vertices, we see that we have a new pair of "outer" edges that must be avoided. Hence, no alternating cycle is present. □

44

4.2.3. Convex graphs

The *convex* graphs are chordal bipartite graphs for which there is a vertex ordering of one side with the property that the neighborhoods of the other side form a collection of intervals. In this section, we will discuss an attempt at building an algorithm for finding maximum connected matchings in convex graphs that is best described in the language of hypergraphs.

Convex graphs can naturally be thought of as the bipartite graph representation of an interval hypergraph. As such, we inquire after connected matchings in convex graphs by considering the k-dominated chain problem for interval hypergraphs.

Proposition 4. *The largest set in a chain dominated by an interval hypergraph H can be chosen to be an interval in the ordering of $V(H)$ derived from the interval representation of H.*

This is obviously true as all sets in the chain are subsets of the interval dominating the top element in the chain. Due to this fact, and the fact that there are only polynomially many subintervals of a finite interval, it suffices to find a polynomial-time algorithm for the problem of finding a *perfect* dominated chain in interval hypergraphs. If this is possible, we need only check every interval for a perfect chain dominated by H.

Dominated k-chain

Input: Hypergraph H, positive integer k

$$\begin{array}{cccccc}
v_1 & v_2 & \boxed{v_3 \quad v_4} & v_5 & v_6 \\
v_1 & \boxed{v_2 \quad v_3 \quad v_4} & v_5 & v_6 \\
\boxed{v_1 \quad v_2 \quad v_3 \quad v_4} & v_5 & v_6 \\
v_1 & v_2 & \boxed{v_3 \quad v_4 \quad v_5 \quad v_6} \\
v_1 & v_2 & \boxed{v_3 \quad v_4 \quad v_5 \quad v_6} \\
v_1 & \boxed{v_2 \quad v_3 \quad v_4 \quad v_5} & v_6 \\
\boxed{v_1 \quad v_2 \quad v_3 \quad v_4 \quad v_5 \quad v_6}
\end{array}$$
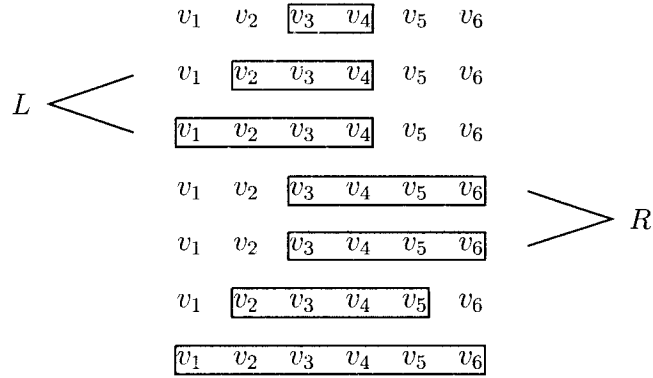
Figure 4.8: Diagram of an interval hypergraph. The cap is $\{v_3, v_4\}$.

> Question: Is there a chain of length k dominated by H?

Let H be an interval hypergraph. We call the intersection of all intervals $C = \bigcap_{e \in E(H)} e$ in an interval hypergraph the *cap* of the hypergraph. We are especially interested in two collections of intervals, $R = \{I \in H : I \geq C\}$, and $L = \{I \in H : I \leq C\}$. As in Figure 4.8, we think of these as the intervals that extend only to the left or right (respectively) of the cap.

Let us index the vertices of H $v_1, v_2, ...v_k$ according to an interval ordering, reading left to right. Suppose there is a chain $C_1, C_2, \ldots, C_k$ on $V(H)$ that is dominated by H. It must be the case that C_1 is a singleton, $C_k = V(H)$, and $|C_i| = |C_{i-1}| + 1$. Define a permutation $\sigma : [k] \to [k]$ so that

$$\sigma(i) = j \text{ when } v_i \in C_j \text{ and } v_i \notin C_{j-1}.$$

In other words, v_i is "added" at the $\sigma(i)$th step in the chain. Our chain is completely determined by σ. To dominate the chain, we suppose that there exists an indexing

46

$E_1, E_2, \ldots, E_k$ of the edges of H so that

$$\{v_i : \sigma(i) \leq j\} \subseteq E_j$$

We construct an algorithm that chooses vertices to add to the chain together with dominating intervals in a "step-by-step" manner. It will be easy to see that if the algorithm succeeds, its output must be a dominated perfect chain.

