

**The edit distance function for graphs: an exploration of the case of forbidden
induced $K_{2,t}$ and other questions**

by

Tracy J. McKay

A dissertation submitted to the graduate faculty
in partial fulfillment of the requirements for the degree of
DOCTOR OF PHILOSOPHY

Major: Mathematics

Program of Study Committee:

Ryan Martin, Major Professor

Maria Axenovich

Oliver Eulenstein

Leslie Hogben

Sung Song

Iowa State University

Ames, Iowa

2012

DEDICATION

To my grandfather (Richards).

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	v
ABSTRACT	vi
CHAPTER 1. INTRODUCTION	1
1.1 Definitions	2
1.1.1 Simple graph terminology and notation	2
1.1.2 Terminology and notation specific to the problem	4
1.2 Select review of the literature and past results	7
1.3 Results	13
CHAPTER 2. PROOFS FOR RESULTS FROM [34]	16
2.1 Preliminary results and observations	16
2.2 Proof of Theorem 18	19
2.3 Proof of Theorem 19	22
2.3.1 Upper bounds	22
2.3.2 Lower bounds	23
2.4 Proofs of Theorems 20, 21 and 22	32
2.4.1 An extension of the known interval for $K_{2,t}$	32
2.4.2 A construction for odd t	33
2.4.3 A general lower bound for t	34
2.5 Upper bound constructions	36
2.5.1 Results from strongly regular graph constructions	36
2.5.2 Cycle construction	37
2.5.3 Füredi constructions	38

2.6	Figures	42
CHAPTER 3. MORE ON CYCLE POWERS AND FÜREDI'S CONSTRUCTION		47
3.1	General cycle constructions	47
3.1.1	The best k for a given value of r	47
3.1.2	Why $r = 2$ is optimal	48
3.2	Füredi's construction from [29] described	49
CHAPTER 4. CONCLUSIONS AND FUTURE DIRECTIONS		50
BIBLIOGRAPHY		52

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my adviser Ryan Martin for all of his guidance, insight, and encouragement throughout the process of completing this thesis. I would also like to thank all of the members of my POS Committee for their help, support, and valuable advice. In particular, I would like to thank Prof. Hogben for her feedback on the manuscript of this thesis, Prof. Song for important discussions about strongly regular graphs (including directing me to Andries Brouwer's website), and Maria Axenovich for pointing me in the direction of Zolán Füredi's $K_{2,t}$ -free constructions. Ed Marchant, Andrew Thomason, and Stephen Hartke also contributed helpful conversations and information to my adviser, which he, in turn, passed along to me.

In addition, I would like to thank the Mathematics Department at Iowa State for giving me the chance to participate in so many enriching opportunities during my time here, and for both travel support and a semester of research on the Wolfe Research Fellowship.

Last, but *certainly* not least, I would like to thank all of the office staff for everything that they have done for me along the way.

ABSTRACT

This thesis examines the edit distance function for principal hereditary properties of the form $\text{Forb}(K_{2,t})$, the hereditary property of graphs containing no induced bipartite subgraph on 2 and t vertices. It explores applications of several methods from the literature for determining these edit distance functions, and also constructions from classical graph theory problems that can be used to create colored regularity graphs leading to upper bounds on the functions. Results include the entire edit distance function when $t = 3$ and 4, as well as bounds for larger values of t , including the result that the maximum value of the function occurs over a nondegenerate interval of values for odd t .

CHAPTER 1. INTRODUCTION

This thesis explores results for the edit distance function for graph properties of the form $\text{Forb}(K_{2,t})$. A simple graph G is described by a 2-tuple $(V(G), E(G))$, where $V(G)$ is the set of vertices of G , and $E(G) \subseteq V(G)^{(2)}$ is the set of edges of G ($V(G)^{(2)}$ denotes the subsets of $V(G)$ with order 2). Given two simple graphs G and G' on the same vertex set V , the edit distance between G and G' , denoted $\text{dist}(G, G')$, is the order of the symmetric difference between $E(G)$ and $E(G')$. Perhaps a more intuitive way to calculate edit distance is to ask the question “How many edges would I need to add or delete in G to turn it into G' ?” Edit distance can be extended to measure the distance from a single graph G to an entire set of graphs, or graph property, in a similar way. The edit distance function for a specific graph property explores the asymptotic (in terms of the order of $V(G)$) behavior of the maximum edit distance of any graph G with edge-density p from the property.

Problems involving adding and/or deleting edges from graphs are called edge-modification problems. Edge-modification problems have a number of potential applications to chemistry, biology, and the social sciences. These problems were one motivation for the initial exploration of the edit distance function (see [5] and [7]). However, determining the edit distance function for any single property can pose unique challenges, which rely on connections to classical graph theory problems, that are interesting in their own right.

This thesis will explore some of the work that has already been done regarding edit distance and the edit distance function, and then describe techniques and constructions used to generate the new results for the properties $\text{Forb}(K_{2,t})$, defined explicitly below. Most of these results also appear in the submitted paper [34], which is joint work with Ryan Martin, though we will also explore some of the constructions in more depth.

1.1 Definitions

We begin with a more rigorous description of key terms and definitions, so that the problem and results (past and present) can be more precisely described.

1.1.1 Simple graph terminology and notation

What follows is a summary of standard definitions and terminology for simple graphs that will be used throughout the thesis.

As mentioned above, a **simple graph** $G(V, E)$, or simply G , is defined by a set of vertices $V(G)$ and edges $E(G)$, where each edge in $E(G)$ is a two element subset of $V(G)$. Note that this definition excludes loops (an edge from a vertex to itself) and multiple edges between the same pair of vertices, and that all edges in a simple graph are undirected. Two vertices $v_1, v_2 \in V(G)$ are said to be **adjacent** if $\{v_1, v_2\} \in E(G)$; otherwise, they are said to be **nonadjacent**. For a vertex $v \in V(G)$, any vertex that is adjacent to v in G is called a **neighbor** of v , and the set of all vertices that are adjacent to v is referred to as the neighborhood of v . A subset of vertices in G that are all adjacent to each other is referred to as a **clique**, and likewise, a set of vertices in G such that no two are adjacent is referred to as a **coclique**. If $V(G)$ is a clique for a graph G , then G is said to be a complete graph. The complete graph on n vertices is denoted by K_n .

If the vertices of a graph G with $|V(G)| = n$ are labeled $1, \dots, n$, then the **adjacency matrix** of G is $A = [a_{ij}]$ so that $a_{ij} = 0$ or 1 if vertices i and j are nonadjacent or adjacent, respectively.

The Erdős-Rényi **random graph** [25], denoted $G(n, p)$, is a random graph with n vertices and each pair of vertices adjacent with probability p . Random graphs play an important role in the exploration of edit distance and the edit distance function as well as other problems in graph theory, and both [10] and [30] are cited in the literature as general sources on the topic.

Two labeled graphs G and G' are said to be **isomorphic** if there exists a one-to-one mapping ϕ from the vertices of G to the vertices of G' such that if $\phi(v) \rightarrow v'$ and $\phi(u) \rightarrow u'$ then $\{u, v\} \in E(G)$ if and only if $\{u', v'\} \in E(G')$.

For a graph G , each subset of the vertices of G , $V' \subseteq V(G)$, defines an **induced subgraph** H of G with vertex set $V(H) = V'$ and edge set $E(H) = \{e \in E(G) : e \subseteq V'\}$. A **(weak) subgraph** of G on the vertex set V' has some edge set $E' \subseteq \{e \in E(G) : e \subseteq V'\}$. At times, it is convenient to refer to the (usually infinite) set of graphs that do not contain a specific induced subgraph, H . Such a set is denoted $\text{Forb}(H)$ because H is said to be a **forbidden induced subgraph** of the set.

Certain simple graphs are especially relevant to our problem. These are

- **Complete Bipartite Graphs:** A graph G is a complete bipartite graph if its vertices can be partitioned into two sets V_1 and V_2 , so that $E(G) = \{\{v_1, v_2\} : v_1 \in V_1 \text{ and } v_2 \in V_2\}$. The complete bipartite graph for which $|V_1| = m$ and $|V_2| = n$ is denoted by $K_{m,n}$.
- **Cycles and Powers of Cycles:** The cycle on n vertices, $\{1, \dots, n\}$, has edge set $\{\{i, j\} : i - j = \pm 1 \pmod n\}$. We denote the cycle on n vertices by C_n , and C_n^r will denote the cycle on n vertices to the r th power, which has the same vertex set and edge set $\{\{i, j\} : i - j = \pm r' \pmod n, \text{ where } 0 < r' \leq r\}$.
- **Books** ([18]): A (triangular) book graph is an n vertex graph that consists of two adjacent vertices $\{v_1, v_2\}$ (considered to be the spine of the “book”) and $n - 2$ other vertices that are adjacent only to v_1 and v_2 (the “pages” of the book). A book with r pages is denoted by B_r .

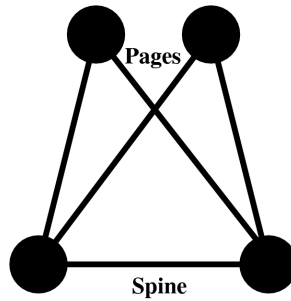


Figure 1.1 The book graph B_2 .

- **Strongly Regular Graphs:** A (k, d, λ, μ) -strongly regular graph is a graph with k vertices such that each vertex has degree d , each pair of adjacent vertices has exactly

λ neighbors in common, and each pair of nonadjacent vertices has exactly μ common neighbors. It should be noted that this notation for parameters is somewhat unorthodox, but was selected to be consistent with notation for colored regularity graphs discussed later in this section.

Certain graphs constructed by Zoltán Füredi [29] will also play a key role in our exploration the edit distance function later in this paper. These graphs are discussed in Sections 2.5.3 and 3.2.

1.1.2 Terminology and notation specific to the problem

The edit distance functions explore the asymptotic maximum normalized edit distance from specific graph properties. In particular, they are well-defined for a subset of graph properties called hereditary graph properties. A **hereditary graph property**, \mathcal{H} , is a set of graphs that is closed under both isomorphism and vertex deletion. Any hereditary graph property can always be described by a (possibly infinite) set of **forbidden induced subgraphs**. A convenient subset of hereditary properties is the properties that can be completely defined by forbidding a single induced subgraph. These are called **principle hereditary properties**, and as mentioned above, each may be denoted by $\text{Forb}(H)$, where H is the forbidden induced subgraph that completely defines the property. The results in this thesis pertain to the principle hereditary properties described by $\text{Forb}(K_{2,t})$ (e.g., $\text{Forb}(K_{2,3})$, $\text{Forb}(K_{2,4})$, $\text{Forb}(K_{2,5})$, etc.).

The edit distance between two graphs G and G' may be described as the minimum number of edge changes (either additions or deletions) necessary to make a graph G the same as G' , and this idea can be extended to describe the edit distance from G to an entire set of graphs. Here we rigorously define these notions of graph edit distance.

Definition 1 (Edit Distance). *Let G and H be simple graphs on the same labeled vertex set, and let \mathcal{H} be a hereditary property, then*

1. $\text{dist}(G, H) = |E(G) \Delta E(H)|$ is the edit distance from G to H ,
2. $\text{dist}(G, \mathcal{H}) = \min\{\text{dist}(G, H) : H \in \mathcal{H} \text{ and } V(G) = V(H)\}$ is the edit distance from G to \mathcal{H} and

3. $\text{dist}(n, \mathcal{H}) = \max\{\text{dist}(G, \mathcal{H}) : |G| = n\}$ is the maximum edit distance from the set of all n -vertex graphs to the hereditary property \mathcal{H} .

The maximum possible edit distance between two labeled graphs on the same set of n vertices is $\binom{n}{2}$, and so when exploring the asymptotic behavior as $n \rightarrow \infty$ of $\text{dist}(n, \mathcal{H})$ it makes sense to normalize by this factor. Thus, $\text{dist}_{\text{norm}}(\cdot, *) = \text{dist}(\cdot, *) / \binom{n}{2}$ from the definition above. We are now ready to define the edit distance function.

Definition 2 (Edit Distance Function). *The **edit distance function** of a hereditary property \mathcal{H} is a function of $p \in [0, 1]$ and is defined as follows:*

$$ed_{\mathcal{H}}(p) = \lim_{n \rightarrow \infty} \max \left\{ \text{dist}_{\text{norm}}(G, \mathcal{H}) : |V(G)| = n, |E(G)| = \lfloor p \binom{n}{2} \rfloor \right\}.$$

That the limit in this definition of $ed_{\mathcal{H}}(p)$ exists is not obvious, but has been shown (see [5, 9]). It can be challenging (if not impossible with the techniques known today) to find the edit distance function for a given hereditary property. When it is possible to find the maximum value of this function for a hereditary property \mathcal{H} , however, this maximum value corresponds to the value of $\lim_{n \rightarrow \infty} \text{dist}_{\text{norm}}(n, \mathcal{H})$. We will refer to this maximum value as $d_{\mathcal{H}}^*$ and the value(s) of p such that $ed_{\mathcal{H}}(p) = d_{\mathcal{H}}^*$ as $p_{\mathcal{H}}^*$. A useful tool for this endeavor is colored regularity graphs.

Definition 3 (Colored Regularity Graph; Alon-Stav [5]). *A **colored regularity graph (CRG)**, K , is a complete graph with vertices colored black or white, and with edges colored black, white or gray.*

A **sub-CRG** of a CRG K is an induced subgraph of K that preserves the coloring of vertices and edges. At times, the language “ K contains a gray subgraph H ” may also be used to describe a situation where there is a set of vertices $V(H)$ in K , so that the edges $E(H)$ corresponding to the subgraph H on these vertices are all gray.

A colored homomorphism from simple graphs into CRGs is defined as follows.

Definition 4 (Colored Homomorphism). *A **colored homomorphism** from a (simple) graph H to a colored regularity graph K is a mapping $\phi : V(H) \mapsto V(K)$, which satisfies the following:*

1. If $uv \in E(H)$, then either $\phi(u) = \phi(v) = m$ and m is colored black, or $\phi(u) \neq \phi(v)$ and the edge $\phi(u)\phi(v)$ is colored black or gray.
2. If $uv \notin E(H)$, then either $\phi(u) = \phi(v) = m$ and m is colored white, or $\phi(u) \neq \phi(v)$ and the edge $\phi(u)\phi(v)$ is colored white or gray.

A colored homomorphism from a simple graph H to a CRG K is sometimes referred to as an embedding of H in K . If no such homomorphism exists for a particular graph H , we say that K forbids H embedding.

As in [9], the sets of white vertices, white edges, black vertices, and black edges are denoted $VW(K)$, $EW(K)$, $VB(K)$, and $EB(K)$ respectively for a given CRG K , and two functions of p are defined as follows:

$$f_K(p) = \frac{1}{k^2} [p(|VW(K)| + 2|EW(K)|) + (1-p)(|VB(K)| + 2|EB(K)|)] \quad (1.1)$$

$$g_K(p) = \min\{\mathbf{u}^T M_K(p) \mathbf{u} : \mathbf{u}^T \mathbf{1} = 1 \text{ and } \mathbf{u} \geq 0\}. \quad (1.2)$$

Here $M_K(p)$ denotes a weighted adjacency matrix assigning a weight of p to the ij th entry if the edge $\{i, j\}$ in the CRG is white or if $i = j$ and the vertex corresponding to i is white, and a weight of $(1-p)$ if the edge $\{i, j\}$ in the CRG is black or if $i = j$ and the vertex corresponding to i is black. If the edge $\{i, j\}$ is gray, then the corresponding value is 0. The significance of these two functions (as shown in [9]) is that

$$ed_{\mathcal{H}}(p) = \inf f_K(p) = \inf g_K(p)$$

where \inf is taken over all of the CRGs K that permit the embedding of graphs in \mathcal{H} only.

