The edit distance from a cycle- and squared cycle-free graph

by

Chelsea Kay Peck

A thesis submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Major: Mathematics

Program of Study Committee:

Ryan Martin, Major Professor

Steven Butler

Daniel Nordman

Iowa State University

Ames, Iowa

2013

DEDICATION

I would like to dedicate my thesis to my husband, Brian, and my daughter, Audra, who have been helpful, supportive, and gracious toward me during my last few months of graduate school.

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGEMENTS

I would like to thank my major professor, Dr. Ryan Martin, for his guidance in all of the stages of the research that has gone into this thesis, from understanding the problem initially to giving input on the final written drafts. He has been incredibly supportive and patient throughout the whole process.

Thank you to my entire POS Committee for their time and feedback.

I would also like to thank Maria Axenovich, who believed in me enough as a freshman undergraduate student to get me started doing graph theory research.

This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. DGE0751279.

ABSTRACT

The edit distance from a hereditary property is the fraction of edges in a graph that must be added or deleted for a graph to become a member of that hereditary property. Let Forb(C_h) and Forb(C_h^2) denote the hereditary properties containing graphs with no induced cycle or squared cycle on h vertices, respectively. The edit distance from Forb(C_h) is found for odd values of h, and the maximum edit distance is found for all values of h. The edit distance is found for Forb(C_h^2) for $h = 8, 9, 10$, and the maximum value is known for $h = 11, 12$, with partial results for other values of h.

CHAPTER 1. INTRODUCTION

1.1 Motivation

Edit distance problems concern the question: how how many edges need to be added or removed (edited) in a graph G so that it will have a certain property? A graph can be edited by adding or deleting edges between vertices. The edit distance between two graphs G and H each on n vertices is the minimum number of edge edits to G needed to make G and H isomorphic. This distance is often normalized based on the total number of possible edges of a graph on n vertices.

Edit distance in graphs is similar to the concept of Hamming distance. The Hamming distance between two binary strings of equal length is the number of digits in the first string that must to be changed to make it identical to the second string. Hamming distance is also used in linear algebra, where it is defined as the number of entries that differ between two matrices of the same size. Since graphs can be represented by adjacency matrices, which are symmetric $\{0, 1\}$ matrices, the edit distance between two graphs would correspond to the Hamming distance between two such matrices [3].

The concept of edit distance may be useful in fields outside of mathematics. Several applications involve constructing graph representations of phenomena, and it is often of interest whether such a graph has a certain property.

In biology, phylogenetic trees are used to see evolutionary relationships between organisms. Smaller phylogenetic trees can be combined into a larger "supertree" in order to see larger-scale relationships. The smaller trees, however, often have discrepancies that must be corrected in order to make a large tree that is most faithful to the information from its component trees. A graph representation of the data's compatibility can be created in the form of a bipartite graph which needs to be edited by adding or deleting edges to create a graph that has no induced path on 5 vertices [3, 6]. Data compatibility is also a factor in biology when DNA fragments are reconstructed. An interval graph can be used to represent this data [11]. The edit distance from the set of all interval graphs was studied by Alon and Stav [2]. Finally, graph representations of metabolic pathways can be constructed. If there is an undesirable metabolic process that corresponds to a small induced graph, edit distance could be used to find the amount of edge editing that must be done to avoid the graph that represents that process [3].

In computer science, several theoretical questions relating to edit distance are of interest [3, 8]. Edit distance is also important in questions of computing. For example, it is favorable to find an acyclic database scheme, which has a unique path for retrieving pieces of data [11]. A graph which represents such a scheme is a chordal graph. The edit distance from the set of chordal graphs has been studied by Alon and Stav [2].

Several other types of graph properties that are of general interest in graph theory and computer science have been studied with regards to edit distance. Examples of these include perfect graphs, split graphs, permutation graphs, and forbidden induced subgraphs [2, 4, 12, 13, 7].

1.2 Basic Definitions

These definitions are based on work by Balogh and Martin [4].

Definition 1. *A hereditary property $\mathcal{H}$ is a class of simple graphs that is closed under vertex deletion.*

Definition 2. *If H is a simple graph, then* $\mathrm{Forb}(H)$ *is the hereditary property that includes all graphs without an induced copy of H. We call such a property a principle hereditary property.*

Any hereditary property may be written as the intersection of principle hereditary properties [4], but here we will restrict our attention to hereditary properties $\mathcal{H}$ which are of the form $\mathrm{Forb}(H)$ for a single graph H.

Definition 3. *The edit distance between two graphs G and H each on n vertices is defined as* *the number of edges that must be added or removed from G to form a graph isomorphic to H. We write* $\operatorname{dist}(G, H) = |E(G) \triangle E(H)|$.