First, note that if $j > i$, and $E_j \subseteq E_i$, then the edge sequence with E_i and E_j interchanged still dominates the chain. That is, we can take an interval "too soon", so long as the interval is contained in the interval it is replacing. After interchanging the two intervals in the edge sequence, the vertex order σ is still intact and corresponds to the dominated chain.

Our algorithm proceeds by sequentially removing edges and vertices. Start with all of H, and attempt to choose a vertex contained in all k edges of H along with one of the edges of H to be designated v_1 and E_1. Then remove this edge and this vertex and do the same for the resulting hypergraph. If we can proceed until all edges and vertices have been chosen without encountering a hypergraph whose cap is empty, then clearly we have exhibited a perfect chain dominated by H. To complete an algorithm for efficiently finding a perfect dominated chain, we must make these choices in such a way that no perfect dominated chains are missed. As of this writing, we have not been able to describe such a method. Nonetheless, the following method avoids many of the pitfalls encountered in our search, and is a reasonable place to begin any further attempts.

We first claim that we can choose any vertex of C to be v_1. Suppose $u \in C$ is not v_1. Because u is in every edge of H, it can act as the first chain element. We then move v_1 to the place in the sequence of chose vertices formerly occupied by u without disturbing the chain. Thus, the first step in our algorithm is

> **Step 0.** *Choose any element of C to be v_1.*

The problem remains to select an edge of H to dominate v_1. By our earlier observation concerning edges that contain "later" edges, if $C \in E(H)$ then we can certainly choose C. So we suppose that $C \notin E(H)$.

> **Step 1.** *If $C \in E(H)$, let $E_1 = C$.*

Looking ahead, we see that we must choose all of R or L before C is exhausted. Until all elements of either R or L have been chosen, the mutual intersection of remaining edges is a subset of C. If C is exhausted, then only the empty subset remains and the algorithm terminates.

> **Step 2.** *If both R and L are smaller than C, then there is no dominated perfect chain.*

One of R or L will be used in its entirety when C is exhausted. We must ensure that the other collection can be used later. Subroutine 1 is designed to tell us if this is possible.

Subroutine 1.

Input: Sequence of vertices $u_l, u_{l-1}, \ldots, u_1, v_1, v_2, \ldots, v_k$; Set R of nested intervals all containing v_1 and not u_1; Set U of intervals containing v_1 and u_1.

1. Assign v_1 to the smallest unassigned R set.

2. If all R sets have been assigned, return **YES**.

3. If v_2 does not exist, return **NO**.

4. If v_2 is not in all sets, reassign v_1 to the smallest unassigned U set that contains v_1 but not v_2. Return **NO** if no such set exists.

5. Subtract 1 from the index of all v-vertices and go to 2.

For the next step, send R, $V(H) - C$ in order and with all u vertices less than C and all v vertices greater than C, and $U = E(H) - (L \cup R)$ as input to Subroutine 1. Then send L with the vertex order suitably reversed.

Step 3. *If both L and R return **NO** from Subroutine 1, there is no dominated perfect chain. If (WLOG) L returns **NO** and R returns **YES**, choose $\min L$ to be E_1.*

We show that if Subroutine 1 is satisfied for R and L, then we are free to choose either direction to build our chain. For brevity, we lean to the left.

Step 4. *If both L and R return YES from Subroutine 1, choose $\min L$ to be* E_1.

Finally, we iterate the process.

Step 5. *Remove v_1 and E_1 from H and begin at Step 0 with the resulting interval hypergraph.*

The hole in this algorithm lies in whether or not choosing L instead of R when both are safe will miss a perfect dominated chain. We have been unable to produce a counterexample or a proof.

5. MODELING WITH CONNECTED MATCHINGS

5.1. Bipartite graphs

In a typical bipartite matching problem, we are attempting to perform an *assignment.* We are given a collection A of one type of object, a collection B of a second type of object, and a set of feasible assignments of objects from B to each object in A. We assume that only one object may be assigned to another object. The problem is to choose from among these possible assignments a set of *actual* assignments that maximizes some desirable property.

In the case of *connected* matchings, we have added another requirement. The assignment must be done in such a way that among the objects in any two assigned pairs, there is some other, unused feasible assignment. This may model redundancy, flexibility, interconnectivity, proximity, or some other quality we wish to demand of the chosen matching.