Definition 5 (p -core CRG). *A p -core CRG is a CRG K' such that for no nontrivial sub-CRG K of K' is it the case that $g_K(p) = g_{K'}(p)$. In other words, if K' is a p -core CRG, and K is a nontrivial sub-CRG of K' , then $g_K(p) > g_{K'}(p)$.*

For p -core CRGs K , it is shown in [32] that there is a unique vector \mathbf{x} such that $\mathbf{x}^T \mathbf{1} = 1$ and $\mathbf{x} \geq 0$, and $g_K(p) = \mathbf{x}^T M_K(p) \mathbf{x}$. This vector \mathbf{x} can be viewed as a function of the vertices

$v \in V(K)$, so that $\mathbf{x}(v)$ is the weight that \mathbf{x} assigns to the vertex v . It is referred to in the literature as the **optimal weight function** of K at p . With such a weight function in place, we will also need the value $d_G(v)$, which is the sum of the weights of vertices adjacent to v in K via gray edges.

Some important CRG constructions for $\text{Forb}(K_{2,t})$ are defined (as in [34]) below.

Definition 6. *Let $K(w, b)$ denote the CRG with w white vertices, b black vertices and only gray edges. In particular:*

1. *Let $K(1, 1)$ be the CRG consisting of a white and black vertex joined by a gray edge.*
2. *Let $K(0, t - 1)$ be the CRG consisting of $t - 1$ black vertices all joined by gray edges.*

1.2 Select review of the literature and past results

Recent papers by Axenovich et al. [7] and Alon and Stav [5] originated current interest in the determination of bounds for $\text{dist}(n, \mathcal{H})$.

The work of Axenovich et al. cites applications of graph editing problems to computer science and bioinformatics, in particular work by Chen et al. in [17], as practical motivations for their work. As is pointed out in both [7] and [5], these types of problems are a one of several natural evolutions of Turán type problems [20, 22, 23, 24, 26, 27, 28], as well as other editing problems, especially those involving more global properties [8, 21].

Their results make use of the so-called binary chromatic number to bound the maximum possible edit distance of any graph from a principle hereditary property $\mathcal{H} = \text{Forb}(H)$. The binary chromatic number, introduced by Prömel and Steger in [37] as the parameter τ and generalized in [13], is defined in [7] as follows:

Definition 7 (Binary Chromatic Number). *The binary chromatic number of a graph G , $\chi_B(G)$ is the least integer $k + 1$ such that, for all $c \in \{0, \dots, k + 1\}$, there exists a partition of $V(G)$ into c cliques and $k + 1 - c$ cocliques.*

In [7], they also define c_{\min} to be the least value of c that does not allow G to be partitioned into c cliques and $k - c$ cocliques, and c_{\max} be the greatest value of c that does not allow such a

partition, where $k = \chi_B(G) - 1$. What follows is a summary of some of their results bounding $\text{dist}(n, \text{Forb}(H))$ using these parameters.

Theorem 8 (Axenovich et al. [7]). *If H is a graph with binary chromatic number $k + 1$, then*

$$\text{dist}(n, \text{Forb}(H)) > (1 - o(1)) \frac{n^2}{4k}$$

Theorem 9 (Axenovich et al. [7]). *Let H be a graph with binary chromatic number $k + 1$. If $c_{\min} \leq k/2 \leq c_{\max}$ then*

$$\text{dist}(n, \text{Forb}(H)) \leq \frac{1}{2k} \binom{n}{2}.$$

Combining these two bounds on $\text{dist}(n, \text{Forb}(H))$ yielded the following nice result for self-complementary H .

Corollary 10 (Axenovich et al. [7]). *If H is a self-complementary graph with the property that $\chi_B(H) = k + 1$, then*

$$\text{dist}(n, \text{Forb}(H)) = (1 + o(1)) \frac{n^2}{4k}.$$

In addition, they determined $\text{dist}(n, \text{Forb}(H))$ for $H \in \{K_3, \overline{K_3}, K_{1,2}, \overline{K_{1,2}}\}$, and bounded $\text{dist}(n, \text{Forb}(H))$ in several other cases.

Meanwhile in [5], Alon and Stav cite algorithmic edge-modification problems in theoretical computer science as a key motivation for their exploration of the problem. Their approach uses several versions of the regularity lemma [2, 31, 39] and colored regularity graph structures (originated by Bollobás and Thomason in [13]). It follows a method, used by Alon and Shapira in [3] to address questions about edit distance and property testing, to achieve their main result:

Theorem 11 (Alon and Stav [5]). *Let \mathcal{H} be an arbitrary hereditary graph property. Then there exists $p^* \in [0, 1]$, such that with high probability*

$$\text{dist}(n, \mathcal{H}) = \text{dist}(G(n, p^*), \mathcal{H}) + o(n^2).$$

In the case of self-complementary graphs addressed in Corollary 10, it has been observed that $p^* = 1/2$ for the above theorem. In general, however, the p^* for a specific hereditary

property is not determined explicitly by the proof of Theorem 11, and in fact, as will be discussed later in this thesis, for some hereditary properties an interval of p^* values may work.

An approach to finding p^* , employed by Balogh and Martin in [9], is to calculate the expected value of $\lim_{n \rightarrow \infty} \text{dist}_{\text{norm}}(G(n, p), \mathcal{H})$ for all $p \in [0, 1]$, and then find the value(s) of p that yield the maximum distance. What follows is the main result from their work. Recall that $f_K(p)$ and $g_K(p)$ are defined by equations 1.1 and 1.2.

Theorem 12 (Balogh and Martin [9]). *For a hereditary property \mathcal{H} , let $\mathcal{K}(\mathcal{H})$ denote all CRGs K that do not permit the embedding of any of the forbidden induced subgraphs associated with \mathcal{H} . Then $d^*(\mathcal{H}) = \lim_{n \rightarrow \infty} \text{dist}_{\text{norm}}(n, \mathcal{H})$ exists. Define*

$$f(p) = \inf_{K \in \mathcal{K}(\mathcal{H})} f_K(p) \quad \text{and} \quad g(p) = \inf_{K \in \mathcal{K}(\mathcal{H})} g_K(p).$$

Then it is the case that $f(p) = g(p)$ for all $p \in [0, 1]$,

$$d_{\mathcal{H}}^* = \max_{p \in [0, 1]} f(p) = \max_{p \in [0, 1]} g(p),$$

and $p_{\mathcal{H}}^$ is the value of p at which f achieves its maximum. In addition, the function $f(p) = g(p)$ is concave.*

Furthermore, for all $p \in (0, 1)$,

$$\max_{G: |E(G)| = p \binom{n}{2}} \{\text{dist}(G, \mathcal{H})\} = f(p) \binom{n}{2} + o(n^2),$$

and for all $\epsilon > 0$, $\text{dist}(G(n, p), \mathcal{H}) \geq f(p) \binom{n}{2} - \epsilon n^2$.

This theorem tells us not only that $ed_{\mathcal{H}}(p)$ is well defined for all p , but that it is actually achieved for each p by $\lim_{n \rightarrow \infty} \mathbb{E}[\text{dist}_{\text{norm}}(G(n, p), \mathcal{H})]$, where $f_K(p)$ and $g_K(p)$ can be viewed as two different approaches to find this quantity. The observed concavity of the function is a helpful tool for determining its maximum, even when the entire function is not computable (which, at least with current techniques, is often the case).

Furthermore, while p^* and d^* were determined by Alon and Stav in [4] for $\mathcal{H} = \text{Forb}(H)$ when $V(H) \leq 4$, the methods used in [9] to determine these parameters with the edit distance

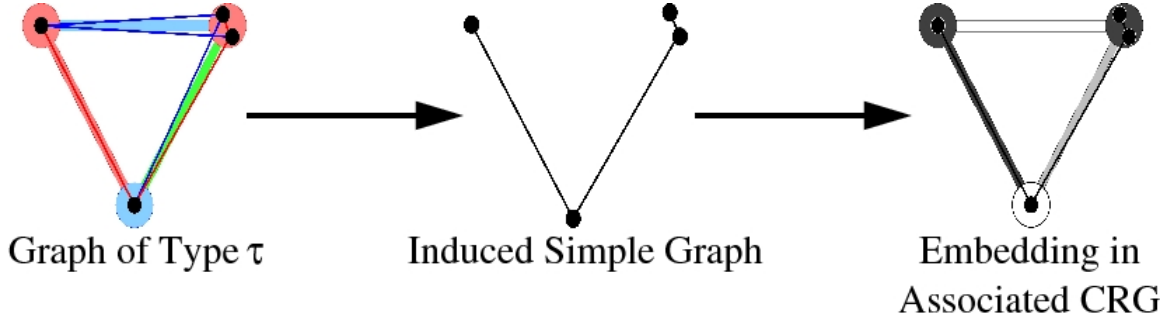


Figure 1.2 The relationship between types, simple graphs, and CRGs.

function and CRGs allows one to avoid direct use of the regularity lemmas. In [9], this method is used to compute p^* and d^* for split graphs (graphs with vertex sets that can be partitioned into two sets V_1 and V_2 , which induce a coclique and clique, respectively) and $K_{3,3}$. Also of interest is an upper bound for the so-called H_9 graph described in the paper, which demonstrates that, for this method, consideration of CRGs with only gray edges is not sufficient.

The study of CRGs in terms of edit distance and the edit distance function is closely related to a more general study of 2-coloured multigraphs by Marchant and Thomason in [32]. A two-coloured multigraph is the union of two simple graphs on H_r and H_b on the same vertex set. Associated with this union is a simple graph, H_u , with edges $\{u, v\}$ colored red, blue, or green depending on whether vertices u and v are adjacent in H_r only, H_b only, or in both H_r and H_b .

While it is not the case in general that the underlying graph is complete, when it is complete, it may be associated with one or more *types*. Types are represented by complete graphs with vertices colored either blue or red, and edges colored blue, red, or green. If H_u is complete, it is associated with a type τ (not to be confused with the parameter in [37]) if there is a mapping $\phi : V(H_u) \rightarrow V(\tau)$ such that if $\{u, v\} \in E(H_u)$ is red, blue, or green, then $\phi(u) = \phi(v) = r$, a red vertex, or $\{\phi(u), \phi(v)\}$ is a red or green edge; $\phi(u) = \phi(v) = b$, a blue vertex, or $\{\phi(u), \phi(v)\}$ is a blue or green edge; or $\{\phi(u), \phi(v)\}$ is a green edge, respectively. In the special case, where we take a simple graph H , and let $H_r = H$ and $H_b = \overline{H}$ (the graph complement of H), then the types of this paradigm correspond to the CRGs defined in the previous section, where red corresponds to black, blue corresponds to white, and gray corresponds to green.

Below are some key results from [32] for our applications, stated in terms of our CRGs and

edit distance functions. For more information on 2-coloured multigraphs, the reader may wish to consult [40], an excellent survey by Thomason of the topic, including its applications to edit distance.

Theorem 13 (Marchant and Thomason [32]). *Let K be a p -core CRG. Then all edges of K are gray, except*

- *if $p < 1/2$, then some edges joining two black vertices might be white, or*
- *if $p > 1/2$, then some edges joining two white vertices might be black.*

While lower bounds in [9] came mainly from $f(p)$, Theorem 13 lays the foundation for developing lower bounds using $g(p)$. What is more, Marchant and Thomason show that, in fact, $\inf_{K \in \mathcal{K}(\mathcal{H})} g_K(p) = \min_{K \in \mathcal{K}(\mathcal{H})} g_K(p)$. Through examples, they also demonstrate the following partial results for $H = \text{Forb}(K_{2,t})$.

Theorem 14 (Lemma 5.14 and Example 5.16 in Marchant and Thomason [32]). *Let $\mathcal{H} = \text{Forb}(K_{2,t})$, then*

- *if $t = 2$, $ed_{\mathcal{H}}(p) = p(1 - p)$.*
- *any CRG K with all black vertices and only white and gray edges so that the gray subgraph does not have $K_{2,t}$ or the book B_{t-1} forbids $K_{2,t}$ embedding.*
- *if $p \geq 1/2$ then for $t > 2$, $ed_{\mathcal{H}}(p) = \frac{1-p}{t-1}$*
- *if $p \leq 1/2$ then for $t > 2$, either $ed_{\mathcal{H}}(p) = \min\{p(1 - p), \frac{1-p}{t-1}\}$ or $ed_{\mathcal{H}}(p) = g_K(p)$ for some CRG $K \in \mathcal{K}(\mathcal{H})$ that has no white vertices.*

In the above theorem, the bound of $p(1 - p)$ comes from the CRG $K(1, 1)$ and the bound $\frac{1-p}{t-1}$ from $K(0, t - 1)$.

Theorem 15 pertains to the value of $g_K(p)$ for certain CRGs with regular gray subgraphs.

Theorem 15 (Marchant and Thomason [32]). *Let $d \geq 2$ be an integer and let G be a connected d -regular graph of order n . Let K be the CRG with vertex set $V(G)$ whose vertices are all black, with the edges of G colored gray and all other edges white.*

If $p \leq 1/(d + 2)$ then K is p -core, and $g_K(p) = \frac{1}{k} + \left(\frac{k-d-2}{k}\right)p$.

This result is interesting because many of the CRG constructions in this thesis are, in fact, regular. In this case $f_K(p) = \frac{1}{k} + \left(\frac{k-d-2}{k}\right)p$ as well, and so it appears to be the case that at least for smaller values of p , an even weighting in such instances is optimal for $g_K(p)$.

In addition to these results, Marchant and Thomason consider several other examples pertaining to hereditary properties, including an alternative example to H_9 in [9] for a property that requires CRGs that do not have all gray edges to determine the maximum value of its edit distance function, and an example of properties that have an infinite number of p -core CRGs K , such that for some fixed p , $ed_{\mathcal{H}}(p) = g_K(p)$. This last example pertains to questions of the *stability* of structures within a property that are likely to be closest to the random graph $G(n, p)$, a question that is also explored for the case when $p = 1/2$ by Alon and Stav in [6].

A final example from Marchant and Thomason that was of particular interest for this thesis was an application of a dense bipartite $K_{3,3}$ -free graph construction by Brown in [16], where in this case, $K_{3,3}$ -free refers to the absence of a weak $K_{3,3}$ subgraph, to improve on other bounds for this construction for small p .

Such constructions for $K_{s,t}$ -free graphs in general are closely related to the Zarankiewicz problem, which asks how many edges can a graph on n vertices have before it must contain a $K_{s,t}$ subgraph. Bipartite versions of these constructions can naturally be converted into CRGs that forbid $K_{2,t}$ embedding, and so one question addressed in [34] that is also discussed in this thesis, is what similar constructions by Füredi in [29] mean for $ed_{\text{Forb}(K_{2,t})}(p)$.

In addition to these constructions for a classic extremal graph theory question, we also look at CRG constructions inspired by known strongly regular graphs as summarized at Brouwer's website [15] and variations on triangle free cycle constructions discussed by Brandt in [14].

Two other results that are essential to establishing lower bounds for the edit distance functions of $\text{Forb}(K_{2,t})$ appears as lemmas in [33].