Definition 4. *The edit distance between a graph G and a hereditary property $\mathcal{H}$ is defined as follows:*

$$\operatorname{dist}(G, \mathcal{H}) = \min\{\operatorname{dist}(G, H) : H \in \mathcal{H}\}. \tag{1.1}$$

Definition 5. *We define the edit distance of a hereditary property $\mathcal{H}$ from graphs on n vertices as*

$$\operatorname{dist}(n, \mathcal{H}) = \max\{\operatorname{dist}(G, \mathcal{H}) : |V(G)| = n\}. \tag{1.2}$$

Definition 6. *The edit distance function is the distance of the furthest graph from the hereditary property, and it is normalized by the possible number of edges in the graph. The edit distance function is defined by Balogh and Martin in [4] as follows:*

$$ed_{\mathcal{H}}(p) = \lim_{n\to\infty} \max\left\{\operatorname{dist}(G, \mathcal{H}) : |V(G)| = n; |E(G)| = \left\lfloor p\tbinom{n}{2}\right\rfloor\right\} / \tbinom{n}{2}. \tag{1.3}$$

In [4], Balogh and Martin proved that the edit distance function may also be expressed as follows:

$$ed_{\mathcal{H}}(p) = \lim_{n\to\infty} \mathbb{E}[\operatorname{dist}(G(n,p), \mathcal{H})]/\tbinom{n}{2}, \tag{1.4}$$

where $\mathbb{E}[X]$ denotes the expectation of a random variable X. This definition makes use of the the Erdős-Rényi random graph $G(n,p)$ on n vertices with each pair of vertices connected by an edge independently with probability p. In fact, the edit distance function is equal to the normalized expected edit distance for $G(n,p)$ from $\mathcal{H}$, as expressed in Equation 1.4.

For a given hereditary property $\mathcal{H}$, the edit distance function realizes a maximum $d^*_{\mathcal{H}}$, often written d^*, that is given by

$$d^*_{\mathcal{H}} = \lim_{n\to\infty} \operatorname{dist}(n, \mathcal{H})/\tbinom{n}{2}. \tag{1.5}$$

In [4] it was shown that these limits exist and that the edit distance function is both continuous and concave down. Thus, this maximum value may be realized at a single value $p^* \in [0, 1]$ or over a nondegenerate subinterval of $[0, 1]$.

1.3 Main Results

The main results in this paper concern the edit distance function $ed_{\mathcal{H}}(p)$ when $\mathcal{H} = \mathrm{Forb}(C_h)$ or $\mathcal{H} = \mathrm{Forb}(C_h^2)$, where C_h is the cycle on h vertices and C_h^2 is the squared cycle on h vertices.

For $\mathcal{H} = \mathrm{Forb}(C_h)$, we split the results into two cases depending on whether the cycle is even or odd.

Theorem 7. *Let C_h be a cycle on $h > 3$ vertices. Let $\mathcal{H} = \mathrm{Forb}(C_h)$. If h is odd, then for $0 \le p \le 1$,*

$$ed_{\mathcal{H}}(p) = \min\left\{\frac{p}{2}, \frac{p(1-p)}{1+p(\lceil h/3\rceil - 2)}, \frac{1-p}{\lceil h/2\rceil - 1}\right\}. \tag{1.6}$$

If h is even, then for $1/\lceil h/3\rceil \le p \le 1$,

$$ed_{\mathcal{H}}(p) = \min\left\{\frac{p(1-p)}{1+p(\lceil h/3\rceil - 2)}, \frac{1-p}{\lceil h/2\rceil - 1}\right\}. \tag{1.7}$$

Moreover, for $0 \le p \le 1$, $ed_{\mathcal{H}}(p) \le \min\left\{\frac{p(1-p)}{1+p(\lceil h/3\rceil - 2)}, \frac{1-p}{\lceil h/2\rceil - 1}\right\}$.

Plots of the function $ed_{\mathrm{Forb}(C_h)}(p)$ for $4 \le h \le 18$ are shown in Figures A.1 through A.15 in Appendix A.

The functions given in Equation 1.6 and Equation 1.7 were proven as being the upper bound for the edit distance function by Martin in [13], and were proven to be the edit distance function for $h < 10$. Thus, the result proven here is identical to the result in [13] for $h = 5, 7, 9$ and is weaker than the result proven for $h = 6, 8$.

The three parts of the equation correspond to three colored regularity graphs (CRGs), algorithms for editing a graph to remove induced copies of C_h, which are each optimal on a different range of p. We define CRGs in Section 2 below.

Note that although the function for even h is known only for large p, the function achieves its maximum value on this range due to its concavity and continuity properties. Also, this formula serves as an upper bound on the edit distance function when p is small.

Corollary 8. *For all h not in $\{4, 7, 8, 10, 16\}$, the edit distance function achieves its maximum at the point*

$$(p^*, d^*) = \left(\frac{1}{\lceil h/2\rceil - \lceil h/3\rceil + 1}, \frac{\lceil h/2\rceil - \lceil h/3\rceil}{(\lceil h/2\rceil - 1)(\lceil h/2\rceil - \lceil h/3\rceil + 1)}\right).$$

For $h \in \{4, 7, 8, 10, 16\}$,

$$(p^*, d^*) = \left(\frac{1}{1 + \sqrt{\lceil h/3 \rceil - 1}}, \frac{1}{\lceil h/3 \rceil + 2\sqrt{\lceil h/3 \rceil - 1}} \right).$$

For most values of h, the maximum point (p^*, d^*) corresponds to the point at which two of the curves which make up the edit distance function intersect. For the five other values of h, the maximum corresponds to the local maximum of the curve $\frac{p(1-p)}{1+p(\lceil h/3 \rceil - 2)}$.

We also prove some results concerning the edit distance from $\mathrm{Forb}(C_h^2)$.