5.1.1. Cloud administration

Let us draw up a hypothetical problem that highlights the above description. Suppose we are administering a cloud-based application. Broadly speaking, the architecture of the application consists of *servers* which store and process data and *clients* which

deliver content and receive commands from end users. At any given time, each client is logged in to the small subset of servers needed to carry out a particular task.

Suppose that we want to allow limited and moderated client-client communication. For privacy, security, or other reasons we do not wish to allow all clients logged in to a given server to send and receive messages from each other. In fact, we require that one side of any transmission be a superuser. This privileged client serves to moderate the communication. However, we do wish for *any* two clients logged into *any* of the servers to be able to get a message from one to the other as quickly as possible.

We begin with a matching problem: how do we assign a privileged (logged-in) client to each server? Each superuser will administer its assigned server, in addition to transmitting messages among clients logged into that server. By itself, this question provides us with a collection of superusers that can moderate all messages on any given server. Furthermore, each message passes through only one intermediary. This is close to what we want, but so far we cannot guarantee that users not mutually logged into any particular server can communicate.

To fully meet our requirements, we must ensure that there is a "safe" path between any two servers. To meet them and ensure rapid communication, we should ensure that any pair of superusers are both logged into some server and can thus pass messages. When we make an assignment that does so, we have actually found a *connected* matching. It is also not difficult to see that any connected matching in the server-client bigraph represents a satisfactory assignment of servers to superusers.

An interesting feature of this particular example is an asymmetry in scaling. As the application grows in users, each individual server machine will be able to serve

greater numbers of clients, provided that the software involved is well designed and scalable. If the number of servers remains essentially fixed as the application grows, then we will still be able to efficiently identify large connected matchings in the server graph.

5.1.2. The bipartite margin shop

Another application of connected matchings in bipartite graphs arises from the work of Oron, Steiner and Timkovsky in [28]. A large brokerage firm may manage tens of millions of customer accounts. Margining these accounts and producing account status slips at the end of each business day presents a serious computational challenge. Oron et. al model the computational task with a graph they term the *bipartite margin shop.*

A bipartite margin shop $G = (A, B; E)$ is a bipartite graph with vertices that represent tasks. Each task v has an associated processing time p_v. Each connected component of this graph is called a *job*. This graph is a *precedence graph* in that an edge ab, with $a \in A$ and $b \in B$, is present when task a must be completed no later than task b. This model is bipartite on the assumption that there are two machines at work, M_A and M_B. Tasks from A are processed on machine M_A, and tasks from B are processed on machine M_B. Interpreting this problem in terms of a brokerage firm margining accounts,we have an edge between the tasks a and b if the account b has a position in the security a. This means the "market" machine M_A must calculate the margin requirement for one unit of the security a before the "account" machine M_B can calculate the margin requirement for the account b.
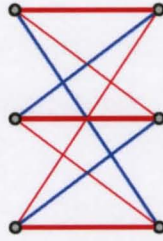
53

Figure 5.1: A red connected matching M with an alternating cycle of length six.

In the margin-shop scheduling problem, we take a bipartite margin shop and try compute to the schedule that completes each job in the minimum time without violating the precedence relationships encoded in the graph. The authors of [28] prove that the bipartite margin shop problem is actually equivalent to the MR problem discussed in Chapter 4.

This is very closely related to the problem of finding a connected matching in the graph of red edges. In fact, a matching as described above with no alternating blue-red cycles would certainly be a connected matching. As we proved in Chapter 4, for chordal bipartite graphs, the problems are equivalent. However for general bipartite graphs it is stronger than the notion of a connected matching. As we see in Figure 5.1, a connected matching in red edges could permit alternating cycles longer than four edges.

5.2. General graphs

Sometimes we may encounter a matching-type problem with only one type of object modeled by nodes. In this case, we cannot assume that the appropriate graph model is bipartite. The relation we wish to model with edges in the graph is assumed to be non-

transitive, as otherwise the graph model degenerates into a collection of disconnected cliques.

5.2.1. Partnership assignments

One type of "assignment between equals" we encounter in the real world is that of working partnerships. In any organization, we can model individuals as vertices in a graph and use edges to represent some type of advantageous (non-transitive) working relationship. By identifying a connected matching, we collect a set of possible partnerships wherein each pair of partners has the advantageous relationship and any two pairs A and B have an indvidual from pair A that has the advantageous relationship with some individual from pair B,

Let us see how this might work in a high school mathematics classroom. Suppose we model the students in the class as vertices. We are interested in the opportunities students have to work together outside of class, so we place an edge between vertices if the corresponding students share a study hall, lunch period, or live within a block of one another. We want to assign a project to groups of two students each, and would like for the students to have time to work on the project outside of class. Assigning groups so that each group corresponds to an edge in the graph we have constructed is a matching problem, and the result is a collection of pairs of students that have opportunities to work together outside of class.