Lemma 16 (Martin [33]). *Let $p \in (0, 1)$ and K be a p -core CRG with optimal weight function \mathbf{x} .*

1. *If $p \leq 1/2$, then $\mathbf{x}(v) = g_K(p)/p$ for all $v \in \text{VW}(K)$ and*

$$d_G(v) = \frac{p - g_K(p)}{p} + \frac{1 - 2p}{p} \mathbf{x}(v), \quad \text{for all } v \in \text{VB}(K).$$

2. If $p \geq 1/2$, then $\mathbf{x}(v) = g_K(p)/(1-p)$ for all $v \in \text{VB}(K)$ and

$$d_G(v) = \frac{1-p-g_K(p)}{1-p} + \frac{2p-1}{1-p}\mathbf{x}(v), \quad \text{for all } v \in \text{VW}(K).$$

Lemma 17 (Martin [33]). *Let $p \in (0, 1)$ and K be a p -core CRG with optimal weight function \mathbf{x} .*

1. If $p \leq 1/2$, then $\mathbf{x}(v) \leq g_K(p)/(1-p)$ for all $v \in \text{VB}(K)$.

2. If $p \geq 1/2$, then $\mathbf{x}(v) \leq g_K(p)/p$ for all $v \in \text{VW}(K)$.

As is evident from Theorem 14, the case when $p \leq 1/2$ is of most interest for our purposes. However, the main results of [33] include bounds on the edit distance function for any hereditary properties that forbid a clique, and exact determination of the edit distance function for principal hereditary properties defined by forbidding K_n , C_n when $n \leq 9$, and C_{10} for $p \geq 1/7$, with a resulting determination of p^* and d^* for each of these principal hereditary properties.

In addition to the recent work on edit distance and the edit distance function summarized above, a number of related questions have also been asked. For instance, how many graphs of order n are there in a specific hereditary property [37, 38], and what is the probability that the random graph $G(n, p)$ does, in fact, fall in a hereditary property [1] (see also [11], [12], [35], [36] for additional work on hereditary properties and forbidden induced subgraphs)? Exploration in these areas helped lay the foundation for this more recent work.

1.3 Results

The following results from [34] (joint work with Ryan Martin) will be discussed in Chapter 2.

Theorem 18. *Let $\mathcal{H} = \text{Forb}(K_{2,3})$. Then $ed_{\mathcal{H}}(p) = \min\{p(1-p), \frac{1-p}{2}\}$ with $p_{\mathcal{H}}^* = \frac{1}{2}$ and $d_{\mathcal{H}}^* = \frac{1}{4}$.*

Theorem 19. *Let $\mathcal{H} = \text{Forb}(K_{2,4})$. Then $ed_{\mathcal{H}}(p) = \min\{p(1-p), \frac{7p+1}{15}, \frac{1-p}{3}\}$ with $p_{\mathcal{H}}^* = \frac{1}{3}$ and $d_{\mathcal{H}}^* = \frac{2}{9}$.*

It should be noted that the values of these functions for when $p \geq 1/2$ from Theorems 18 and 19 follow directly from Theorem 14. The function value of $\frac{1-p}{t-1}$ for $p \geq 1/2$ in Theorem 14 from [32] are extended for general t below.

Theorem 20. *Let $t \geq 4$, $p \geq 2/(t+1)$ and $\mathcal{H} = \text{Forb}(K_{2,t})$, then $ed_{\mathcal{H}}(p) = (1-p)/(t-1)$.*

Theorem 21. *Let $t \geq 3$ and $p < 1/2$. If K is a black-vertex, p -core CRG with white and gray edges such that the gray edges have neither a $K_{2,t}$ nor a book B_{t-2} , then*

$$g_K(p) \geq p - \frac{t-1}{4t-5} \left[3p - 2 + 2\sqrt{1 - 3p + (t+1)p^2} \right]. \quad (1.3)$$

Theorem 22. *For odd $t \geq 5$ and $\mathcal{H} = \text{Forb}(K_{2,t})$,*

$$d_{\mathcal{H}}^* = 1/(t+1) \quad \text{and} \quad p_{\mathcal{H}}^* \supseteq \left[\frac{2t-1}{t(t+1)}, \frac{2}{t+1} \right].$$

Theorem 22 is the first principal hereditary property known to have a nondegenerate interval of possible values for p^* .

The next two results follow from constructions by Füredi [29].

Theorem 23. *For $\mathcal{H} = \text{Forb}(K_{2,t})$, the edit distance function $ed_{\mathcal{H}}(p) \leq \frac{t-1+p(2q^2-q(t-1)-2t)}{2(q^2-1)}$ for any prime power q such that $t-1$ divides $q-1$.*

Corollary 24. *For $t \geq 9$, there exists a value q_0 , so that if $q > q_0$, then $\frac{t-1+p(2q^2-q(t-1)-2t)}{2(q^2-1)} < p(1-p)$ for some values of p , which approach 0 as q increases. That is, arbitrarily close to $p = 0$, there is some value for p such that $ed_{\mathcal{H}}(p) < p(1-p)$.*

A strongly regular graph construction provides the upper bound $\frac{7p+1}{15}$ for $ed_{\text{Forb}(K_{2,4})}(p)$. Such constructions continue to be relevant for larger t values, and so we have the following general result for such graphs.

Theorem 25. *For any (k, d, λ, μ) -strongly regular graph, there exists a corresponding CRG, K , such that*

$$f_K(p) = \frac{1}{k} + \left(\frac{k-d-2}{k} \right) p.$$

If $\lambda \leq t-3$ and $\mu \leq t-1$, then K forbids $K_{2,t}$ embedding, and when equality holds for both λ and μ ,

$$f_K(p) = \frac{t-1}{t-1+d(d+1)} + \left(1 - \frac{(d+2)(t-1)}{t-1+d(d+1)} \right) p. \quad (1.4)$$

There is an interesting connection between strongly regular graphs and the lower the result in Theorem 21. Namely, the upper bound resulting from a $(k, d, t - 3, t - 1)$ -strongly regular graph is always a line tangent to this lower bound.

The following general upper bound arises from a CRG construction involving the second power of cycles. The reasoning for our attention to these particular cycle powers will be discussed further in Chapter 3.

Theorem 26. *For $\mathcal{H} = \text{Forb}(K_{2,t})$,*

$$ed_{\mathcal{H}}(p) \leq \frac{3p + 1}{5 + t}.$$

Below are the known upper bounds for when $5 \leq t \leq 8$ based on the Theorems above and knowledge of strongly regular graphs from [15]. Knowledge of the existence of strongly regular graphs with certain parameters is still incomplete, and so improvements are possible with further advancement with regard to this question.

Theorem 27. *Let $\mathcal{H} = \text{Forb}(K_{2,t})$.*

- *If $t = 5$, then*

$$ed_{\mathcal{H}}(p) \leq \min \left\{ p(1 - p), \frac{1 + 75p}{96}, \frac{1 + 26p}{40}, \frac{1 + 5p}{13}, \frac{1}{6}, \frac{1 - p}{4} \right\}.$$

- *If $t = 6$, then*

$$ed_{\mathcal{H}}(p) \leq \min \left\{ p(1 - p), \frac{1 + 63p}{85}, \frac{1 + 14p}{26}, \frac{1 + 7p}{17}, \frac{1 + 2p}{10}, \frac{1 - p}{5} \right\}.$$

- *If $t = 7$, then*

$$ed_{\mathcal{H}}(p) \leq \min \left\{ p(1 - p), \frac{1 + 124p}{156}, \frac{1 + 76p}{100}, \frac{1 + 44p}{64}, \frac{1 + 31p}{49}, \frac{1 + 20p}{36}, \frac{1 + 5p}{16}, \frac{1}{8}, \frac{1 - p}{6} \right\}.$$

- *If $t = 8$, then*

$$ed_{\mathcal{H}}(p) \leq \min \left\{ p(1 - p), \frac{1 + 124p}{156}, \frac{1 + 95p}{125}, \frac{1 + 53p}{76}, \frac{1 + 20p}{36}, \frac{1 + 11p}{25}, \frac{1 + 5p}{16}, \frac{3p + 1}{13}, \frac{1 - p}{7} \right\}.$$

CHAPTER 2. PROOFS FOR RESULTS FROM [34]

This chapter contains proofs of the results from Section 1.3 that also appear in [34], a joint work with Ryan Martin. While most of the arguments appear as they were written in sections 3-7 and the appendices of the paper, there is some modifications from the original text for the sake of clarity and continuity.

2.1 Preliminary results and observations

We begin with some notation used throughout the chapter.

Definition 28. *Let K be a black-vertex, p -core CRG with $g_K(p) \leq p(1-p)$ and optimal weight function \mathbf{x} :*

- $N_G(v) = \{y \in V(K) : vy \in EG(K)\},$
- u_0 is a fixed vertex in K such that $\mathbf{x}(u_0) \geq \mathbf{x}(v),$ for all $v \in V(K),$ and $x = \mathbf{x}(u_0)$ is its weight,
- $U = N_G(u_0)$ and $|U| = \ell,$
- u_1 is a fixed vertex with maximum weight in $U,$ and $x_1 = \mathbf{x}(u_1),$
- W is the set of all vertices in K that are neither $u_0,$ nor contained in $U;$ or equivalently, W is the set of all vertices in the white neighborhood of $u_0,$ and
- $\mathbf{x}(S) = \sum_{y:y \in S} \mathbf{x}(y)$ for some set $S \subseteq V(K).$

Partitioning the vertices in a black-vertex, p -core CRG that forbids a $K_{2,t}$ embedding into the three sets $\{u_0\}, U$ and W as seen in Figure 2.1, illustrates some interesting features of its

optimal weight function when the gray neighborhoods of these vertices are considered in the context of Lemma 16. One such feature is the upper bounds in Proposition 29 for x_1 .

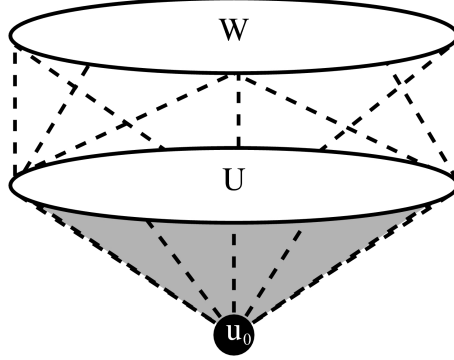


Figure 2.1 A partition of the vertices in a black-vertex, p -core CRG, K . Dashed lines and gray background represent gray edges. White edges are omitted, as are edges within subsets.

Proposition 29. *Let $K \in [\mathcal{K}(\text{Forb}(K_{2,3})) \cup \mathcal{K}(\text{Forb}(K_{2,4}))]$ be a black-vertex, p -core CRG. If either $p < 1/3$ or both $p < 1/2$ and the gray sub-CRG of K is triangle-free, then*

$$x_1 \leq x \quad \text{and} \quad x_1 \leq p - x$$

where $x = \mathbf{x}(u_0)$ is the maximum weight of a vertex in K , and $x_1 = \mathbf{x}(u_1)$ is the maximum weight of a vertex in that vertex's gray neighborhood.

Proof. The inequality $x_1 \leq x$ follows directly from definitions of x_1 and x , since x is the greatest weight in K . To justify the inequality $x_1 \leq p - x$, we break the problem into two cases:

Case 1: u_0 and u_1 have no common gray neighbor.

Recall that u_1 is a vertex with maximum weight in the gray neighborhood of u_0 , a vertex with maximum weight in all of K , and assume that $x + x_1 > p$. Then applying Lemma 16 and Theorem 14,

$$d_G(u_0) + d_G(u_1) \geq \left\lceil p + \frac{1-2p}{p}x \right\rceil + \left\lceil p + \frac{1-2p}{p}x_1 \right\rceil = 2p + \left(\frac{1-2p}{p} \right) (x + x_1) > 2p + (1-2p).$$

This is a contradiction because in Case 1, $N_G(u_0) \cap N_G(u_1) = \emptyset$. Thus, $d_G(u_0) + d_G(u_1) \leq 1$, since the sum of the weights of the vertices in K must be 1.

This completes the proof of Proposition 29 for $K \in \mathcal{K}(\text{Forb}(K_{2,3}))$, since, in this case, no $K \in \mathcal{K}$ contains a gray triangle (book B_1). So we may assume that $K \in \mathcal{K}(\text{Forb}(K_{2,4}))$.

Case 2: u_0 and u_1 have a common gray neighbor and $p < 1/3$.

In this case, u_1 has a single neighbor u_2 in U because any more such neighbors would result in a gray book B_2 (contradicting Theorem 14). Furthermore, we note that in order to avoid a gray book B_2 , the common neighborhood of u_1 and u_2 in W must be empty. Consequently, $d_G(u_1) + d_G(u_2) \leq \mathbf{x}(W) + 2x + x_1 + \mathbf{x}(u_2)$.

Applying similar reasoning to that in Case 1,

$$d_G(u_0) + d_G(u_1) + d_G(u_2) \geq \left\lceil p + \frac{1-2p}{p}x \right\rceil + \left\lceil p + \frac{1-2p}{p}x_1 \right\rceil + \left\lceil p + \frac{1-2p}{p}\mathbf{x}(u_2) \right\rceil.$$

So,

$$\begin{aligned} d_G(u_0) + (\mathbf{x}(W) + 2x + x_1 + \mathbf{x}(u_2)) &\geq \left\lceil p + \frac{1-2p}{p}x \right\rceil + \left\lceil p + \frac{1-2p}{p}x_1 \right\rceil + \left\lceil p + \frac{1-2p}{p}\mathbf{x}(u_2) \right\rceil \\ \mathbf{x}(U) + (\mathbf{x}(W) + 2x + x_1 + \mathbf{x}(u_2)) &\geq 3p + \frac{1-2p}{p}(x + x_1 + \mathbf{x}(u_2)) \\ \mathbf{x}(U) + \mathbf{x}(W) + x &\geq 3p + \frac{1-3p}{p}(x + x_1 + \mathbf{x}(u_2)) \\ 1 &\geq 3p + \frac{1-3p}{p}(x + x_1 + \mathbf{x}(u_2)). \end{aligned}$$

With $p < 1/3$ and $x + x_1 \geq p$, we have a contradiction. \square

Applying the pigeon-hole principle, we also have the following lower bound for ℓ :

Fact 30. *In a CRG, if u_0 is a vertex with maximum weight, $x = \mathbf{x}(u_0)$, the maximum weight in the gray neighborhood of u_0 is x_1 , and the order of the gray neighborhood of u_0 is ℓ , then $\ell \geq d_G(u_0)/x_1 \geq d_G(u_0)/x$.*

While simple, when combined with Lemma 16 and Proposition 29 along with the observation that $\mathbf{x}(u_0) + \mathbf{x}(U) + \mathbf{x}(W) = 1$, this fact forces a balance between the weights of the vertex u_0 , the vertices in U , and the vertices in W , that is a powerful tool for bounding $g_K(p)$.

2.2 Proof of Theorem 18

In this section, we establish the value of $ed_{\text{Forb}(K_{2,3})}(p)$ for $p \in (0, 1/2)$, determining the entire function via continuity and Theorem 14, from which we know that $ed_{\text{Forb}(K_{2,3})}(p) = (1 - p)/2$ for $p \in [1/2, 1]$.

For the following discussion, we will assume that K is a p -core CRG on all black vertices into which $K_{2,3}$ may not be embedded and that $g_K(p) \leq p(1 - p)$. The following lemma yields a useful restriction of the order of U .