Definition 9. *Let h be an integer. The square of a cycle of length h, C_h^2, is the graph formed by taking a cycle of length h and adding edges between vertices which have only one vertex between them on the cycle.*

Theorem 10. *Let C_h^2 denote the squared cycle on h vertices. Then*

i. $ed_{\mathrm{Forb}(C_8^2)}(p) = \min\left\{\frac{p}{3}, \frac{p(1-p)}{2-p}, \frac{1-p}{2}\right\}$, $(p^*, d^*) = (2 - \sqrt{2}, 3 - 2\sqrt{2})$

ii. $ed_{\mathrm{Forb}(C_9^2)}(p) = \min\left\{\frac{p(1-p)}{2-p}, \frac{p(1-p)}{1+p}\right\}$, $(p^*, d^*) = (1/2, 1/6)$

iii. $ed_{\mathrm{Forb}(C_{10}^2)}(p) = \min\left\{\frac{1-p}{3}, \frac{p}{3}\right\}$, $(p^*, d^*) = (1/2, 1/6)$

iv. Let $p \leq 1/2$. *Then* $ed_{\mathrm{Forb}(C_{11}^2)}(p) = \min\left\{\frac{1-p}{3}, \frac{p}{3}, \frac{p(1-p)}{2}\right\}$, $(p^*, d^*) = (1/2, 1/8)$.

v. Let $p \leq 1/2$. *Then* $ed_{\mathrm{Forb}(C_{12}^2)}(p) = \min\left\{\frac{p(1-p)}{2}, \frac{1-p}{3}\right\}$, $(p^*, d^*) = (1/2, 1/8)$.

Plots of the function $ed_{\mathrm{Forb}(C_h^2)}(p)$ for $8 \leq h \leq 12$ are shown in Figures B.1 through B.5 in Appendix B.

In addition to the results in Theorem 10, several results about upper bounds for the edit distance from $\mathrm{Forb}(C_h^2)$ are proven for general values of h.

CHAPTER 2. BACKGROUND

2.1 Colored Regularity Graphs (CRGs)

In this section we will introduce the main tool that is used to find the edit distance function. The definitions given here are based on the work of Balogh and Martin [4].

Definition 11. *A colored regularity graph (CRG), K, is a complete graph whose vertices are colored black and white and whose edges are colored black, white, and gray. The vertex set $V(K)$ of the CRG can be partitioned into a set of black vertices $VB(K)$ and a set of white vertices $VW(K)$. Likewise, the edge set $E(K)$ of K can be partitioned into a set of black edges $EB(K)$, a set of white edges $EW(K)$, and a set of gray edges $EG(K)$.*

Definition 12. *Given a simple graph H and a CRG K, we say that H embeds in K if there exists a function $\phi : V(H) \to V(K)$ such that for any $v, w \in V(H)$:*

- *If vw is an edge of H, then either v and w map to the same black vertex in K or to different vertices in K which are connected by an edge that is either black or gray (formally, either $\phi(v) = \phi(w) \in VB(K)$ or $\phi(v)\phi(w) \in EB(K) \cup EG(K)$).*

- *If vw is a nonedge of H, then either v and w map to the same white vertex in K or to different vertices in K which are connected by an edge that is either white or gray (formally, either $\phi(v) = \phi(w) \in VW(K)$ or $\phi(v)\phi(w) \in EW(K) \cup EG(K)$).*

Informally, an embedding of H in K maps the adjacent vertices of H to black and gray in K, while it maps the nonadjacent vertices of H to white and gray in K.

Definition 13. *Given a hereditary property $\mathcal{H}$ such that $\mathcal{H} = \mathrm{Forb}(H)$, $\mathcal{K}(\mathcal{H})$ denotes the set of all CRGs K such that H does not embed in K.*

Note that if H does not embed in a CRG K, then any graph with H as an induced subgraph will not embed in K. In other words, if G embeds in some $K \in \mathcal{K}(\mathcal{H})$, then $G \in \mathcal{H}$. Thus, if a graph can be edited so that it embeds in a CRG contained in $\mathcal{K}(\mathcal{H})$, it will be in the hereditary property.

The notation $K(a, b)$ is used to denote a CRG with all gray edges that has a white vertices and b black vertices.

Example 14. *Let $H = C_7$, the cycle on 7 vertices. Consider the following CRGs that are not in the set $\mathcal{K}(\text{Forb}(C_7))$ (i.e., we are considering CRGs for which there is an embedding of C_7):*

i. *$K(2, 5)$, the CRG with 2 white vertices, 5 black vertices, and all gray edges. The embedding is achieved by mapping any vertex of C_7 to a vertex in the CRG.*

ii. *$K(0, 4)$, the CRG with 4 black vertices and all gray edges. The simple graph C_7 embeds into $K(0, 4)$ as shown in Figure 2.1. The embedding is accomplished by first choosing 3 nonadjacent edges. These three edges are 2-cliques that embed into 3 of the CRG's black vertices. The remaining vertex is a 1-clique that embeds into the last black vertex of $K(2, 0)$. (In the figure, the four cliques of C_7 which embed into the four black vertices of $K(0, 4)$ are circled by dotted lines on the graph C_7.)*

iii. *$K(3, 0)$, the CRG with 4 white vertices and all gray edges. The chromatic number of C_7 is 3, so its vertices may be partitioned into 3 independent sets, which are mapped to each of the white vertices of $K(3, 0)$. (See Figure 2.1. The vertices in each independent set has the same shape in the graph C_7.)*