Now let's suppose that the content of each pair's assigned project is a piece of a larger class project. In light of this, we would like to additionally require that each pair of two-student groups has a chance to share their findings and collaborate outside

of class. In order for this to happen, the matching must also be connected. That is to say, in any two groups A and B, there is a student from A who shares a study hall, lunch period, or neighborhood with a student from B.

6. RELATED PROBLEMS

.

In the course of studying connected matchings, some problems have arisen that are quite interesting in their own right. While these approaches did not lead to results related to connected matchings, they are closely related to this investigation in their own ways, and may point the way to further research. The first is a notion related to the distance approach to detecting connected matchings. The second offers a systematic way of "reducing" a chordal bipartite graph.

6.1. Characterizing the "connected matching graph"

A graph G induces a proximity partition described by the collection of *distance-k graphs* of G. We may ask if there is a characterization of $\mathcal{H}_k$ where

$$\mathcal{H}_k = \{H : H \text{ is the distance-}k \text{ graph of some graph } G\}$$

In the case of $\mathcal{H}_2$, we can do so. Let $A(G)$ denote the adjacency matrix of a graph G. In [27] Mukhopdhyay characterizes graphs that have a *square root*, which is to say graphs H such that for some graph G, $A(H) = A(G)^2$. This is equivalent to H possessing an edge between any pair of vertices u, v that satisfy $d_G(u, v) \leq 2$.

Theorem 14 (Mudkhopdhyay). *A connected graph G with n vertices $v_1, v_2, \ldots, v_n$ has a square root if and only if some set of n complete subgraphs of G whose union is G can be labeled $C_1, C_2, \ldots, C_n$ so that, for all $i, j = 1, 2, \ldots, n$ the following conditions hold:*

1. C_i contains v_i,

2. C_i contains v_j if and only if $C_{i,j}$ contains v_i.

An alternate definition of $\mathcal{H}_2$ is

$$\mathcal{H}_2 = \{H : A(H) = A(G)^2 - A(G)\}$$

and we have the following characterization in the spirit of Mukhopdhyay's theorem.

Theorem 15. *A graph G with n vertices $v_1, v_2, \ldots, v_n$ is the distance 2 graph of some graph H if and only if some set of n subgraphs of G whose union is G can be labeled $C_1, C_2, \ldots, C_n$ so that*

1. $v_a \notin C_a$

2. For every pair of vertices $v_i, v_j \in C_k$, exactly one of the following holds:

(a) $v_i v_j \in E(C_k)$

(b) If $v_i \in V(C_j)$, then $v_j \in V(C_i)$

(c) $v_i \in C_j$ and $v_j \in C_i$

3. If $C_i \cap C_j \neq \emptyset$, then $v_i, v_j \in V(C_k)$ for some k.

The proof of this theorem rests on the same central realization as Mudkhopdhay's theorem, which is that the enumerated subgraphs correspond to the neighborhoods (closed neighborhoods in Mudkhopdhay's theorem, open ones in Theorem 15) in the underlying graph.

Proof. Suppose we have a graph G on vertices $v_1, v_2, \ldots, v_n$ and subgraphs $C_1, C_2, \ldots, C_n$ that satisfy the above conditions. We construct a graph H on the same vertex set by adding edges in two steps for each C_i.

Step 1. Add the complement of C_i.

Step 2. Add all edges from v_i to C_i.

We claim that G is now the distance 2 graph of H. Suppose that $d_H(v_i, v_j) = 2$. We want to show that $v_i v_j \in E(G)$. Since $d_H(v_i, v_j) \leq 2$, there is some vertex v_k so that $v_i v_k, v_j v_k \in E(H)$. If both edges were added in step 2, then one of the following occurs

Case 1. $v_i, v_j \in C_k$

Case 2. $v_k \in C_i$, $v_k \in C_j$

Case 3. $v_k \in C_i, C_j$

In case 1, distance 2 implies $v_i v_j \notin E(H)$. In particular, this edge was not added in step 2, so $v_i \notin C_j$ and $v_j \notin C_i$. Condition 2 then implies that if $v_i, v_j \in C_k$ for some C_k, then $v_i v_j \in E(G)$. In case 2, $V(C_i)$ and $V(C_j)$ intersect, implying (by condition 3) again that there is a $V(C_l)$ containing both v_i and v_j. In case 3, condition 4 requires that $v_i, v_j \in C_k$. In any event, $v_i v_j \in E(G)$.