Lemma 31. *Let K be a black-vertex, p -core CRG with $p \in (0, 1/2)$, no gray triangles, no gray $K_{2,3}$ and $g_K(p) \leq p(1 - p)$. If u_0 is a vertex of maximum weight, x , in K , and $\ell = |N_G(u_0)|$, then*

$$\ell \leq \frac{2(1 - x) - \frac{1}{p}d_G(u_0)}{p - x}.$$

Proof. Let u_1, \dots, u_ℓ be an enumeration of the vertices in U , the gray neighborhood of u_0 . Observe that K cannot contain a K_3 with all gray edges, and so U contains no gray edges. Therefore, with the exception of u_0 , the entire gray neighborhood of each u_i is contained in W . Furthermore, if any three vertices in U had a common gray neighbor in W , then K would contain a gray $K_{2,3}$. That is, each vertex in W is adjacent to at most two vertices in U via a gray edge. Applying these observations,

$$\sum_{i=1}^{\ell} (d_G(u_i) - x) \leq 2\mathbf{x}(W).$$

Using Lemma 16 and the assumption that $\frac{p - g_K(p)}{p} \geq p$,

$$\sum_{i=1}^{\ell} \left(p - x + \frac{1 - 2p}{p} \mathbf{x}(u_i) \right) \leq 2\mathbf{x}(W).$$

The fact that $\mathbf{x}(W) = 1 - x - d_G(u_0)$, gives

$$\begin{aligned} \ell(p-x) + \frac{1-2p}{p}d_G(u_0) &\leq 2(1-x-d_G(u_0)) \\ \ell(p-x) &\leq 2-2x-\frac{1}{p}d_G(u_0) \\ \ell &\leq \frac{2(1-x)-\frac{1}{p}d_G(u_0)}{p-x}. \end{aligned}$$

□

The following technical lemma is an important tool in the proof of Theorem 18.

Lemma 32. *Let K be a black-vertex, p -core CRG for $p \in (0, 1/2)$ with no gray triangles, no gray $K_{2,3}$ and $g_K(p) \leq p(1-p)$. If x and x_1 are defined as in Proposition 29, then*

$$\left[p + \frac{1-2p}{p}x \right] \left[\frac{1}{x_1} + \frac{1}{p(p-x)} \right] \leq \frac{2(1-x)}{p-x}.$$

Proof. By Fact 30, $\ell \geq \frac{d_G(u_0)}{x_1}$, and by Lemma 31, $\ell \leq \frac{2(1-x)-\frac{1}{p}d_G(u_0)}{p-x}$. Therefore,

$$\frac{d_G(u_0)}{x_1} \leq \frac{2(1-x)-\frac{1}{p}d_G(u_0)}{p-x}.$$

After combining the $d_G(u_0)$ terms we get,

$$d_G(u_0) \left[\frac{1}{x_1} + \frac{1}{p(p-x)} \right] \leq \frac{2(1-x)}{p-x},$$

and then applying Lemma 16,

$$\left[p + \frac{1-2p}{p}x \right] \left[\frac{1}{x_1} + \frac{1}{p(p-x)} \right] \leq \frac{2(1-x)}{p-x}.$$

□

Proof of Theorem 18. Let $p \in (0, 1/2)$, and K be a black-vertex, p -core CRG with $g_K(p) < p(1-p)$ and no gray triangle (i.e., the book B_1) or gray $K_{2,3}$.

With the above assumptions, we will show that there is no possible value for x , the value of the largest vertex-weight. To do so, we break the problem into 2 cases: $x \geq \frac{p}{2}$ and $x < \frac{p}{2}$.

Case 1: $x \geq p/2$.

We start with the inequality from Lemma 32,

$$\left[p + \frac{1-2p}{p}x \right] \left[\frac{1}{x_1} + \frac{1}{p(p-x)} \right] \leq \frac{2(1-x)}{p-x},$$

and apply the bound $x_1 \leq p-x$ from Proposition 29 to get

$$\left[p + \frac{1-2p}{p}x \right] \left[\frac{1}{p-x} + \frac{1}{p(p-x)} \right] \leq \frac{2(1-x)}{p-x}.$$

From Lemma 17, $p-x > 0$, and so

$$\begin{aligned} \left[p + \frac{1-2p}{p}x \right] \left[1 + \frac{1}{p} \right] &\leq 2(1-x) \\ x \left(\frac{1-p}{p^2} \right) &\leq 1-p \\ x &\leq p^2, \end{aligned}$$

a contradiction, since $\frac{p}{2} > p^2$ for $p \in (0, 1/2)$.

Case 2: $x < p/2$.

We again apply Lemma 32, only now we employ the trivial bound $x_1 \leq x$ from Proposition 29:

$$\begin{aligned} \left[p + \frac{1-2p}{p}x \right] \left[\frac{1}{x} + \frac{1}{p(p-x)} \right] &\leq \frac{2(1-x)}{p-x} \\ \left[p + \frac{1-2p}{p}x \right] [p(p-x) + x] &\leq 2px(1-x) \\ (4p^2 - 3p + 1)x^2 - (3p^3)x + p^4 &\leq 0. \end{aligned}$$

Observe that $4p^2 - 3p + 1$ is always positive, and therefore the parabola $(4p^2 - 3p + 1)x^2 - (3p^3)x + p^4$, in the variable x , is concave up, so the range of x values for which this inequality is satisfied is $x \in [x', x'']$ where

$$x' = \frac{3p^3 - \sqrt{-4p^4 + 12p^5 - 7p^6}}{2(1 - 3p + 4p^2)} \quad \text{and} \quad x'' = \frac{3p^3 + \sqrt{-4p^4 + 12p^5 - 7p^6}}{2(1 - 3p + 4p^2)}.$$

If $p < (6 - 2\sqrt{2})/7$, then neither x' nor x'' is real, and so the inequality is never satisfied. For $p \in \left[\frac{6-2\sqrt{2}}{7}, \frac{1}{2}\right)$, routine calculations show that $\frac{p}{2} < x'$, a contradiction to the assumption that $x < \frac{p}{2}$.

Hence, there is no possible value for x if $ed_{\text{Forb}(K_{2,3})}(p) < p(1 - p)$, so the proof is complete. \square

2.3 Proof of Theorem 19

This section addresses the case of $ed_{\text{Forb}(K_{2,4})}(p)$.

2.3.1 Upper bounds

Recall that from Theorem 14 we already know that $ed_{\text{Forb}(K_{2,4})}(p) \leq \min\{p(1 - p), \frac{1-p}{3}\}$. For the remaining upper bound, we turn to strongly regular graphs.

Lemma 33. *Let $\mathcal{H} = \text{Forb}(K_{2,t})$. If there exists a (k, d, λ, μ) -strongly regular graph with $\lambda \leq t - 3$ and $\mu \leq t - 1$, then*

$$ed_{\mathcal{H}}(p) \leq \frac{1}{k} + \frac{k - d - 2}{k}p.$$

Proof. Let G be the aforementioned strongly regular graph. We construct a CRG, K , on k black vertices with gray edges in K corresponding to adjacent vertices in G and white edges in K corresponding to nonadjacent vertices in G .

No pair of adjacent vertices has $t - 2 > \lambda$ common neighbors, so there is no book B_{t-2} in the gray subgraph, and no pair of vertices has $t > \mu, \lambda$ common neighbors, so there is no $K_{2,t}$ in the gray subgraph. Thus, by Theorem 14, K forbids $K_{2,t}$ embedding. Furthermore,

$$f_K(p) = \frac{1}{k^2} \left[(1 - p)k + 2p \left(\binom{k}{2} - \frac{dk}{2} \right) \right] = \frac{1}{k} + \frac{k - d - 2}{k}p.$$

\square

In fact, there is a $(15, 6, 1, 3)$ -strongly regular graph [15]. It is a so-called “generalized quadrangle,” $\text{GQ}(2, 2)$. As a result,

$$ed_{\text{Forb}(K_{2,4})}(p) \leq \min \left\{ p(1-p), \frac{1+7p}{15}, \frac{1-p}{3} \right\}.$$

2.3.2 Lower bounds

Because the edit distance function is both continuous and concave down, it is sufficient to verify that $ed_{\text{Forb}(K_{2,4})}(p) \geq p(1-p)$ for $p \in (0, 1/5)$ and that $ed_{\text{Forb}(K_{2,4})}(p) \geq (1-p)/3$ for $p \in (1/3, 1/2)$. This is because the line determined by the bound $\frac{1+7p}{15}$ passes through the points $(1/5, 4/25)$ and $(1/3, 2/9)$. Furthermore, by Theorem 14, we need only consider CRGs that have black vertices and white and gray edges.

Lemmas 34 and 37 address the cases where $p \in (1/3, 1/2)$ and where $p \in (0, 1/5)$, respectively.

Lemma 34. *Let $p \in (1/3, 1/2)$. If K is a black-vertex, p -core CRG that does not contain a gray book B_2 or a gray $K_{2,4}$, then $g_K(p) \geq \frac{1-p}{3}$, with equality occurring only if K is a gray triangle (i.e., $K \approx K(0, 3)$).*

Proof. We break this into two cases: when K does and does not have a gray triangle.

Case 1: K has a gray triangle.

Let the gray subgraph of K contain a triangle whose vertices are v_1, v_2 and v_3 with optimal weights y_1, y_2 and y_3 , respectively. Because K has no gray B_2 , we know that no pair of the vertices v_1, v_2, v_3 have a common gray neighbor other than the remaining vertex in the triangle. Letting $g = g_K(p)$, we have the following because the sum of the optimal weights on all vertices in K is 1:

$$y_1 + y_2 + y_3 + \sum_{i=1}^3 [d_G(v_i) - (y_1 + y_2 + y_3 - y_i)] \leq 1.$$

Then, applying Lemma 16,

$$\begin{aligned} y_1 + y_2 + y_3 + 3 \left(\frac{p-g}{p} \right) + \frac{1-2p}{p} (y_1 + y_2 + y_3) - 2(y_1 + y_2 + y_3) &\leq 1 \\ 3 \left(\frac{p-g}{p} \right) + \frac{1-3p}{p} (y_1 + y_2 + y_3) &\leq 1, \end{aligned}$$

and so

$$\frac{2p-3g}{p} \leq \left(\frac{3p-1}{p} \right) (y_1 + y_2 + y_3) \leq \frac{3p-1}{p}.$$

Consequently, $g \geq (1-p)/3$ with equality if and only if $y_1 + y_2 + y_3 = 1$; i.e., K itself is a gray triangle.

Case 2: K has no gray triangle.

Let u_0 be a vertex of largest weight, $x = \mathbf{x}(u_0)$, and let $U = N_G(u_0)$. The absence of a gray triangle means that there are no gray edges between pairs of vertices in U . Furthermore, no vertex in W can be adjacent to more than three vertices in U via a gray edge, since by Theorem 14, the gray subgraph of K does not contain a $K_{2,4}$.

Let u_1, \dots, u_ℓ be an enumeration of the vertices in U with weights x_1, \dots, x_ℓ , respectively, and $g = g_K(p)$. Then

$$\begin{aligned} \sum_{i=1}^{\ell} (d_G(u_i) - x) &\leq 3\mathbf{x}(W) \\ &\leq 3(1 - x - \mathbf{x}(U)), \end{aligned}$$

and applying Lemma 16 to compute $d_G(u_i)$,

$$\begin{aligned} \sum_{i=1}^{\ell} \left(\frac{p-g}{p} + \frac{1-2p}{p} x_i - x \right) &\leq 3(1 - x - \mathbf{x}(U)) \\ \ell \left(\frac{p-g}{p} - x \right) + \frac{1-2p}{p} \mathbf{x}(U) &\leq 3(1 - x) - 3\mathbf{x}(U) \\ \ell \left(\frac{p-g}{p} - x \right) &\leq 3(1 - x) - \frac{1+p}{p} \mathbf{x}(U). \end{aligned} \tag{2.1}$$

First, suppose $\ell \geq 5$. Then, from inequality (2.1), we have

$$5 \left(\frac{p-g}{p} - x \right) \leq 3(1 - x) - \frac{1+p}{p} \mathbf{x}(U),$$

and applying Lemma 16 again,

$$\begin{aligned} 5 \left(\frac{p-g}{p} - x \right) &\leq 3(1-x) - \frac{1+p}{p} \left(\frac{p-g}{p} + \frac{1-2p}{p} x \right) \\ \frac{1+6p}{p} \cdot \frac{p-g}{p} - 3 &\leq \left(5 - 3 - \frac{1+p}{p} \cdot \frac{1-2p}{p} \right) x \\ p(1+3p) - g(1+6p) &\leq x(4p^2 + p - 1). \end{aligned}$$

If $4p^2 + p - 1 < 0$, then we may use the fact that $x > 0$,

$$g > \frac{p(1+3p)}{1+6p} = \frac{1-p}{3} + \frac{(3p-1)(1+5p)}{3(1+6p)}.$$

If $4p^2 + p - 1 \geq 0$, then we use Lemma 17 and substitute $x = g/(1-p)$,

$$\begin{aligned} p(1+3p) - g(1+6p) &\leq \frac{g}{1-p} (4p^2 + p - 1) \\ p(1+3p) &\leq g \left(\frac{6p-2p^2}{1-p} \right) \\ \frac{1-p}{3} + \frac{(1-p)(11p-3)}{6(3-p)} &\leq g. \end{aligned}$$

Regardless of the value of $p \in (1/3, 1/2)$, if $\ell \geq 5$, then $g > (1-p)/3$. Therefore, we may assume that $\ell \leq 4$.

Second, suppose $\ell \leq 2$. Then by Fact 30 we have $\ell \geq \mathbf{x}(U)/x$, yielding

$$\mathbf{x}(U)/x \leq \ell \leq 2,$$

and so bounding $\mathbf{x}(U)$ using Lemma 16,

$$\begin{aligned} \frac{1}{x} \left(\frac{p-g}{p} + \frac{1-2p}{p} x \right) &\leq 2 \\ \frac{p-g}{p} &\leq \frac{4p-1}{p} x. \end{aligned}$$

Using Lemma 17, $x \leq \frac{g}{1-p}$ yields

$$\begin{aligned} \frac{p-g}{p} &\leq \frac{4p-1}{p} \cdot \frac{g}{1-p} \\ p(1-p) &\leq 3pg, \end{aligned}$$

and so if $\ell \leq 2$, then $g \geq (1-p)/3$, with equality if and only if $x = g/(1-p)$, and consequently,

K is a gray triangle. So, we may further assume that $\ell \in \{3, 4\}$.

Third, suppose $\ell = 3$. Then

$$\begin{aligned} \mathbf{x}(U)/x &\leq 3 \\ \frac{p-g}{p} &\leq \frac{5p-1}{p}x \\ \frac{p-g}{5p-1} &\leq x. \end{aligned}$$

Returning to inequality (2.1), we have

$$\begin{aligned} 3 \left(\frac{p-g}{p} - x \right) &\leq 3(1-x) - \frac{1+p}{p} \mathbf{x}(U) \\ \frac{1+4p}{p} \cdot \frac{p-g}{p} - 3 &\leq - \left[\frac{1+p}{p} \cdot \frac{1-2p}{p} \right] x \\ p(1+p) - g(1+4p) &\leq -(1+p)(1-2p) \left(\frac{p-g}{5p-1} \right) \\ p(1+p)(5p-1) + p(1+p)(1-2p) &\leq g [(1+p)(1-2p) + (1+4p)(5p-1)] \\ \frac{1+p}{6} &\leq g \\ \frac{1-p}{3} + \frac{3p-1}{6} &\leq g. \end{aligned}$$

If $\ell = 3$, then $g > (1-p)/3$.