Example 15. *Let $H = C_7$, the cycle on 7 vertices. Consider the following CRGs that are in the set $\mathcal{K}(\text{Forb}(C_7))$ (i.e., the set of CRGs into which C_7 does not embed):*

i. *$K(2, 0)$, the CRG on two white vertices with a gray edge. Since the chromatic number of C_7 is 3, there is no way to partition the vertices of C_7 into 2 independent sets. Therefore, C_7 does not embed into $K(2, 0)$ and so $K(2, 0) \in \mathcal{K}(\text{Forb}(C_7))$.*

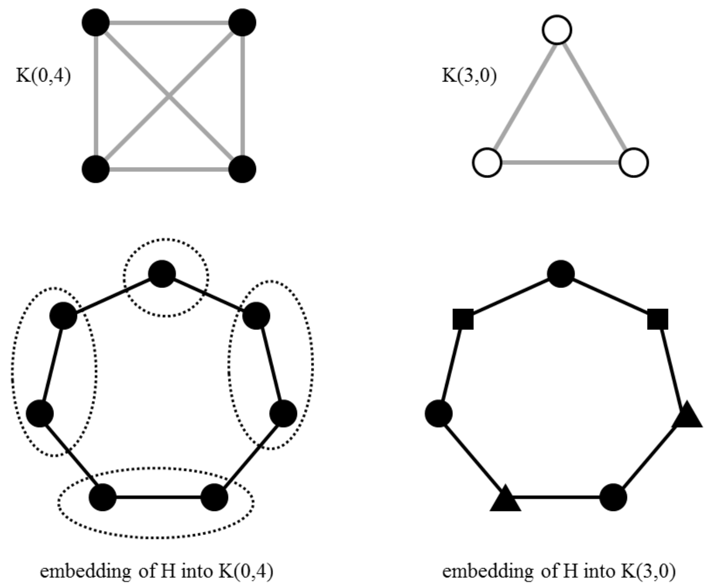

Figure 2.1 Embedding of C_7 into the CRGs $K(0,4)$ and $K(3,0)$.

ii. $K(0,3)$, *the CRG on 3 black vertices with all gray edges. Since the largest clique size in* C_7 *is 2, we cannot partition the 7 vertices of the graph into 3 cliques. So* C_7 *does not embed into* $K(0,3)$ *and thus* $K(0,3) \in \mathcal{K}(\mathrm{Forb}(C_7))$.

iii. $K(1,2)$, *the CRG with 1 white vertex, 2 black vertices, and all gray edges. The largest independent set in* C_7 *has size 3, and the largest clique in* C_7 *has size 2. So if* C_7 *were to embed into this CRG, we would need to partition its vertices into an independent set of size 3 and two cliques of size 2. However, the only independent set of size 3 excludes an edge and two nonadjacent vertices. Thus there is no partition as we desire, and* $K(1,2) \in \mathcal{K}(\mathrm{Forb}(C_7))$.

In fact, these results were generalized for $h > 3$ by Martin [13].

Proposition 16. *[13]*

Let $\mathcal{K}(\mathrm{Forb}(C_h))$ *denote the set of all CRGs* K *such that* C_h *does not embed in* K. *For* $h > 3$,

i. $K(0, \lceil h/2 \rceil - 1) \in \mathcal{K}(\mathrm{Forb}(C_h))$,

ii. $K(1, \lceil h/3 \rceil - 1) \in \mathcal{K}(\mathrm{Forb}(C_h))$, *and* *iii. for h odd, $K(2,0) \in \mathcal{K}(\mathrm{Forb}(C_h))$.*

Definition 17. *[13] Let $\mathcal{H} = \mathrm{Forb}(H)$ be a hereditary property for some graph H. The clique spectrum of $\mathcal{H}$ is the set*

$$\Gamma(\mathcal{H}) = \{(r,s) : H \not\mapsto K(r,s)\}.$$

The point (r,s) is an extreme point in the clique spectrum if $(r,s) \in \Gamma$ but $(r+1,s) \notin \Gamma$ and $(r,s+1) \notin \Gamma$. The set of extreme points is denoted Γ^*.

2.2 The f and g functions

Since the ability of a graph G to embed into a CRG K in $\mathcal{K}(\mathcal{H})$ implies its membership in the hereditary property $\mathcal{H}$, we are interested in the amount of editing that must be done to enable a graph to embed into some $K \in \mathcal{K}(\mathcal{H})$.

Each CRG $K \in \mathcal{K}(\mathcal{H})$ is itself a 'recipe' for editing a graph G to make it a member of the hereditary property $\mathcal{H}$. This recipe is as follows: For a $K \in \mathcal{K}(\mathcal{H})$ on k vertices, partition $V(G)$ into k sets. Each of these sets corresponds to a vertex $v \in V(K)$. Let S be the set in the partition of $V(G)$ which corresponds to the vertex v in the CRG. If v is black, then G is edited by adding all possible edges between vertices in S. If v is white, then G is edited by deleting all edges between vertices in S. The edges of the CRG contribute to the recipe in the following way: let S_1 and S_2 be sets of the partition of $V(G)$ which correspond to the vertices v_1 and v_2, respectively. Then if the edge $v_1v_2 \in E(K)$ is black, all possible edges with one vertex in S_1 and the other in S_2 are added. If v_1v_2 is white, all edges which have one vertex in S_1 and the other in S_2 are deleted. If v_1v_2 is gray, then no editing is done to the edges and non-edges with one vertex in S_1 and the other in S_2. A graph G edited according to this recipe given by the CRG K is a member of the hereditary property $\mathcal{H}$.