Now we may assume (WLOG) that $v_i v_k$ was added to H in step 1. Suppose $v_j v_k$ was added in step 2. Then either $v_k \in C_j$, implying $C_i \cap C_j \neq \emptyset$, or $v_j \in C_k$ implying $v_i, v_j \in C_k$. The only remaining possibility is that both $v_i v_k$ and $v_j v_k$ were added in step 1. Following from two applications of condition 2(b), both v_i and v_j are then in $V(C_k)$. This completes the proof of sufficiency.

For the proof of necessity, we take a graph H and show that the distance-two graph $D_2(H)$ has a collection of subgraphs with the necessary properties. For each vertex v_i, let $V(C_i) = N(v_i)$. That condition 1 holds is immediate. The vertices of any given C_i are at most distance two apart. Whenever there is a nonedge in a particular C_i and 2(a) fails, the vertices must be adjacent in H, and condition 2(b) holds. The symmetric property of the neighbor relation shows that conditions 3 and 4 hold as well. Finally, all distance 2 edges occur between vertices with a common neighbor, so $\bigcup C_i = G$. $\qquad\qquad\square$

In a distance 2 graph, the enumerated subgraphs are actually the *complements* of the graphs induced by the open neighborhoods of the underlying graphs. This allows us to characterize the distance 2 graphs of graphs with local characterizations. The following is an example.

Proposition 5. *If every enumerated subgraph C_i of a distance 2 graph H has independence number 2, then it is the distance 2 graph of a claw-free graph.*

In the case of connected matchings, we are most interested in the distance 2 graphs of line graphs. Unfortunately, line graphs do not have a local characterization. The best local characterization that approximates them is the *locally co-bipartite* graphs.

A graph is locally co-bipartite if each open neighborhood induces a graph whose complement is bipartite. Hence, the following is the best we have in the direction of characterizing the distance 2 graphs of line graphs.

Proposition 6. *If every enumerated subgraph C_i of a distance 2 graph H is bipartite, then it is the distance 2 graph of a locally co-bipartite graph.*

6.2. Totally balanced hypergraphs, basic trees, and chordal bipartite graphs.

In chapter 4, we discussed the chordal bipartite graphs at some length. There is a characterization of chordal bipartite graphs that we have not yet discussed. The *incidence bigraph* of a hypergraph H is a bipartite graph with one partite set comprised of the edges of H, and the other comprised of the vertices of H. We draw an edge from v to e whenever $v \in e$ in H. The chordal bipartite graphs are characterized as the incidence bigraphs of a special class of hypergraphs known as *totally balanced* hypergraphs.

A *cycle*[1] in a hypergraph is an analog of the graph-theoretic notion of a cycle. The vertices $v_1, v_2, \ldots, v_k$ of of a hypergraph H form a cycle if the pairs

$$v_1 v_2, v_2 v_3, \ldots, v_{k-1} v_k, v_k v_1$$

are all adjacent in H. A hypergraph is totally balanced if and only if every cycle of length greater than two has an edge containing three vertices of the cycle. It is not

---

[1]This is not the only defintion of a cycle in the theory of hypergraphs. This particular definition is a generalization of cycles in graphs, but the reader should be aware of other definitions in the literature.

difficult to see that through the lens of an incidence bigraph, this is equivalent to requiring that cycles of length 6 have a chord.

This connection between totally balanced hypergraphs and chordal bipartite graphs led to a new line of inquiry growing out of work by Lehel. In [25], Lehel describes a reduction process achievable on totally balanced hypergraphs. Totally balanced hypergraphs are a special case of *tree* hypergraphs. A tree hypergraph is one for which there is a *basic tree*. A basic tree is a tree T_H on the vertices of the hypergraph with the property that every edge of H induces a subtree of T_H.

Let us describe Lehel's method of "reducing" a tree hypergraph modulo a basic tree. For a hypergraph H and a choice of basic tree T, let the edges of T be the vertices of a new hypergraph H/T. For each edge e of H, we construct an edge e' of H/T by collecting all of the edges of T for which both endpoints are in e.