Fourth, and finally, suppose $\ell = 4$. Then

$$\begin{aligned} \mathbf{x}(U)/x &\leq 4 \\ \frac{p-g}{p} &\leq \frac{6p-1}{p}x \\ \frac{p-g}{6p-1} &\leq x. \end{aligned}$$

Returning to inequality (2.1), we have

$$\begin{aligned} 4 \left(\frac{p-g}{p} - x \right) &\leq 4(1-x) - \frac{1+p}{p} \mathbf{x}(U) \\ \frac{1+5p}{p} \cdot \frac{p-g}{p} - 4 &\leq \left[4-3 - \frac{1+p}{p} \cdot \frac{1-2p}{p} \right] x \\ p(1+2p) - g(1+5p) &\leq [3p^2 + p - 1] x. \end{aligned}$$

If $3p^2 + p - 1 < 0$, then we use the fact that $x \geq (p - g)/(6p - 1)$:

$$\begin{aligned}
p(1 + 2p) - g(1 + 5p) &\leq [3p^2 + p - 1] \left[\frac{p - g}{6p - 1} \right] \\
p(1 + 2p) - \frac{p(3p^2 + p - 1)}{6p - 1} &\leq g \left[1 + 5p - \frac{3p^2 + p - 1}{6p - 1} \right] \\
\frac{1 + 3p}{9} &\leq g \\
\frac{1 - p}{3} + \frac{2(3p - 1)}{9} &\leq g.
\end{aligned}$$

If $3p^2 + p - 1 \geq 0$, then we use Fact 30 to bound $x \leq \frac{g}{1-p}$,

$$\begin{aligned}
p(1 + 2p) - g(1 + 5p) &\leq [3p^2 + p - 1] \left[\frac{g}{1 - p} \right] \\
p(1 + 2p) &\leq g \left[1 + 5p + \frac{3p^2 + p - 1}{1 - p} \right] \\
\frac{(1 - p)(1 + 2p)}{5 - 2p} &\leq g \\
\frac{1 - p}{3} + \frac{2(4p - 1)(1 - p)}{3(5 - 2p)} &\leq g.
\end{aligned}$$

Regardless of the value of $p \in (1/3, 1/2)$, if $\ell = 4$, then $g > (1 - p)/3$.

This ends Case 2 and the proof of the lemma. \square

Before proving Lemma 37, we need two propositions that are used in several cases.

Proposition 35. *Let $p \in (0, 1/2)$, and let K be a black-vertex, p -core CRG with no gray book B_2 and no gray $K_{2,4}$. If $g = g_K(p)$, $U = N_G(u_0)$, $\ell = |U|$ and $U_1 \subseteq U$ is the set of vertices in U that are incident to a gray edge in U , then*

$$\ell \left(\frac{p - g}{p} - x \right) \leq 3 - 3x - \frac{1 + p}{p} \mathbf{x}(U) + \mathbf{x}(U_1) \leq 3 - 3x - \frac{1}{p} \mathbf{x}(U).$$

Proof. Let u_1, \dots, u_ℓ be an enumeration of the vertices of U . Then

$$\sum_{i=1}^{\ell} (d_G(u_i) - x) - \mathbf{x}(U_1) \leq 3(1 - x - \mathbf{x}(U)),$$

and applying Lemma 16,

$$\sum_{i=1}^{\ell} \left(\frac{p - g}{p} + \frac{1 - 2p}{p} \mathbf{x}(u_i) - x \right) - \mathbf{x}(U_1) \leq 3(1 - x - \mathbf{x}(U)).$$

Simplification yields the first inequality. The second inequality results from observing that $\mathbf{x}(U_1) \leq \mathbf{x}(U)$. \square

Proposition 36. *Let $p \in (0, 1/2)$, and let K be a black-vertex, p -core CRG with no gray book B_2 and no gray $K_{2,4}$. If $g_K(p) \leq p(1-p)$, then both*

$$p \geq \frac{9 - 4\sqrt{3}}{11} \quad \text{and} \quad x \geq \frac{p^2}{2(1 - 3p + 5p^2)} \left[1 + 3p - \sqrt{-3 + 18p - 11p^2} \right] \geq \frac{1}{25}.$$

Proof. We begin with Proposition 35 and then use $\ell \geq \mathbf{x}(U)/x$ from Fact 30:

$$\begin{aligned} \ell \left(\frac{p-g}{p} - x \right) &\leq 3 - 3x - \frac{1}{p} \mathbf{x}(U) \\ \frac{\mathbf{x}(U)}{x} \left(\frac{p-g}{p} - x \right) &\leq 3 - 3x - \frac{1}{p} \mathbf{x}(U) \\ \mathbf{x}(U) \left(\frac{p-g}{px} - 1 + \frac{1}{p} \right) &\leq 3 - 3x \\ \left[\frac{p-g}{p} + \frac{1-2p}{p} x \right] \left[\frac{p-g}{p} + \frac{1-p}{p} x \right] &\leq 3x - 3x^2. \end{aligned} \tag{2.2}$$

Recall that $(p-g)/p \geq p$ because $g \leq p(1-p)$, so

$$\begin{aligned} \left[p + \frac{1-2p}{p} x \right] \left[p + \frac{1-p}{p} x \right] &\leq 3x - 3x^2 \\ p^2 - (1+3p)x + \frac{1-3p+5p^2}{p^2} x^2 &\leq 0. \end{aligned}$$

The quadratic formula gives that not only must the discriminant be nonnegative (requiring $p \geq (9 - 4\sqrt{3})/11$), but also

$$x \geq \frac{p^2}{2(1 - 3p + 5p^2)} \left[1 + 3p - \sqrt{-3 + 18p - 11p^2} \right].$$

For $p \in [(9 - 4\sqrt{3})/11, 1/2)$, this expression is at least $1/25$, achieving that value uniquely at $p = 1/5$. \square

Lemma 37. *Let $p \in (0, 1/5)$. If K is a black-vertex, p -core CRG that does not contain a gray book B_2 or a gray $K_{2,4}$, then $g_K(p) > p(1-p)$.*

Proof. We assume that $g_K(p) \leq p(1-p)$.

Case 1: $\ell \geq 8$.

According to Proposition 35,

$$\begin{aligned} 8 \left(\frac{p-g}{p} - x \right) &\leq \ell \left(\frac{p-g}{p} - x \right) \leq 3 - 3x - \frac{1}{p} \left(\frac{p-g}{p} + \frac{1-2p}{p} x \right) \\ (1-2p-5p^2)x &\leq 3p^2 - (p-g)(1+8p), \end{aligned}$$

and since $x \geq 1/25$ and $p-g \geq p^2$,

$$\begin{aligned} \frac{1-2p-5p^2}{25} &\leq 3p^2 - p^2(1+8p) \\ (1-5p)^2(1+8p) &\leq 0, \end{aligned}$$

a contradiction. So, $\ell < 8$.

Case 2: $\ell \leq 7$ and $x < p^2/(9p-1)$.

Using Fact 30, and then Lemma 16

$$7 \geq \ell \geq \frac{\mathbf{x}(U)}{x} \geq \frac{p}{x} + \frac{1-2p}{p} > \frac{9p-1}{p} + \frac{1-2p}{p} = 7,$$

a contradiction.

Case 3: $\ell \leq 7$ and $p^2/(9p-1) \leq x \leq p/3$.

First we bound ℓ :

$$\ell \geq \frac{\mathbf{x}(U)}{x} \geq \frac{p}{x} + \frac{1-2p}{p} \geq 3 + \frac{1}{p} - 2 > 6.$$

So, $\ell = 7$. Since ℓ is odd, $\mathbf{x}(U_1) \leq 6x$. By Lemma 35,

$$\ell \left(\frac{p-g}{p} - x \right) \leq 3 - 3x - \frac{1+p}{p} \mathbf{x}(U) + \mathbf{x}(U_1),$$

and applying Lemma 16,

$$\begin{aligned}
7 \left(\frac{p-g}{p} - x \right) &\leq 3 - 3x - \frac{1+p}{p} \left[\frac{p-g}{p} + \frac{1-2p}{p} x \right] + 6x \\
\frac{1-p-12p^2}{p^2} x &\leq 3 - \frac{1+8p}{p} \cdot \frac{p-g}{p} \\
\frac{1-p-12p^2}{p^2} \left[\frac{p^2}{9p-1} \right] &\leq 3 - \frac{1+8p}{p} \cdot p \\
\frac{(1-4p)(1+3p)}{9p-1} &\leq 2(1-4p) \\
\frac{1+3p}{9p-1} &\leq 2,
\end{aligned}$$

which implies $p \geq 1/5$, a contradiction.

Case 4: $\ell \leq 7$ and $x > p/3$.

Now we compute a stronger bound on U_1 . Let u_1 and u_2 be vertices in U that are adjacent via a gray edge, and let their weights be x_1 and x_2 , respectively. Then

$$x + \mathbf{x}(U) + (d_G(u_1) - x - x_2) + (d_G(u_2) - x - x_1) \leq 1$$

because u_1 and u_2 have no common gray neighbor other than u_0 and because they can have no additional gray neighbor in U . Applying Lemma 16,

$$\begin{aligned}
x + \frac{p-g}{p} + \frac{1-2p}{p} x + 2 \frac{p-g}{p} - 2x + \frac{1-3p}{p} (x_1 + x_2) &\leq 1 \\
\frac{1-3p}{p} (x_1 + x_2) &\leq \frac{3g-2p}{p} - \frac{1-3p}{p} x,
\end{aligned}$$

and since $p(1-p) \geq g$,

$$x_1 + x_2 \leq p - x.$$

We can bound the number of vertices in $U - U_1$ by using the fact that $(\ell - \ell_1)x \geq \mathbf{x}(U) - \mathbf{x}(U_1)$. Returning to Proposition 35,

$$\begin{aligned}
\ell \left(\frac{p-g}{p} - x \right) &\leq 3 - 3x - \frac{1+p}{p} \mathbf{x}(U) + \mathbf{x}(U_1) \\
\left[\ell_1 + \frac{1}{x} \mathbf{x}(U) - \frac{1}{x} \mathbf{x}(U_1) \right] \left(\frac{p-g}{p} - x \right) &\leq 3 - 3x - \frac{1+p}{p} \mathbf{x}(U) + \mathbf{x}(U_1) \\
\mathbf{x}(U) \left(\frac{p-g}{px} - 1 + \frac{1+p}{p} \right) - 3 + 3x &\leq \mathbf{x}(U_1) \left(1 + \frac{p-g}{px} - 1 \right) - \ell_1 \left(\frac{p-g}{p} - x \right).
\end{aligned}$$

If $\ell_1 = |U_1|$, then $\mathbf{x}(U_1) \leq (\ell_1/2)(p-x)$. Of course, $\mathbf{x}(U)$ is bounded below by Lemma 16.

$$\begin{aligned}
\left[\frac{p-g}{p} + \frac{1-2p}{p}x \right] \left(\frac{p-g}{px} + \frac{1}{p} \right) - 3 + 3x &\leq \frac{\ell_1}{2}(p-x) \left(\frac{p-g}{px} \right) - \ell_1 \left(\frac{p-g}{p} - x \right) \\
\left[\frac{p-g}{p} + \frac{1-2p}{p}x \right] \left(\frac{p-g}{px} + \frac{1}{p} \right) - 3 + 3x &\leq \ell_1 \left[x - \frac{p-g}{p} \cdot \frac{3x-p}{2x} \right] \\
\left[p + \frac{1-2p}{p}x \right] \left(\frac{p}{x} + \frac{1}{p} \right) - 3 + 3x &\leq \ell_1 \left[x - \frac{p(3x-p)}{2x} \right] \\
p^2 - (1+2p)x + \frac{1-2p+3p^2}{p^2}x^2 &\leq \ell_1 \frac{(p-x)(p-2x)}{2}.
\end{aligned}$$

Now, we bound ℓ_1 , depending on the sign of $p-2x$, requiring two more cases.

Case 4a: $\ell \leq 7$ and $x > p/3$ and $p-2x \geq 0$.

Here we use the bound $\ell_1 \leq 6$:

$$\begin{aligned}
p^2 - (1+2p)x + \frac{1-2p+3p^2}{p^2}x^2 &\leq 3(p-x)(p-2x) \\
-2p^2 + (7p-1)x + \frac{1-2p-3p^2}{p^2}x^2 &\leq 0.
\end{aligned}$$

By Proposition 36, we may restrict our attention to $p \geq (9-4\sqrt{3})/11 > 1/7$ and so we may substitute the smallest possible value for x , which still maintains the inequality.

$$\begin{aligned}
-2p^2 + (7p-1) \left(\frac{p}{3} \right) + \frac{1-2p-3p^2}{p^2} \left(\frac{p}{3} \right)^2 &< 0 \\
-18p^2 + 3(7p-1)p + (1-2p-3p^2) &< 0 \\
1-5p &< 0,
\end{aligned}$$

a contradiction.

Case 4b: $\ell \leq 7$ and $x > p/3$ and $p-2x < 0$.

Here we use the bound $\ell_1 \geq 0$ and then replace x with $\frac{p^2(1+2p)}{2-4p+6p^2}$, the value that minimizes

the left-hand side:

$$\begin{aligned} p^2 - (1 + 2p)x + \frac{1 - 2p + 3p^2}{p^2}x^2 &\leq 0 \\ p^2 - \frac{(1 + 2p)^2 p^2}{4(1 - 2p + 3p^2)} &\leq 0 \\ \frac{p^2(3 - 12p + 8p^2)}{4(1 - 2p + 3p^2)} &\leq 0. \end{aligned}$$

This, too, is a contradiction for $p \in (0, 1/5)$, completing the proof of Lemma 37. \square

2.4 Proofs of Theorems 20, 21 and 22

This section extends the generally known interval for $ed_{\text{Forb}(K_{2,t})}(p)$ from $p \in [1/2, 1]$ to $p \in [\frac{2}{t+1}, 1]$. With a new CRG construction, this extension is sufficient to determine $d_{\mathcal{H}}^*$ and a subset of $p_{\mathcal{H}}^*$ for odd t . Subsection 2.4.1 contains the proof of Theorem 20, while the remaining subsections address Theorems 21 and 22.

2.4.1 An extension of the known interval for $K_{2,t}$

Proof of Theorem 20. Let K be a p -core CRG for $p \in [\frac{2}{t+1}, 1]$ that does not permit $K_{2,t}$ embedding for $t \geq 5$. If we assume that $g_K(p) < g_{K(0,t-1)}(p) = (1 - p)/(t - 1)$, then by Theorem 14, K has only black vertices and no black edges.

Again, we partition the vertices of K into three sets $\{u_0\}$, $U = \{u_1, \dots, u_\ell\}$ and W , where u_0 is a fixed vertex with maximum weight x ; U is the set of all vertices in the gray neighborhood of u_0 with u_1 a vertex of maximum weight x_1 in U ; and W is the set of all remaining vertices, or those vertices adjacent to u_0 via white edges. Finally, let $d_G(u_i)$ signify the sum of the weights of all vertices in the gray neighborhood of u_i .

Then by Theorem 14, the total weight of the vertices in W is at least

$$d_G(u_1) - (t - 3)x_1 - x,$$

since no vertex in U can be adjacent to more than $t - 3$ other vertices in U without forming a book B_{t-2} gray subgraph with u_0 . Thus,

$$x + d_G(u_0) + [d_G(u_1) - (t - 3)x_1 - x] \leq 1.$$

Applying Lemma 17 and letting $g_K(p) = g$,

$$\begin{aligned}
2 \left(\frac{p-g}{p} \right) + \frac{1-2p}{p}x + \left[\frac{1-2p}{p} - (t-3) \right] x_1 &\leq 1 \\
2(p-g) - p + (1-2p)x &\leq [p(t-1) - 1] x_1 \\
2(p-g) - p + (1-2p)x &\leq [p(t-1) - 1] x \\
p - 2g &\leq [p(t+1) - 2] x.
\end{aligned}$$

Since $p \geq \frac{2}{t+1}$ and $x \leq \frac{g}{1-p}$ by Lemma 17,

$$\begin{aligned}
p - 2g &\leq [p(t+1) - 2] \frac{g}{1-p} \\
\frac{1-p}{t-1} &\leq g.
\end{aligned}$$

By Theorem 14, $g \leq \frac{1-p}{t-1}$, for $p \in [\frac{2}{t+1}, 1]$, so $ed_{\text{Forb}(K_{2,t})}(p) = \frac{1-p}{t-1}$. \square

We will now show that this result is enough to determine the maximum value of $ed_{\text{Forb}(K_{2,t})}(p)$ for odd t .