Recall from Equation 1.4 that the edit distance can be computed based on the expected distance of the Erdős-Rényi random graph $G(n,p)$ from the hereditary property. Thus, we can use CRGs to define a function that expresses the expectation of how much editing must be done to $G(n,p)$ in order for it to embed into a member of $\mathcal{K}(\mathcal{H})$. In other words, the function expresses the amount of editing that is done when the graph G is edited according to the recipe given by the CRG K. These functions were defined by Balogh and Martin [4].

The first function, $f_K(p)$, corresponds to the editing that must be done to $G(n,p)$ in order for it to embed into a CRG K by an equipartition. An equipartition is a partition in which the difference in the size of each part is at most one. The equipartition has the same number of parts as K has vertices. Each part with vertices from $G(n,p)$ is embedded into a vertex of the CRG.

The function describes the amount of editing that must be done in order for $G(n,p)$ to embed into a given CRG, K. If a set of vertices from $G(n,p)$ is embedded into a white vertex, all of the edges must be deleted. Edges are present in $G(n,p)$ with probability p, so p times the number of possible edges will need to be edited. Similarly, all possible edges must be added to vertex sets embedded in black vertices of K. Thus, $(1-p)$ times the number of possible edges will be added. From this information we may derive the formula for $f_K(p)$ as follows:

$$f_K(p) = \frac{1}{k^2}[p(|VW(K)| + 2|EW(K)|) + (1-p)(|VB(K)| + 2|EB(K)|). \tag{2.1}$$

However, the editing may be done more efficiently if the vertices are partitioned optimally rather than equipartitioned. The function that expresses the editing that must be done in this optimal partitioning is given by a quadradtic program:

$$g_K(p) = \begin{cases} \min & \mathbf{x}^T\mathbf{M}_K(p)\mathbf{x} \\ \text{s.t.} & \mathbf{x}^T\mathbf{1} = 1 \\ & \mathbf{x} \geq \mathbf{0} \end{cases} \tag{2.2}$$

where $\mathbf{x}$ is the optimal vertex weighting vector. Each entry in $\mathbf{x}$ corresponds to a vertex of the CRG, and its entry is the weighting of that vertex that minimizes the amount of editing that would need to be done by using the corresponding CRG to edit a random graph. The weights must be positive and sum to one. The matrix $M = \mathbf{M}_K(p)$ is essentially a weighted adjacency matrix based on the coloring of the CRG K. For vertices $i, j \in V(K)$, $i \neq j$, the ij^{th} entry of M is p if ij is a white edge, $(1-p)$ if ij is a black edge, and 0 if ij is a gray edge. The i^{th} diagonal entry of M is p if i is a white vertex and $(1-p)$ if i is a black vertex.

There are many ways to compute the $g_K(p)$ function. In simple cases it may be computed directly, but the method of Lagrange multipliers may also be used. Additionally, the following are helpful in some cases:

Definition 18. *Let K' be a sub-CRG of the CRG K. Then K' is a component of K if all edges between $V(K')$ and $V(K)-V(K')$ are gray.*

Thus, components for CRGs are analogous to components in simple graphs (if the white and black edges in the CRG are thought of as edges in the simple graph and the gray CRG edges are thought of as non-edges in the simple graph). If the function $g_{K'}(p)$ is known for each component K' of a CRG K, then we may use the following lemma to compute $g_K(p)$:

Lemma 19. *[13] Let $K^{(1)}, ..., K^{(l)}$ be the components of the CRG K. Then*

$$(g_K(p))^{-1} = \sum_{i=1}^{l} (g_{K^{(i)}}(p))^{-1}.$$

For a CRG with all gray edges, we recognize that a CRG composed of a single white vertex has a g function of p and a CRG composed of a single black vertex has a g function of $1-p$ to obtain the following corollary:

Corollary 20. *[13] Let K be a CRG with only gray edges. Then*

$$g_K(p) = \left(\frac{|VW(K)|}{p} + \frac{|VB(K)|}{1-p} \right)^{-1}.$$

The functions $f_K(p)$ and $g_K(p)$ are directly related to the edit distance function for $\mathcal{H}$ when $K \in \mathcal{K}(\mathcal{H})$. The following relationship was proven by Alon and Stav [1], with the last equality proven by Marchant and Thomason [10]:

$$ed_{\mathcal{H}}(p) = \inf_{K\in\mathcal{K}} \{f_K(p)\} = \inf_{K\in\mathcal{K}} \{g_K(p)\} = \min_{K\in\mathcal{K}} \{g_K(p)\}. \tag{2.3}$$

These equalities give that for each $p \in [0,1]$ there is some CRG $K \in \mathcal{K}$ such that the edit distance function equals $g_K(p)$.