Unfortunately, this line of investigation did not yield substantial results on the question of computing connected mathings in chordal bipartite graphs. However, a key property of Lehel's reduction vis a vis basic trees was uncovered, as well as an interesting tree mapping which we describe below.

A (simple) *enveloping* of a graph $G = (V, E(G))$ into a graph $H = (V, E(H))$ is a one-to-one mapping $\phi : E(G) \rightarrow E(H)$ with the property that for every $uv \in E(G)$, $\phi(uv)$ is on a u, v path in H.

Theorem 16. *For any pair of trees T_1 and T_2 on a vertex set V, there is a simple enveloping ϕ of T_1 into T_2.*

Proof. We work by induction on the number of vertices in V. For the case of $|V| = 2$,

an identity map is an enveloping.

Now suppose we have an enveloping between trees on up to k vertices. Let T_1 and T_2 be a pair of trees on $k+1$ vertices. Start by taking any pendant edge uv of T_1 with leaf vertex u, and mapping it to uv' where v' is the next vertex on the u, v path in T_2. Now remove u from T_1 and identify u and v' as v' in T_2. We now have two trees $T_1 - u$ and T_2' on a vertex set $V' = V - \{u\}$ with k vertices. Hence there is an enveloping ϕ of $T_1 - \{u\}$ into T_2'. Furthermore, ϕ can be extended as follows to an enveloping ϕ' of T_1 into T_2:

1. uv maps to uv' as above.

2. If $\phi(x, y) = wv'$ then either wu or wv' is an edge of T_2. Set $\phi'(x, y) = wu$ or wv' as appropriate.

$\square$

This theorem tells us that between any two trees on a set of vertices, one can be mapped onto the other in such a way that edges are mapped into the paths defined by their endpoints. Tree envelopings may be a topic worthy of investigation in their own right.

Using this result on envelopings, we prove that the reduction proposed by Lehel is insensitive to the choice of basic tree.

Theorem 17. *If T_1 and T_2 are basic trees of a tree hypergraph H, then*

$$H/T_1 \simeq H/T_2$$

Proof. We show that an enveloping ϕ of T_1 into T_2 induces an isomorphism between H/T_1 and H/T_2.

Suppose first that $uv \in N(e_i)$ in H/T_1. Then $\{u, v\} \subseteq e_i$ in H. Since e_i then induces a subtree of T_2, the vertices of u, v path of T_2 must be in e_i. Hence, the endpoints of $\phi(uv)$ are in e_i, and $\phi(uv) \in N(e_i)$ in H/T_2.

Now we need to show that $|N(e_i)|$ in H/T_1 is the same as $|N(e_i)|$ in H/T_2. This is as simple as noting that the neighborhood of e_i induces a subtree in both T_1 and T_2, and both of these subtrees have $|e_i| - 1$ edges. Hence $|N(e_i)| = |e_i| - 1$ in both H/T_1 and H/T_2. $\qquad \square$

6.3. Conclusion

We have laid a great deal of groundwork in computing maximum connected matchings in certain families of bipartite graphs. With additional attention, it is likely this can provide the basis for determining the complexity of finding a maximum connected matching in convex or chordal bipartite graphs.

We have also demonstrated the equivalence of the maximum connected matching problem to certain cases of an important practical optimization problem: the bipartite margin shop. It would be interesting to see how well chordal bipartite graphs approximate the actual account/security relationships in investment firms. Conversely, it would be also be worthwhile to investigate how closely a graph must approximate a chordal bipartite graph (whether this is measured in edit distance, or total number

of long cycles) in order to apply the results found herein.

In pursuit of the extremal question, we have determined that counterexamples to the key conjecture of Gyárfás, Füredi and Simonyi (Conjecture 5) have both high connectivity and no large cliques. Furthermore, we have determined that any infinite family of counterexamples must not have the *smallest* possible clique numbers for graphs with independence number two. This excludes from consideration the most probable graphs arising from the triangle-free process.

The study of connected matchings is alive and well, and it is our hope that this dissertation will be helpful in moving this area forward. Many appealing computational problems remain, and the close connection to Hadwiger's conjecture is a fascinating aspect of the structural study of connected matchings. We look forward to investigating these ideas in the future.