2.4.2 A construction for odd t

Proposition 38. *Let $\mathcal{H} = \text{Forb}(K_{2,t})$ for odd t . Then $ed_{\mathcal{H}}(p) \leq 1/(t+1)$.*

Proof. Let K be the CRG consisting of $t+1$ black vertices with white subgraph forming a perfect matching and all other edges gray. The CRG, K , does not contain a gray $K_{2,t}$ or book B_{t-2} , and so by Theorem 14, K forbids a $K_{2,t}$ embedding.

The CRG, K , contains exactly $(t+1)/2$ white edges, so by Equation (1.1),

$$f_K(p) = \frac{1}{(t+1)^2} \left[p \left(2 \cdot \frac{t+1}{2} \right) + (1-p)(t+1) \right] = \frac{1}{t+1}.$$

Therefore, $ed_{\mathcal{H}}(p) \leq 1/(t+1)$. \square

Since by Theorem 20, $ed_{\text{Forb}(K_{2,t})}(\frac{2}{t+1}) = \frac{1}{t+1}$, and by Proposition 38, $ed_{\text{Forb}(K_{2,t})} \leq \frac{1}{t+1}$, we have that $d_{\text{Forb}(K_{2,t})}^* = \frac{1}{t+1}$ for odd $t \geq 5$.

2.4.3 A general lower bound for t

We conclude this section by determining a general lower bound for the edit distance function of $\text{Forb}(K_{2,t})$. It is the lower bound from Theorem 21, and it allows us to make the claim in Theorem 22 that, in the case of odd t , there is a nondegenerate interval $p_{\mathcal{H}}^*$ that achieves the maximum value of the function.

Proof of Theorem 21. Here we use the standard bounds from Propositions 16 and 17. Let $g = g_K(p)$, where K is a black-vertex, p -core CRG, and let $N_G(v)$ denote the gray neighborhood of a given vertex v in K . Then if u_1, \dots, u_ℓ are the vertices in the gray neighborhood, U , of a fixed vertex of maximum weight, u_0 ,

$$\sum_{i=1}^{\ell} [d_G(u_i) - x - \mathbf{x}(N_G(u_i) \cap N_G(u_0))] \leq (t-1)(1-x-d_G(u_0)).$$

The left-hand side of this inequality calculates the weight of the total gray neighborhood of each vertex in U that must be contained in W , the set of all vertices not in U or u_0 . On the right-hand side we make use of the facts that $\mathbf{x}(W) = 1-x-d_G(u_0)$ and that no vertex in W may be adjacent to more than $(t-1)$ vertices in U without violating Theorem 14 by forming a gray $K_{2,t}$ with u_0 . Thus, applying Lemma 16,

$$\sum_{i=1}^{\ell} \left[\frac{p-g}{p} - x + \frac{1-2p}{p} \mathbf{x}(u_i) \right] - \sum_{i=1}^{\ell} \mathbf{x}(N_G(u_i) \cap N_G(u_0)) \leq (t-1)(1-x-d_G(u_0)).$$

Again considering Theorem 14 reveals that no vertex $u_i \in U$ can have more than $t-3$ gray neighbors in U without inducing a gray book B_{t-2} with u_0 . Therefore,

$$\begin{aligned} \ell \left[\frac{p-g}{p} - x \right] + \frac{1-2p}{p} d_G(u_0) - (t-3)d_G(u_0) &\leq (t-1)(1-x-d_G(u_0)) \\ \ell \left[\frac{p-g}{p} - x \right] &\leq (t-1)(1-x) - \frac{1}{p} d_G(u_0). \end{aligned}$$

Recalling that by Lemma 17, $\frac{p-g}{p} \geq x$, we use the pigeon-hole bound from Fact 30 $\ell \geq$

$d_G(u_0)/x$ to get

$$\begin{aligned} \frac{d_G(u_0)}{x} \left[\frac{p-g}{p} - x \right] &\leq (t-1)(1-x) - \frac{1}{p} d_G(u_0) \\ d_G(u_0) \left[\frac{p-g}{p} - x \right] &\leq (t-1)x(1-x) - \frac{x}{p} d_G(u_0) \\ d_G(u_0) \left[\frac{p-g}{p} + \frac{1-p}{p} x \right] &\leq (t-1)x(1-x). \end{aligned}$$

By Lemma 16,

$$\left[\frac{p-g}{p} + \frac{1-2p}{p} x \right] \left[\frac{p-g}{p} + \frac{1-p}{p} x \right] \leq (t-1)x(1-x).$$

Collecting terms yields,

$$\left(\frac{p-g}{p} \right)^2 + \left[\left(\frac{p-g}{p} \right) \left(\frac{2-3p}{p} \right) - (t-1) \right] x + \left[\left(\frac{1-2p}{p} \right) \left(\frac{1-p}{p} \right) + (t-1) \right] x^2 \leq 0,$$

and so minimizing the left-hand side of the inequality with respect to x , we have

$$\begin{aligned} \left(\frac{p-g}{p} \right)^2 - \frac{\left[(t-1) - \left(\frac{p-g}{p} \right) \left(\frac{2-3p}{p} \right) \right]^2}{4 \left[\left(\frac{1-2p}{p} \right) \left(\frac{1-p}{p} \right) + (t-1) \right]} &\leq 0 \\ \left(\frac{p-g}{p} \right)^2 (4t-5) + 2 \left(\frac{p-g}{p} \right) (t-1) \left(\frac{2-3p}{p} \right) - (t-1)^2 &\leq 0. \end{aligned}$$

Using the quadratic formula,

$$\begin{aligned} \frac{p-g}{p} &\leq \frac{-2(t-1) \left(\frac{2-3p}{p} \right) + \sqrt{4(t-1)^2 \left(\frac{2-3p}{p} \right)^2 + 4(t-1)^2(4t-5)}}{2(4t-5)} \\ p-g &\leq \frac{t-1}{4t-5} \left[3p-2 + \sqrt{(2-3p)^2 + (4t-5)p^2} \right] \\ g &\geq p - \frac{t-1}{4t-5} \left[3p-2 + 2\sqrt{1-3p+(t+1)p^2} \right]. \end{aligned}$$

□

The function in (1.3) achieves its maximum at $p = \frac{2t-1}{t^2+t}$, and that maximum is, in fact, $\frac{1}{t+1}$. Hence $ed_{\text{Forb}(K_{2,t})}(p)$ is at least $\frac{1}{t+1}$ at $p = \frac{2t-1}{t(t+1)}$ and is at least $\frac{1}{t+1}$ at $p = \frac{2}{t+1}$. As a result of concavity,

$$ed_{\text{Forb}(K_{2,t})}(p) \geq \frac{1}{t+1} \quad \text{for} \quad p \in \left[\frac{2t-1}{t(t+1)}, \frac{2}{t+1} \right].$$

Equality holds whenever t is odd because, in that case, Proposition 38 gives that $ed_{\text{Forb}(K_{2,t})}(p) \leq 1/(t+1)$, so $p_{\mathcal{H}}^*$ must be an interval. This concludes the proof of Theorem 22.

If $t \geq 5$, then we can analyze the first and second derivatives, with respect to p , of

$$p(1-p) - \left(p - \frac{t-1}{4t-5} \left[3p - 2 + 2\sqrt{1 - 3p + (t+1)p^2} \right] \right). \quad (2.3)$$

The maximum difference between $p(1-p)$ and the lower bound in Theorem 21 on the interval $[0, \frac{2}{t+1}]$ is $\frac{1}{t+1}$ and occurs when $p = \frac{2t-1}{t(t+1)}$. We can also see that (2.3) is bounded below by $\left(\frac{1}{2} - \frac{1}{t-1}\right)p(1-p)$.

2.5 Upper bound constructions

In this section we look at upper bound constructions for when $t \geq 5$.

Subsection 2.5.1 revisits the work in Section 2.3.1 on strongly regular graphs. Subsection 2.5.2 gives some constructions inspired by the analysis of triangle-free graphs in [14].

2.5.1 Results from strongly regular graph constructions

Recall that a *strongly regular graph* with parameters (k, d, λ, μ) is a d -regular graph on k vertices such that each pair of adjacent vertices has λ common neighbors, and each pair of nonadjacent vertices has μ common neighbors. Here we develop a function based on the existence of a strongly regular graph.

Suppose that K is a CRG with all vertices black and all edges white or gray that is derived from a (k, d, λ, μ) -strongly regular graph so that the edges of the strongly regular graph correspond to gray edges of K . In such a case we recall from Section 2.3.1 that

$$f_{S_{k,d,\lambda,\mu}}(p) = \frac{1}{k} + \left(\frac{k-d-2}{k} \right) p.$$

As is commonly known (see [41], for instance), if a strongly regular graph with parameters (k, d, λ, μ) exists then it is necessary, though not sufficient, for

$$d(d - \lambda - 1) = \mu(k - d - 1).$$

If we substitute $\lambda = t - 3$ and $\mu = t - 1$ in this equation and then solve for k , we find that

$$k = \frac{t-1 + d(d+1)}{t-1},$$

and substituting these values into $f_{S_{k,d,\lambda,\mu}}(p)$ yields

$$f_{S_{k,d,\lambda,\mu}}(p) = \frac{t-1}{t-1+d(d+1)} + \left(1 - \frac{(d+2)(t-1)}{t-1+d(d+1)}\right)p.$$

Fixing p and minimizing $f_{S_{k,d,\lambda,\mu}}(p)$ with respect to d gives the following expression:

$$\frac{p(t-2) + 2(t-1)}{4t-5} - \frac{2(t-1)}{4t-5} \sqrt{1 - 3p + (t+1)p^2}, \quad (2.4)$$

which is equal to the lower bound from (1.3) in Theorem 21.

Of course, in order to even have a chance of actually attaining (2.4) with a strongly regular graph construction, both d and $k = \frac{t-1+d(d+1)}{t-1}$ must be integers. This equation, however, provides something of a best case scenario for strongly regular graphs, and if there is a CRG, K , derived from a $(k, d, t-3, t-1)$ -strongly regular graph that realizes equation (2.4), then $f_K(p)$ is tangent to the lower bound in (1.3) at

$$p = \frac{2d+1}{(d+1)(d+3)-t},$$

determining the value of $ed_{\text{Forb}(K_{2,t})}(p)$ exactly.

The remaining upper bounds in Theorem 27 are the result of checking constructions from the known strongly regular graphs listed at [15]. Figure 2.2 (see Section 2.6) is a chart of the relevant parameters and $f_K(p)$ functions for $5 \leq t \leq 8$.

There is an additional construction defining the upper bound for $t = 8$ in Theorem 26, described in the following section, and explored in more depth in Chapter 3.

2.5.2 Cycle construction

Let C_k^r be the cycle on k vertices raised to the r th power. Define $C_{k,r}$ to be the CRG on k black vertices with white edges corresponding to those in C_k^r and gray edges corresponding to those in the complement of C_k^r . Let EW denotes the set of white edges for $C_{k,r}$. Then

$$\begin{aligned} f_{C_{k,r}}(p) &= \frac{1}{k^2}[(1-p)k + 2p|\text{EW}|] \\ &= \frac{1}{k^2}[(1-p)k + 2p(rk)] \\ &= \left(\frac{2r-1}{k}\right)p + \frac{1}{k}. \end{aligned}$$

Proposition 39. $C_{5+t,2}$ forbids a $K_{2,t}$ embedding, and therefore $ed_{\text{Forb}(K_{2,t})}(p) \leq \frac{3p+1}{5+t}$.

Proof. First, we check that $C_{5+t,2}$ does not contain a gray $K_{2,t}$. If u_1 and u_2 are any two vertices in C_{5+t}^2 , then $|(N(u_1) \cup N(u_2)) - \{u_1, u_2\}| \geq 4$. This inequality is justified by observing that two vertices u_1 and u_2 that are neighbors in C_{5+t} have the smallest possible number of total neighbors in C_{5+t}^2 , and this common neighborhood has order 4. It then follows that $|N(u_1) \cap N(u_2)| \leq t-1$ in the complement of C_{5+t}^2 . Thus, $C_{5+t,2}$ does not contain a gray $K_{2,t}$.

Second, we check that $C_{5+t,2}$ does not contain a gray B_{t-2} . If u_1 and u_2 are any two nonadjacent vertices in C_{5+t}^2 , then $|(N(u_1) \cup N(u_2)) - \{u_1, u_2\}| \geq 6$. Therefore, by reasoning similar to above, $|N(u_1) \cap N(u_2)| \leq t-3$ in the complement of C_{5+t}^2 , implying $C_{5+t,2}$ does not contain a gray B_{t-2} .

Thus, by Theorem 14, $C_{5+t,2}$ forbids a $K_{2,t}$ embedding, and therefore $ed_{\text{Forb}(K_{2,t})}(p) \leq f_{C_{5+t,2}}(p) = \frac{3p+1}{5+t}$. \square

While there are several other orders and powers of cycles that would also lead to a construction forbidding $K_{2,t}$ embedding, it is shown in Section 3.1 that none of them have a corresponding $f_K(p)$ value that beats the upper bound $\min\{p(1-p), \frac{3p+1}{5+t}, \frac{1-p}{t-1}\}$, so we restrict our interest to this one.

For $t \geq 5$, $f_{C_{5+t,2}}(p)$ is always an improvement on the bound $\min\{p(1-p), \frac{1-p}{t-1}\}$ from Theorem 14, though it is improved upon or made irrelevant by bounds from strongly regular graphs, for $t \leq 7$. When $t = 4$, the function $f_{C_{9,2}}(p)$ is tangent to the edit distance function at $p = 1/3$, where the edit distance function achieves its maximum value.

2.5.3 Füredi constructions

As is observed in Theorem 14 and used in the exploration of the past two constructions, graphs that forbid $K_{2,t}$ and B_{t-2} as subgraphs are of interest when looking for CRGs that forbid $K_{2,t}$ embedding. The following results come from examining the bipartite versions of $K_{2,t}$ -free graph constructions described by Füredi in [29] (see Chapter 3). This strategy mimics the one used in [32] with Brown's $K_{3,3}$ -free construction.

Proof of Theorem 23. We take the construction described in [29] for a $K_{2,t}$ -free graph G on $n = (q^2 - 1)/(t - 1)$ vertices, each with degree q , where q is a prime power so that $t - 1$ divides $q - 1$. We should note here that in the original construction from [29], loops were omitted, reducing the degree of some vertices to $q - 1$. It is to our advantage, however, to leave the loops in so that the final construction will be q -regular. By the same proof as in [29], the graph with loops still retains the property that no two vertices have a common neighborhood greater than $t - 1$ even when a looped vertex is considered to be in its own neighborhood.

Next, we create a CRG, K , by taking two copies of the vertex set $\{v_1, \dots, v_n\}$ from the $K_{2,t}$ -free graph with loops described above: $\{v'_1, \dots, v'_n\}, \{v''_1, \dots, v''_n\}$. Color all of these $k = 2n$ vertices black, and let $\text{EG}(K) = \{v'_i v''_j : v_i v_j \in E(G)\}$ with all edges not in $\text{EG}(K)$ white.

The gray subgraph of K is bipartite, so it cannot contain a B_{t-2} , and since no two vertices v_i and v_j from the original construction have more than $t - 1$ common neighbors, the common neighborhood of two vertices in the gray subgraph of K is also at most $t - 1$. Thus by Theorem 14, K forbids a $K_{2,t}$ embedding.