Example 21. *Consider the CRGs from Proposition 16. We will compute the f and g functions that correspond to these CRGs:*

i. $K = K(2,0)$

Since $K(2,0)$ has all gray edges, $|EW(K(2,0))| = |EB(K(2,0))| = 0$. We also have that $|VW(K(2,0))| = 2$ and $|VB(K(2,0))| = 0$. The CRG has 2 total vertices, so $k = 2$. Thus,

$$\begin{aligned} f_K(p) &= \frac{1}{k^2}[p(|VW(K)| + 2|EW(K)|) + (1-p)(|VB(K)| + 2|EB(K)|) \\ &= \frac{1}{4}[2p] \\ &= p/2. \end{aligned}$$

To calculate the g function, we must first construct the matrix M:

$$M = \begin{bmatrix} p & 0 \\ 0 & p \end{bmatrix}.$$

Thus,

$$\begin{aligned} g_K(p) &= \min\{\mathbf{x}^T\mathbf{M}_K(p)\mathbf{x} : \mathbf{x}^T\mathbf{1} = 1; \mathbf{x} \geq \mathbf{0}\} \\ &= \min\left\{[x_1 x_2]\begin{bmatrix} p & 0 \\ 0 & p \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} : x_1 + x_2 = 1; x_1, x_2 \geq 0\right\} \\ &= \min\left\{px_1^2 + px_2^2 : x_1 + x_2 = 1; x_1, x_2 \geq 0\right\} \\ &= \min\{2px_1^2 - 2px_1 + p : x_1 \geq 0\}. \end{aligned}$$

To solve this, we find the local minimum of the quadratic equation occurs at $x_1 = 1/2$. Thus, $g_K = p/2$ and $\mathbf{x} = \begin{bmatrix} 1/2 \\ 1/2 \end{bmatrix}$.

The method of Lagrange multipliers could also have been used here, and it is often a useful approach to solve for the g function directly.

In this case, the f and g functions are equal, which implies that equipartitioning the vertices yields the optimal weighting.

ii. $K = K(0, \lceil h/2 \rceil - 1)$

$$\begin{aligned} f_K(p) &= \frac{1}{k^2}[p(|VW(K)|+2|EW(K)|)+(1-p)(|VB(K)|+2|EB(K)|) \\ &= \frac{1}{(\lceil h/2\rceil-1)^2}[(1-p)(\lceil h/2\rceil-1)] \\ &= \frac{1-p}{\lceil h/2\rceil-1} \end{aligned}$$

To calculate the g function, we observe that the matrix $M=(1-p)I$ since K is composed of all black vertices and gray edges. However, the g function may be easily calculated using the result from Corollary 20.

$$g_K(p)=\left(\frac{|VW(K)|}{p}+\frac{|VB(K)|}{1-p}\right)^{-1}=\frac{1-p}{\lceil h/2\rceil-1}$$

In this case the g function is again equivalent to the f function, implying that the equipartition is optimal.

iii. $K=K(1,\lceil h/3\rceil-1)$

$$\begin{aligned} f_K(p) &= \frac{1}{k^2}[p(|VW(K)|+2|EW(K)|)+(1-p)(|VB(K)|+2|EB(K)|) \\ &= \frac{1}{(\lceil h/3\rceil)^2}[p+(1-p)(\lceil h/3\rceil-1)] \end{aligned}$$

In this case we may compute that the matrix M is as follows:

$$M=\begin{bmatrix} p & 0 & \dots & \dots & 0 \\ 0 & 1-p & 0 & \dots & 0 \\ \vdots & 0 & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & 0 & \dots & 0 & 1-p \end{bmatrix}$$

Instead of calculating $\min\{\mathbf{x}^T\mathbf{M}_K(p)\mathbf{x}:\mathbf{x}^T\mathbf{1}=1;\mathbf{x}\geq\mathbf{0}\}$ *directly or with Lagrange multipliers, we can again use Corollary 20 to find that:*

$$g_K(p)=\left(\frac{|VW(K)|}{p}+\frac{|VB(K)|}{1-p}\right)^{-1}=\left(\frac{1}{p}+\frac{\lceil h/3\rceil-1}{1-p}\right)^{-1}=\frac{p(1-p)}{1+p(\lceil h/3\rceil-2)}.$$

If for some simple graph H a CRG K is in $\mathcal{K}(\mathrm{Forb}(H))$, then any graph which embeds in K will be in the hereditary property $\mathrm{Forb}(H)$. Thus, the amount of editing that must be done so that a graph embeds in such a K will be an upper bound for the edit distance of that graph from $\mathrm{Forb}(H)$. In other words, $g_K(p)$ serves as an upper bound for the edit distance function of H.