REFERENCES

[1] Martin Aigner and Günter M. Ziegler, *Proofs from The Book*, 4th ed., Springer-Verlag, Berlin, 2010. MR2569612 (2010m:00001)

[2] József Balogh and Alexandr V. Kostochka, *Large minors in graphs with given independence number*, Discrete Math. **311** (2011), no. 20, 2203–2215, DOI 10.1016/j.disc.2011.07.003. MR2825665 (2012i:05275)

[3] Jonah Blasiak, *A special case of Hadwiger's conjecture*, J. Combin. Theory Ser. B **97** (2007), no. 6, 1056–1073, DOI 10.1016/j.jctb.2007.04.003. MR2354718 (2009a:05064)

[4] Tom Bohman, *The triangle-free process*, Adv. Math. **221** (2009), no. 5, 1653–1677, DOI 10.1016/j.aim.2009.02.018. MR2522430 (2010h:05271)

[5] Béla Bollobás, Paul Catlin, and Paul A. Erdős, *Hadwiger's conjecture is true for almost every graph*, European J. Combin. **1** (1980), no. 3, 195-199. MR593989 (82b:05107)

[6] Kathie Cameron, *Connected matchings*, Combinatorial optimization—Eureka, you shrink!, Lecture Notes in Comput. Sci., vol. 2570, Springer, Berlin, 2003, pp. 34–38. MR2163948 (2006c:90072)

[7] _____, *Induced matchings*, Discrete Appl. Math. **24** (1989), no. 1-3, 97–102, DOI 10.1016/0166-218X(92)90275-F. First Montreal Conference on Combinatorics and Computer Science, 1987. MR1011265 (90g:05139)

[8] Paul A. Catlin, *Hajós' graph-coloring conjecture: variations and counterexamples*, J. Combin. Theory Ser. B **26** (1979), no. 2, 268–274, DOI 10.1016/0095-8956(79)90062-5. MR532593 (81g:05057)

[9] Maria Chudnovsky, Neil Robertson, Paul Seymour, and Robin Thomas, *The strong perfect graph theorem*, Ann. of Math. (2) **164** (2006), no. 1, 51–229, DOI 10.4007/annals.2006.164.51. MR2233847 (2007c:05091)

[10] Gabriel A. Dirac, *A property of 4-chromatic graphs and some remarks on critical graphs*, J. London Math. Soc. **27** (1952), 85–92. MR0045371 (13,572f)

[11] Pierre Duchet and Henry Meyniel, *On Hadwiger's number and the stability number*, Graph theory (Cambridge, 1981), North-Holland Math. Stud., vol. 62, North-Holland, Amsterdam, 1982, pp. 71–73. MR671905 (84h:05074)

[12] Jack Edmonds, *Paths, trees, and flowers*, Canad. J. Math. **17** (1965), 449–467. MR0177907 (31 #2165)

[13] Jacob Fox, *Complete minors and independence number*, SIAM J. Discrete Math. **24** (2010), no. 4, 1313–1321, DOI 10.1137/090766814. MR2735925 (2011m:05293)

[14] Alexandra Fradkin, *Clique minors in claw-free graphs*, J. Combin. Theory Ser. B **102** (2012), no. 1, 71–85, DOI 10.1016/j.jctb.2011.04.005. MR2871768 (2012m:05332)

[15] Zoltán Füredi, András Gyárfás, and Gábor Simonyi, *Connected matchings and Hadwiger's conjecture*. Problem Section, Combin. Probab. Comput. **14** (2005), 435–438.

[16] Michael R. Garey and David S. Johnson, *Computers and intractability*, W. H. Freeman and Co., San Francisco, Calif., 1979. A guide to the theory of NP-completeness; A Series of Books in the Mathematical Sciences. MR519066 (80g:68056)

[17] Martin Charles Golumbic, *Algorithmic graph theory and perfect graphs*, 2nd ed., Annals of Discrete Mathematics, vol. 57, Elsevier Science B.V., Amsterdam, 2004. With a foreword by Claude Berge. MR2063679 (2005e:05061)

[18] Martin Charles Golumbic and Clinton F. Goss, *Perfect elimination and chordal bipartite graphs*, J. Graph Theory **2** (1978), no. 2, 155–163, DOI 10.1002/jgt.3190020209. MR493395 (80d:05037)

[19] András Gyárfás, Miklós Ruszinkó, Gábor N. Sárközy, and Endre Szemerédi, *One-sided coverings of colored complete bipartite graphs*, Topics in discrete mathematics, Algorithms Combin., vol. 26, Springer, Berlin, 2006, pp. 133–144. MR2249267 (2008c:05120)

[20] Hugo Hadwiger, *Über eine Klassifikation der Streckenkomplexe*, Vierteljschr. Naturforsch. Ges. Zürich **88** (1943), 133–142 (German). MR0012237 (6,281c)