The CRG, K , has $k = 2n = 2(q^2 - 1)/(t - 1)$ vertices and $q(q^2 - 1)/(t - 1)$ gray edges, so by equation (1.1), $f_K(p)$ is as described in the statement of Theorem 23. \square

Remark 40. *Though the property of being bipartite is sufficient to exclude a B_{t-2} subgraph, using a bipartite $K_{2,t}$ -free construction may not be the optimal choice. A more efficient CRG may be constructed from another graph that has a gray subgraph that is both $K_{2,t}$ - and B_{t-2} -free, but, for instance, still contains triangles.*

Nevertheless, we can discover more about the potential for these constructions to improve upon the bounds for $\text{ed}_{\text{Forb}(K_{2,t})}(p)$ by fixing p and considering the general formula in Theorem 23 as a continuous function with respect to q .

Lemma 41. *Let $t \geq 3$, and let $q_0 < q$ be prime powers such that $t - 1$ divides both $q_0 - 1$ and $q - 1$. If the CRG, K_0 , is constructed according to the proof of Theorem 23 with parameter q_0 and if the CRG, K , is constructed according to the proof of Theorem 23 with parameter q , then $f_{K_0}(p) \leq f_K(p)$ for $p \in \left[\frac{2}{4+q_0}, \frac{1}{3}\right)$.*

Proof. We begin the proof by fixing p and t and analyzing $\phi(q) = \frac{t-1+p(2q^2-q(t-1)-2t)}{2(q^2-1)}$. Note that $f_{K_0}(p) = \phi(q_0)$, and $f_K(p) = \phi(q)$. Consider when the derivative

$$\phi'(q) = \frac{(t-1)(q^2p + p + 4qp - 2q)}{2(q^2-1)^2}$$

is positive and, therefore, ϕ is increasing. Since the greater value of q that makes $q^2p + p + 4qp - 2q = 0$ (note that the leading term is nonnegative) occurs at $q = \frac{(1-2p) + \sqrt{(1-2p)^2 - p^2}}{p}$, it follows that $\phi'(q) \geq 0$ when $q \geq \frac{(1-2p) + \sqrt{(1-2p)^2 - p^2}}{p}$. If $p < 1/3$ and $q_0 \geq \frac{2(1-2p)}{p}$, then

$$q > q_0 \geq \frac{2(1-2p)}{p} > \frac{(1-2p) + \sqrt{(1-2p)^2 - p^2}}{p}.$$

Thus, $\phi'(q) \geq 0$ for $\frac{2}{4+q_0} \leq p < 1/3$. Therefore $f_{K_0}(p) \leq f_K(p)$ for p in this interval. \square

Additionally, we can make some statements about when we can expect constructions that originate from the $K_{2,t}$ -free graphs described by Füredi [29] to improve upon the bound $p(1-p)$ for any q .

Lemma 42. *Fix $t \geq 9$, and let q be a prime power such that $t-1$ divides $q-1$. Let K be the CRG with parameter q described in the proof of Theorem 23, hence $f_K(p) = \frac{t-1+p(2q^2-q(t-1)-2t)}{2(q^2-1)}$. Then for any sufficiently large prime power q and corresponding K , there is an interval of values of p on which $f_K(p) < p(1-p)$. Moreover as $q \rightarrow \infty$ the left-hand endpoints of these open intervals approach 0.*

That is, we can find an infinite sequence of CRG constructions that improve upon the known bounds for $\text{Forb}(K_{2,t})$ when $t \geq 9$, and the intervals on which these improvements occur get arbitrarily close to 0.

Proof. We begin by observing that $f_K(p) = \frac{t-1+p(2q^2-q(t-1)-2t)}{2(q^2-1)} = p - \frac{p(q(t-1)+2t-2)-(t-1)}{2(q^2-1)}$. Thus if $f_K(p) < p(1-p)$,

$$\begin{aligned} p - \frac{p(q(t-1)+2t-2)-(t-1)}{2(q^2-1)} &< p - p^2 \\ 2p^2(q^2-1) &< p(q(t-1)+2(t-1)) - (t-1) \\ 2p^2(q^2-1) - p(t-1)(q+2) + (t-1) &< 0. \end{aligned} \tag{2.5}$$

The minimum value of $2p^2(q^2 - 1) - p(t - 1)(q + 2) + (t - 1)$ occurs when $p = \frac{(t-1)(q+2)}{4(q^2-1)}$.

Therefore, the inequality above is satisfied for some q and p values if and only if

$$\begin{aligned} 2 \left[\frac{(t-1)(q+2)}{4(q^2-1)} \right]^2 (q^2 - 1) - \left[\frac{(t-1)(q+2)}{4(q^2-1)} \right] (t-1)(q+2) + (t-1) &< 0 \\ (t-1) \left(1 - \frac{(t-1)(q+2)^2}{8(q^2-1)} \right) &< 0. \end{aligned}$$

That is, $f_K(p)$ from the constructions in [29] is less than $p(1-p)$ for some value of p if and only if $1 - \frac{(t-1)(q+2)^2}{8(q^2-1)} < 0$. For positive q , it is always the case that $(q+2)^2 > q^2 - 1$, and so any q satisfying the constraints of the original construction will improve upon the upper bound established by $p(1-p)$ for some p when $t \geq 9$. Furthermore, for a fixed prime power q for which $t-1$ divides $q-1$, it is a definite improvement for some open neighborhood around $p = \frac{(t-1)(q+2)}{4(q^2-1)}$. This value approaches 0 as $q \rightarrow \infty$, and there are an infinite number of prime powers q such that $t-1$ divides $q-1$ (see [29]). Thus, it is the case that for arbitrarily small p , we can find some q such that $f_K(p) < p(1-p)$. □

Lemma 43 is an analysis of the Füredi constructions when $t \leq 8$. We then show that our bounds from these constructions do not have an effect on the value of $ed_{\text{Forb}(K_{2,t})}(p)$ for $t \leq 8$.

Lemma 43. *Fix $5 \leq t \leq 8$, and let q be a prime power such that $t-1$ divides $q-1$. Let K be the CRG with parameter q described in the proof of Theorem 23, hence $f_K(p) = \frac{t-1+p(2q^2-q(t-1)-2t)}{2(q^2-1)}$. Then*

$$q < \frac{(t-1) + \sqrt{(t-1)^2 + (9-t)(t+1)}}{\frac{1}{2}(9-t)}. \quad (2.6)$$

Proof. Returning to inequality (2.5) and performing a similar analysis to that in the proof of Lemma 42, we see that if $t \leq 8$, then $2p^2(q^2 - 1) - p(t - 1)(q + 2) + (t - 1) < 0$ for some value of p if and only if

$$\frac{(t-1) - \sqrt{(t-1)^2 + (9-t)(t+1)}}{\frac{1}{2}(9-t)} < q < \frac{(t-1) + \sqrt{(t-1)^2 + (9-t)(t+1)}}{\frac{1}{2}(9-t)}.$$

The lower bound for q described above is immaterial since for $t \leq 8$ it is always negative. The upper bound completes the proof of Lemma 43. □

Using Lemma 43, we generated the following table of possible q values that obey the inequality in (2.6). Since we have already determined the entire edit distance function for $\text{Forb}(K_{2,3})$ and $\text{Forb}(K_{2,4})$, only $t = 5, 6, 7, 8$ needed to be considered:

t	possible q values
5	5
6	none
7	7, 13
8	8, 29

A case analysis of the $f_K(p)$ functions corresponding to these q values finds no improvement to the bounds established by $\min\{p(1-p), \frac{3p+1}{t+5}, \frac{1-p}{t-1}\}$, except in the cases when $t = 7$ and $q = 13$, and $t = 8$ and $q = 29$. In these cases, we see an improvement for the approximate ranges $p \in (0.125, 0.1358)$ and $p \in (0.0625, 0.06667)$, respectively, but even these improvements are surpassed by results from strongly regular graph constructions.

2.6 Figures

t values	parameters	$f_K(p)$
$t \geq 5$	(13, 6, 2, 3) (40, 12, 2, 4) (96, 19, 2, 4)	$(1 + 5p)/13$ $(1 + 26p)/40$ $(1 + 75p)/96$
$t \geq 6$	(10, 6, 3, 4) (17, 8, 3, 4) (26, 10, 3, 4) (85, 20, 3, 5)	$(1 + 2p)/10$ $(1 + 7p)/17$ $(1 + 14p)/26$ $(1 + 63p)/85$

t values	parameters	$f_K(p)$
$t \geq 7$	(16, 9, 4, 6) (36, 14, 4, 6) (49, 16, 3, 6) (64, 18, 2, 6) (100, 22, 0, 6) (156, 30, 4, 6)	$(1 + 5p)/16$ $(1 + 20p)/36$ $(1 + 31p)/49$ $(1 + 44p)/64$ $(1 + 76p)/100$ $(1 + 124p)/156$
$t \geq 8$	(25, 12, 5, 6) (76, 21, 2, 7) (125, 28, 3, 7)	$(1 + 11p)/25$ $(1 + 53p)/76$ $(1 + 95p)/125$

Figure 2.2 Above are the known parameters (see [15]) and $f_K(p)$ functions from strongly regular graphs that provide an improvement upon the known upper bound for $ed_{\mathcal{H}}(p)$ for some interval of p values, where $\mathcal{H} = \text{Forb}(K_{2,t})$ for $5 \leq t \leq 8$. Parameters with resulting bounds surpassed by other strongly regular graph constructions are omitted.

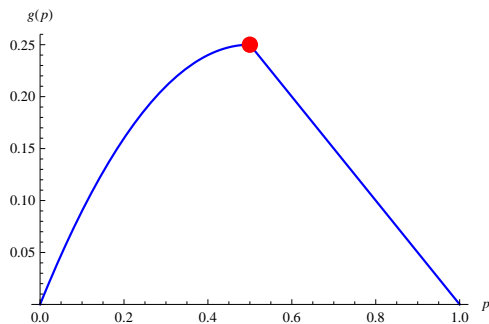


Figure 2.3 Plot of $ed_{\text{Forb}(K_{2,3})}(p) = \min\{p(1-p), (1-p)/2\}$. The point $(p^*, d^*) = (1/2, 1/4)$ is indicated.

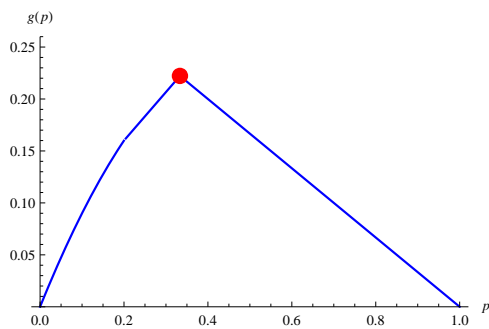


Figure 2.4 Plot of $ed_{\text{Forb}(K_{2,4})}(p) = \min\{p(1-p), (1+7p)/15, (1-p)/3\}$. The point $(p^*, d^*) = (1/3, 2/9)$ is indicated.

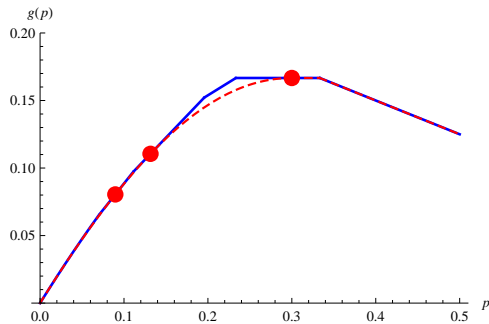


Figure 2.5 Upper and lower bounds (in solid and dashed respectively) for $ed_{\text{Forb}(K_{2,5})}(p)$. Points indicate tangency.

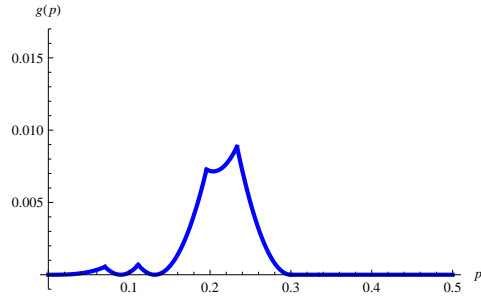


Figure 2.6 Difference between upper and lower bounds for $ed_{\text{Forb}(K_{2,5})}(p)$.

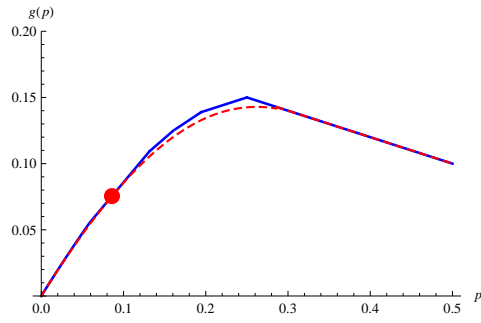


Figure 2.7 Upper and lower bounds (in solid and dashed respectively) for $ed_{\text{Forb}(K_{2,6})}(p)$. Points indicate tangency.

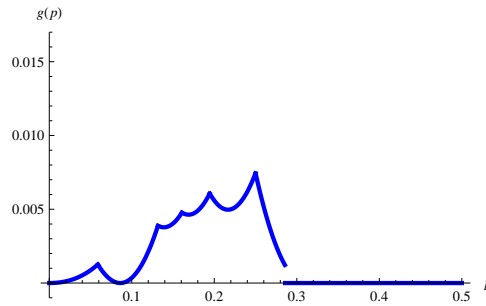


Figure 2.8 Difference between upper and lower bounds for $ed_{\text{Forb}(K_{2,6})}(p)$.

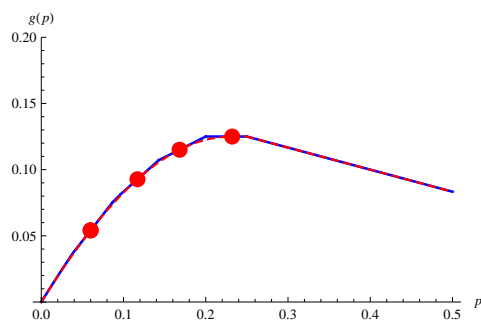


Figure 2.9 Upper and lower bounds (in solid and dashed respectively) for $ed_{\text{Forb}(K_{2,7})}(p)$. Points indicate tangency.

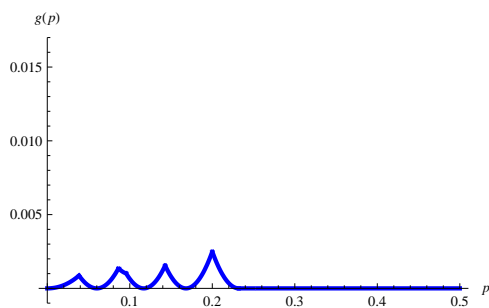


Figure 2.10 Difference between upper and lower bounds for $ed_{\text{Forb}(K_{2,7})}(p)$.

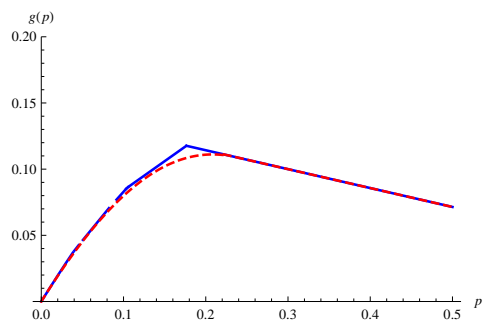


Figure 2.11 Upper and lower bounds (in solid and dashed respectively) for $ed_{\text{Forb}(K_{2,8})}(p)$. In this instance there are no points of tangency.