Definition 22. *[13] Let Γ be the clique spectrum of $\mathcal{H}$. Then the function $\gamma_{\mathcal{H}}(p)$ is defined as*

$$\gamma = \min\{g_{K(r,s)}(p) : (r,s) \in \Gamma\} = \min\{g_{K(r,s)}(p) : (r,s) \in \Gamma^*\}$$

Example 23. *We can construct the function $\gamma_{\mathrm{Forb}(C_h)}$ which is the minimum of the $g_K(p)$ functions computed in Example 21 for the CRGs from Proposition 16. As noted above, this function will serve as an upper bound for the edit distance function of $\mathrm{Forb}(C_h)$.*

$\gamma_{\mathrm{Forb}(C_h)}(p) = \min\left\{\frac{p}{2}, \frac{p(1-p)}{1+p(\lceil h/3\rceil - 2)}, \frac{1-p}{\lceil h/2\rceil - 1}\right\}$ *for h odd and*

$\gamma_{\mathrm{Forb}(C_h)}(p) = \min\left\{\frac{p(1-p)}{1+p(\lceil h/3\rceil - 2)}, \frac{1-p}{\lceil h/2\rceil - 1}\right\}$ *for h even.*

Since each g function corresponding to a CRG in $\mathcal{K}$ serves as an upper bound for the edit distance function, we have that

$ed_{\mathrm{Forb}(C_h)}(p) \leq \gamma_{\mathrm{Forb}(C_h)}(p)$.

2.3 p-core CRGs

There is a special class of CRGs defined by Marchant and Thomason in [10] called p-cores that limits the number of CRGs to be considered in computing the edit distance function. In order to define p-core CRGs, we must first discuss what it means for one CRG to be a sub-CRG of another.

Definition 24. *Given two CRGs K and K', we say that K' is a sub-CRG of K if K' can be obtained from K by vertex deletion.*

It follows from the definition that if K' is a sub-CRG of K and H embeds in K', then H embeds in K. The same embedding used for the sub-CRG works in the larger CRG, which contains the complete vertex set of the sub-CRG.

McKay and Martin observed that $g_{K'}(p) \geq g_K(p)$ if K' is a sub-CRG of K [12]. Intuitively, it is easier to optimally partition the vertices of a simple graph into a CRG for editing when vertices are added.

Definition 25. *A CRG K is a p-core CRG if for every sub-CRG K' of K, $g_K(p) < g_{K'}(p)$.*

Based on Definition 25 and the assertion following Equation 2.3 that the value of the edit distance function at each point corresponds to a particular CRG, we may conclude more strongly that the value of the $ed_{\mathcal{H}}(p)$ corresponds to a p-core CRG at each p.

Marchant and Thomason also recognized a structural requirement for p-core CRGs:

Lemma 26. *[10] Let K be a p-core CRG. Then K has all gray edges with the exception that*

- *if $p < 1/2$, two black vertices may be connected by a white edge, and*
- *if $p > 1/2$, two white vertices may be connected by a black edge.*

The following result concerning p-core CRGs from the work by Martin will be useful in the proofs in Chapter 4:

Lemma 27. *[13] Let $0 < p < 1/2$. Let K be a p-core CRG that has all black vertices and white or gray edges. Then*

i. If there is no gray 3-cycle in K, then $g_K(p) > p/2$.

ii. If there is no gray 4-cycle in K, then for $0 < p < 1/3$, $g_K(p) > p(1-p)$.

iii. If K contains a gray 3-cycle but does not contain 4 vertices that induce 5 gray edges, then $g_K(p) > \min\{2p/3, (1-p)/3\}$.

2.4 Symmetrization Techniques

Some useful results were obtained by Martin using symmetrization techniques [13].

In working with p-core CRGs, it often becomes helpful to consider a subgraph of a CRG that has only gray edges. The following definitions are helpful shorthand in these cases.

Definition 28. *The gray subgraph of a CRG K is the simple graph with vertex set $V(K)$ and edge set $EG(K)$. A cycle in the gray subgraph of K is a gray cycle in K, a path in the gray subgraph of K is a gray path in K, and a neighbor of a vertex in the gray subgraph of K is a gray neighbor of the same vertex in K.*

Definition 29. *The weighted gray degree of a vertex v in a CRG $V(K)$ is the sum of the weights of the vertices in $V(K)$ that are adjacent to v by a gray edge. The weighted gray degree of v is denoted $d_G(v)$.*

Lemma 30. *[13] Let $p \in (0, 1/2]$ and let K be a p-core CRG. Let the function $g_K(p)$ be defined by the quadratic program*

$$g_K(p) = \min\{\mathbf{x}^T \mathbf{M}_K(p)\mathbf{x} : \mathbf{x}^T \mathbf{1} = 1, \mathbf{x} \geq \mathbf{0}\}.$$

such that K has optimal vertex weight vector $\mathbf{x}$ and weighted adjacency matrix $\mathbf{M}_K(p)$. For any $v \in V(K)$, let $d_G(v)$ denote the sum of the weights of the neighbors that are adjacent to v by a gray edge. Then for any $v \in VB(K)$,

$$d_G(v) = \frac{p - g_K(p)}{p} + \frac{1-2p}{p}\mathbf{x}(v). \tag{2.4}$$

Using this lemma and the fact that $d_G(v) + \mathbf{x}(v) \leq 1$, we can also find an upper bound on the weights of individual vertices:

Lemma 31. *[13] Let K be a CRG that is p-core with optimal vertex weight vector $\mathbf{x}$. If $p \in (0, 1/2]$ and $v \in VB(K)$, then*

$$\mathbf{x}(v) \leq \frac{g_K(p)}{1-p}. \tag{2.5}$$

2.5 Previous results on $\mathrm{Forb}(C_h)$

In this section we summarize previously known results from the work of Martin in [13] on the edit distance from $\mathrm{Forb}(C_h)$.