[21] Ken-ichi Kawarabayashi, Michael D. Plummer, and Bjarne Toft, *Improvements of the theorem of Duchet and Meyniel on Hadwiger's conjecture*, J. Combin. Theory Ser. B **95** (2005), no. 1, 152–167, DOI 10.1016/j.jctb.2005.04.001. MR2156345 (2006b:05118)

[22] Andreĭ Kotlov, *Matchings and Hadwiger's conjecture*, Discrete Math. **244** (2002), no. 1-3, 241–252, DOI 10.1016/S0012-365X(01)00087-5. Algebraic and topological methods in graph theory (Lake Bled, 1999). MR1844036 (2002k:05087)

[23] Jeong Han Kim, *The Ramsey number $R(3,t)$ has order of magnitude $t^2/\log t$*, Random Structures Algorithms **7** (1995), no. 3, 173–207, DOI 10.1002/rsa.3240070302. MR1369063 (96m:05140)

[24] Matthias Kriesell, *On Seymour's strengthening of Hadwiger's conjecture for graphs with certain forbidden subgraphs*, Discrete Mathematics (2010), 435–438.

[25] Jenő Lehel, *A characterization of totally balanced hypergraphs*, Discrete Math. **57** (1985), no. 1-2, 59–65, DOI 10.1016/0012-365X(85)90156-6. MR816048 (87f:05123)

[26] Frank Maffray and Henry Meyniel, *On a relationship between Hadwiger and stability numbers*, Discrete Math. **64** (1987), no. 1, 39–42, DOI 10.1016/0012-365X(87)90238-X. MR882610 (88g:05076)

[27] Amar Mukhopadhyay, *The square root of a graph*, J. Combinatorial Theory **2** (1967), 290–295. MR0210616 (35 #1502)

[28] Daniel Oron, George Steiner, and Vadim G. Timkovsky, *The bipartite margin shop and maximum red matchings free of blue-red alternating cycles*, Discrete Optim. **6** (2009), no. 3, 299–309, DOI 10.1016/j.disopt.2009.03.001. MR2532467 (2010g:90052)

68

[29] Anders Sune Pedersen and Bjarne Toft, *A basic elementary extension of the Duchet-Meyniel theorem*, Discrete Math. **310** (2010), no. 3, 480–488, DOI 10.1016/j.disc.2009.03.023. MR2564800 (2011c:05320)

[30] Michael D. Plummer, Michael Stiebitz, and Bjarne Toft, *On a special case of Hadwiger's conjecture*, Discuss. Math. Graph Theory **23** (2003), no. 2, 333–363. MR2070161 (2005e:05055)

[31] Frank P. Ramsey, *On a Problem of Formal Logic*, Proc. London Math. Soc. **S2-30**, no. 1, 264, DOI 10.1112/plms/s2-30.1.264. MR1576401

[32] Neil Robertson, Paul Seymour, and Robin Thomas, *Hadwiger's conjecture for K_6-free graphs*, Combinatorica **13** (1993), no. 3, 279–361, DOI 10.1007/BF01202354. MR1238823 (94i:05037)

[33] Bjarne Toft, *A survey of Hadwiger's conjecture*, Congr. Numer. **115** (1996), 249–283. Surveys in graph theory (San Francisco, CA, 1995). MR1411244 (97i:05048)

[34] Paul Turán, *Eine Extremalaufgabe aus der Graphentheorie*, Mat. Fiz. Lapok **48** (1941), 436–452 (Hungarian, with German summary). MR0018405 (8,284j)

[35] Douglas B. West, *Introduction to graph theory*, Prentice Hall Inc., Upper Saddle River, NJ, 1996. MR1367739 (96i:05001)

INDEX

CURRICULUM VITAE

NAME: Christopher Caraganis

ADDRESS: 3913 Southern Pkwy
                 Louisville, KY 40214

DOB: Louisville, KY - August 4, 1980

EDUCATION
& TRAINING   B.S., Mathematics
                 University of Louisville
                 2004-2007

                 M.A., Mathematics
                 University of Louisville
                 2007-2009

                 Ph.D., Mathematics
                 Univerity of Louisville
                 2009-2012

AWARDS: Logistics and Distribution Institute
                 Graduate Research Assistantship
                 2010

                 College of Arts and Sciences Outstanding Graduate Award
                 2009

                 Graduate Dean's Citation
                 2009