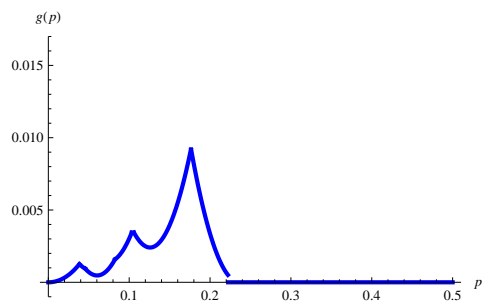


Figure 2.12 Difference between upper and lower bounds for $ed_{\text{Forb}(K_{2,8})}(p)$.

CHAPTER 3. MORE ON CYCLE POWERS AND FÜREDI'S CONSTRUCTION

3.1 General cycle constructions

As mentioned in Section 2.5.2, cycle constructions are one possible type of CRG that can be used to provide an upper bound for the edit distance function of $K_{2,t}$. Recall that if C_k^r is the cycle on k vertices raised to the r th power, then $C_{k,r}$ is defined to be the CRG on k black vertices with white edges corresponding to those in C_k^r , and gray edges corresponding to those in the complement of C_k^r . Furthermore,

$$f_{C_{k,r}}(p) = \left(\frac{2r-1}{k} \right) p + \frac{1}{k}.$$

In Chapter 2, the bound $ed_{\text{Forb}(K_{2,t})}(p) \leq \frac{3p+1}{5+t}$ is found by demonstrating that $C_{5+t,2}$ forbids a $K_{2,t}$ embedding. In this section, we explore what other CRG cycle constructions forbid $K_{2,t}$ embedding, and demonstrate that $C_{5+t,2}$ is indeed the best one for our purposes.

3.1.1 The best k for a given value of r

$f_{C_{k,r}}(p)$ is monotone decreasing with respect to k . Thus, for a given r and t , the goal is to find the largest k so that $C_{k,r}$ contains neither a gray $K_{2,t}$ nor a gray book B_{t-2} (see Theorem 14). With the goal of maximizing k with respect to this constraint in mind, it is reasonable to assume that $k \geq 3r + 2$.

Let $N_G(v)$ denote the neighborhood of v in $C_{k,r}$ via gray edges and $N_W(v)$ denote the neighborhood via white edges. If v_1, v_2 are any two vertices in $C_{k,r}$, then $|N_W(v_1) \cup N_W(v_2) \setminus \{v_1, v_2\}| \geq 2r$. This is because a single vertex in the white subgraph of $C_{k,r}$ has $2r$ neighbors. The largest common neighborhood for v_1 and v_2 occurs when they were neighbors in the original cycle C_k , in

which case $|N_W(v_1) \cup N_W(v_2)| \{v_1, v_2\} = 2r$. Therefore, $|N_G(v_1) \cup N_G(v_2)| \{v_1, v_2\} \leq k - 2r - 2$. Thus to forbid a gray $K_{2,t}$, we want $k - r - 2 < t \Rightarrow k \leq 2r + t + 1$.

To forbid a gray B_{t-2} , if v_1, v_2 are any two vertices adjacent via a gray edge in $C_{k,r}$, then by similar reasoning, $|N_G(v_1) \cup N_G(v_2)| \{v_1, v_2\} \leq k - 3r - 2$.

Thus, comparing these two inequalities yields that the optimal values of k for fixed r such that $C_{k,r}$ still forbids $K_{2,t}$ embedding are

$$\begin{aligned} k &= 3r + t - 1 = 2 + t \text{ for } r = 1 \\ k &= 2r + t + 1 \text{ for } r \geq 2 \end{aligned}$$

3.1.2 Why $r = 2$ is optimal

From the above equations, we have that for fixed r , the optimal $f_{C_{k,r}}(p)$ functions are

$$\begin{aligned} f_{C_{2+t,r}}(p) &= \left(\frac{1}{2+t} \right) p + \frac{1}{2+t} \text{ for } r = 1 \\ f_{C_{2r+t+1,r}}(p) &= \left(\frac{2r-1}{2r+t+1} \right) p + \frac{1}{2r+t+1} \text{ for } r \geq 2 \end{aligned}$$

In the case where $r \geq 2$, $r = 2$ is always the optimal choice to minimize $f_{C_{2r+t+1,r}}(p)$. One way to see this is to observe that $f_{C_{2r+t+1,r}}(p)$ is linear with respect to p and that all such lines pass through the point $p = \frac{1}{2+t}$ with slopes that increase with r . What is more, these functions only improve upon $p(1-p)$ for $p \in [\frac{2+t-\sqrt{-8r+t^2}}{2(1+2r+t)}, \frac{2+t+\sqrt{-8r+t^2}}{2(1+2r+t)}]$, and $\frac{1}{2+t} \leq \frac{2+t-\sqrt{-8r+t^2}}{2(1+2r+t)}$, so the least value of r is best. A similar type of argument also shows that $r = 2$ is always better than $r = 1$.

Thus, $f_{C_{t+5,2}}(p)$ is always optimal.

3.2 Füredi's construction from [29] described

The proof of Theorem 23 alludes to the construction employed by Füredi in [29] to address the Zarankiewicz problem for $K_{2,t}$. For completeness, the construction is outlined below.

Construction 44 (Füredi [29]). *The Construction of a $K_{2,t}$ -free graph G as described in [29] can be accomplished as follows:*

1. Find a prime power q such that $t - 1$ divides $q - 1$.
2. Let \mathbb{F} be the finite field of order q and H be a subgroup of the multiplicative group of $\mathbb{F} \setminus \{0\}$ so that $|H| = t - 1$.
3. Divide the ordered pairs $(a, b) \in \mathbb{F} \times \mathbb{F}$ such that $(a, b) \neq (0, 0)$ into equivalence classes with $(a, b) \sim (a', b')$ if $(a', b') = (ha, hb)$ for some $h \in H$. These equivalence classes are the vertices of G .
4. To define the edges of G , let the vertices corresponding to the distinct equivalence classes of (a, b) and (x, y) be adjacent if and only if $ax + by \in H$.

By demonstrating that the system of equations

$$\begin{aligned} ax + by &= h^\alpha \\ a'x + b'y &= h^\beta \end{aligned}$$

has at most $t - 1$ equivalence classes in its solution set, [29] shows that G cannot contain a $K_{2,t}$. While in [29] graphs do not contain loops, and therefore the vertices of G had degree q if $a^2 + b^2 \notin H$ and degree $q - 1$ if $a^2 + b^2 \in H$, if we allow loops in G for vertices of degree $q - 1$, then the common neighborhood of each pair of vertices is still of order at most $t - 1$ for the same reason as before. Thus, we can construct a $K_{2,t}$ -free, q -regular bipartite graph G' as described in the proof of Theorem 23.

CHAPTER 4. CONCLUSIONS AND FUTURE DIRECTIONS

The results and methods in this thesis demonstrate how a combination of symmetrization techniques from [33], observations from the study of 2-coloured multigraphs, and classical and extremal graph theory constructions can be applied to learn more about the edit distance function for the hereditary property $\text{Forb}(K_{2,t})$.

While $\text{Forb}(K_{2,t})$ is only one of many types of hereditary properties for which this function is unknown, the results in this specific case are interesting for several reasons.

First, the connection between our lower bound in Theorem 21 and upper bounds from $(k, d, t-3, t-1)$ -strongly regular graphs begs the question of not only what advances in knowledge of strongly regular graphs can do to help determine the edit distance function for these hereditary properties, but also whether more knowledge of the edit distance for these properties could lend some insight into the existence of these types of graphs. To the best of my knowledge, it is unknown whether there is a finite or infinite number of strongly regular graphs of this form [19].

Second, the emerging impact of Füredi's constructions for large t leads to the question of what is happening to the optimal p -core CRG constructions for small p as t gets large?

Third, what is the significance of entire intervals of p^* values corresponding to the maximum of the edit distance function for odd t , and when can we expect these intervals in general?

There are also still a number of other open questions and avenues for future work in this area. For instance, questions of stability: how many different p -core colored regularity graphs witness the exact value of this or other hereditary properties for a fixed p and \mathcal{H} ? One way to view these p -core constructions is as a template for an optimal way to edit the edges of the random graph $G(n, p)$ to make it a member of a hereditary property. Therefore, it is natural to ask questions about how many such templates exist.

Another avenue for future work is to answer the question of what can be said about analogous functions for graph metrics other than edit distance? Can we define a similar such function using these metrics, and if so, will it be continuous and/or concave?

BIBLIOGRAPHY

- [1] Alekseev, V. E. (1992). Range of values of entropy of hereditary classes of graphs. *Diskret. Mat.*, 4(2):148–157.
- [2] Alon, N., Fischer, E., Krivelevich, M., and Szegedy, M. (2000). Efficient testing of large graphs. *Combinatorica*, 20(4):451–476.
- [3] Alon, N. and Shapira, A. (2008). A characterization of the (natural) graph properties testable with one-sided error. *SIAM J. Comput.*, 37(6):1703–1727.
- [4] Alon, N. and Stav, U. (2008a). The maximum edit distance from hereditary graph properties. *J. Combin. Theory Ser. B*, 98(4):672–697.
- [5] Alon, N. and Stav, U. (2008b). What is the furthest graph from a hereditary property? *Random Structures Algorithms*, 33(1):87–104.
- [6] Alon, N. and Stav, U. (2009). Stability-type results for hereditary properties. *J. Graph Theory*, 62(1):65–83.
- [7] Axenovich, M., Kézdy, A., and Martin, R. (2008). On the editing distance of graphs. *J. Graph Theory*, 58(2):123–138.
- [8] Axenovich, M. and Martin, R. (2006). Avoiding patterns in matrices via a small number of changes. *SIAM J. Discrete Math.*, 20(1):49–54 (electronic).
- [9] Balogh, J. and Martin, R. (2008). Edit distance and its computation. *Electron. J. Combin.*, 15(1):Research Paper 20, 27.
- [10] Bollobás, B. (2001). *Random graphs*, volume 73 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, second edition.

- [11] Bollobás, B. and Thomason, A. (1995). Projections of bodies and hereditary properties of hypergraphs. *Bull. London Math. Soc.*, 27(5):417–424.
- [12] Bollobás, B. and Thomason, A. (1997). Hereditary and monotone properties of graphs. In *The mathematics of Paul Erdős, II*, volume 14 of *Algorithms Combin.*, pages 70–78. Springer, Berlin.
- [13] Bollobás, B. and Thomason, A. (2000). The structure of hereditary properties and colourings of random graphs. *Combinatorica*, 20(2):173–202.
- [14] Brandt, S. (1999). On the structure of dense triangle-free graphs. *Combin. Probab. Comput.*, 8(3):237–245.
- [15] Brouwer, A. E. Parameters of strongly regular graphs. <http://www.win.tue.nl/~aeb/graphs/srg/srgtab.html> (accessed July 1, 2010).
- [16] Brown, W. G. (1966). On graphs that do not contain a Thomsen graph. *Canad. Math. Bull.*, 9:281–285.
- [17] Chen, D., Eulenstein, O., Fernández-Baca, D., and Sanderson, M. (2002). Supertrees by flipping. In *Computing and combinatorics*, volume 2387 of *Lecture Notes in Comput. Sci.*, pages 391–400. Springer, Berlin.
- [18] Chen, G., Faudree, R. J., and Gould, R. J. (2008). Saturation numbers of books. *Electron. J. Combin.*, 15(1):Research Paper 118, 12.
- [19] Elzinga, R. J. (2003). Strongly regular graphs: values of λ and μ for which there are only finitely many feasible (v, k, λ, μ) . *Electron. J. Linear Algebra*, 10:232–239 (electronic).
- [20] Erdős, P. (1964/1965). On an extremal problem in graph theory. *Colloq. Math.*, 13:251–254.
- [21] Erdős, P., Faudree, R., Pach, J., and Spencer, J. (1988). How to make a graph bipartite. *J. Combin. Theory Ser. B*, 45(1):86–98.

- [22] Erdős, P., Füredi, Z., Rothschild, B. L., and Sós, V. T. (1999). Induced subgraphs of given sizes. *Discrete Math.*, 200(1-3):61–77. Paul Erdős memorial collection.
- [23] Erdős, P., Gyárfás, A., and Ruszinkó, M. (1998). How to decrease the diameter of triangle-free graphs. *Combinatorica*, 18(4):493–501.
- [24] Erdős, P., Győri, E., and Simonovits, M. (1992). How many edges should be deleted to make a triangle-free graph bipartite? In *Sets, graphs and numbers (Budapest, 1991)*, volume 60 of *Colloq. Math. Soc. János Bolyai*, pages 239–263. North-Holland, Amsterdam.
- [25] Erdős, P. and Rényi, A. (1959). On random graphs. I. *Publ. Math. Debrecen*, 6:290–297.
- [26] Erdős, P. and Simonovits, M. (1966). A limit theorem in graph theory. *Studia Sci. Math. Hungar*, 1:51–57.
- [27] Erdős, P. and Stone, A. H. (1946). On the structure of linear graphs. volume 52, pages 1087–1091.
- [28] Füredi, Z. (1991). Turán type problems. In *Surveys in combinatorics, 1991 (Guildford, 1991)*, volume 166 of *London Math. Soc. Lecture Note Ser.*, pages 253–300. Cambridge Univ. Press, Cambridge.
- [29] Füredi, Z. (1996). New asymptotics for bipartite Turán numbers. *J. Combin. Theory Ser. A*, 75(1):141–144.
- [30] Janson, S., Luczak, T., and Rucinski, A. (2000). *Random graphs*. Wiley-Interscience Series in Discrete Mathematics and Optimization. Wiley-Interscience, New York.
- [31] Komlós, J. and Simonovits, M. (1996). Szemerédi’s regularity lemma and its applications in graph theory. In *Combinatorics, Paul Erdős is eighty, Vol. 2 (Keszthely, 1993)*, volume 2 of *Bolyai Soc. Math. Stud.*, pages 295–352. János Bolyai Math. Soc., Budapest.
- [32] Marchant, E. and Thomason, A. (2010). Extremal graphs and multigraphs with two weighted colours. In *Fete of combinatorics and computer science*, volume 20 of *Bolyai Soc. Math. Stud.*, pages 239–286. János Bolyai Math. Soc., Budapest.

- [33] Martin, R. The edit distance function and symmetrization. [arXiv:1007.1897v3](#).
- [34] Martin, R. and McKay, T. On the edit distance from $K_{2,t}$ -free graphs. Submitted.
- [35] Prömel, H. J. and Steger, A. (1991). Excluding induced subgraphs: quadrilaterals. *Random Structures Algorithms*, 2(1):55–71.
- [36] Prömel, H. J. and Steger, A. (1992). Excluding induced subgraphs. III. A general asymptotic. *Random Structures Algorithms*, 3(1):19–31.
- [37] Prömel, H. J. and Steger, A. (1993). Excluding induced subgraphs. II. Extremal graphs. *Discrete Appl. Math.*, 44(1-3):283–294.
- [38] Scheinerman, E. R. and Zito, J. (1994). On the size of hereditary classes of graphs. *J. Combin. Theory Ser. B*, 61(1):16–39.
- [39] Szemerédi, E. (1978). Regular partitions of graphs. In *Problèmes combinatoires et théorie des graphes (Colloq. Internat. CNRS, Univ. Orsay, Orsay, 1976)*, volume 260 of *Colloq. Internat. CNRS*, pages 399–401. CNRS, Paris.
- [40] Thomason, A. Graphs, colours, weights and hereditary properties. Preprint.
- [41] West, D. B. (2001). *Introduction to Graph Theory*. Prentice Hall, Upper Saddle River, NJ, 2nd edition.