Theorem 32. *[13] Let C_3 be a cycle on 3 vertices. Then*

$$ed_{\mathrm{Forb}(C_3)}(p) = \frac{p}{2}. \tag{2.6}$$

Theorem 33. *[13] Let C_h be a cycle on $h > 3$ vertices with h odd. Then,*

$$ed_{\mathrm{Forb}(C_h)}(p) \le \gamma_{\mathrm{Forb}(C_h)}(p) = \min\left\{\frac{p}{2}, \frac{p(1-p)}{1+p(\lceil h/3\rceil - 2)}, \frac{1-p}{\lceil h/2\rceil - 1}\right\} \tag{2.7}$$

and for $h < 10$, $ed_{\mathrm{Forb}(C_h)}(p) = \gamma_{\mathrm{Forb}(C_h)}(p)$.

Furthermore, if there is a $K \in \mathcal{K}(\mathrm{Forb}(C_h))$ such that $g_K(p) < \gamma_{\mathrm{Forb}(C_h)}(p)$, then $p < 1/2$ and K has all black vertices.

Theorem 34. *[13] Let C_h be a cycle on h vertices with h even. Then,*

$$ed_{\mathrm{Forb}(C_h)}(p) \le \gamma_{\mathrm{Forb}(C_h)}(p) = \min\left\{\frac{p(1-p)}{1+p(\lceil h/3\rceil - 2)}, \frac{1-p}{\lceil h/2\rceil - 1}\right\} \tag{2.8}$$

and for $h < 10$, $ed_{\mathrm{Forb}(C_h)}(p) = \gamma_{\mathrm{Forb}(C_h)}(p)$.

Furthermore, if there is a $K \in \mathcal{K}(\mathrm{Forb}(C_h))$ such that $g_K(p) < \gamma_{\mathrm{Forb}(C_h)}(p)$, then $p < 1/2$ and K has all black vertices.

The function $\gamma_{\mathrm{Forb}(C_h)}(p)$ was discussed in Example 23. It corresponds to the CRGs discussed in Example 21.

From these last two theorems we can see that if the edit distance function for $\mathrm{Forb}(C_h)$ is less than $\gamma_{\mathrm{Forb}(C_h)}$ for some $p \in (0, 1/2)$, then there must be a CRG $K \in \mathcal{K}(\mathrm{Forb}(C_h))$ with $g_K(p) < \gamma_{\mathrm{Forb}(C_h)}(p)$ such that K is p-core and has all black vertices.

One further result from [13] is helpful in verifying the edit distance function:

Lemma 35. *[13] Let C_h be a cycle on h vertices with $h \ge 4$. If there is a p-core CRG $K \in \mathcal{K}(\mathrm{Forb}(C_h))$ such that $g_K(p) < \gamma_{\mathrm{Forb}(C_h)}(p)$ and if $p < 1/2$, then K has no gray cycles with length in $\{\lceil h/2\rceil, ..., h\}$.*

The edit distance function for $\mathrm{Forb}(C_h)$ was determined for $h \in \{3,4,5,6,7,8,9\}$. The edit distance function for $\mathrm{Forb}(C_{10})$ was also determined for all but small values of p, and the maximum value for this principle hereditary property was also obtained.

Theorem 36. *[13] Let $\mathcal{H} = \mathrm{Forb}(C_h)$.*

i. Let $h = 3$. Then $ed_{\mathcal{H}}(p) = p/2$, $(p^*, d^*) = (1, 1/2)$.

ii. Let $h = 4$. Then $ed_{\mathcal{H}}(p) = p(1-p)$, $(p^*, d^*) = (1/2, 1/4)$.

iii. *Let* $h = 5$. *Then* $ed_{\mathcal{H}}(p) = \min\left\{\frac{p}{2}, \frac{1-p}{2}\right\}$, $(p^*, d^*) = (1/2, 1/4)$.

iv. *Let* $h = 6$. *Then* $ed_{\mathcal{H}}(p) = \min\left\{p(1-p), \frac{1-p}{2}\right\}$, $(p^*, d^*) = (1/2, 1/4)$.

v. *Let* $h = 7$. *Then* $ed_{\mathcal{H}}(p) = \min\left\{\frac{p}{2}, \frac{p(1-p)}{1+p}, \frac{1-p}{3}\right\}$, $(p^*, d^*) = (\sqrt{2}-1, 3-2\sqrt{2})$.

vi. *Let* $h = 8$. *Then* $ed_{\mathcal{H}}(p) = \min\left\{\frac{p(1-p)}{1+p}, \frac{1-p}{3}\right\}$, $(p^*, d^*) = (\sqrt{2}-1, 3-2\sqrt{2})$.

vii. *Let* $h = 9$. *Then* $ed_{\mathcal{H}}(p) = \min\left\{\frac{p}{2}, \frac{1-p}{4}\right\}$, $(p^*, d^*) = (1/3, 1/6)$.

viii. *Let* $h = 10$ *and let* $p \in [1/7, 1]$. *Then* $ed_{\mathcal{H}}(p) = \min\left\{\frac{p(1-p)}{1+2p}, \frac{1-p}{4}\right\}$, $(p^*, d^*) = ((\sqrt{3}-1)/2, (2-\sqrt{3})/2)$.

Theorem 7 comes from verifying that $\gamma_{\mathrm{Forb}(C_h)}(p) = ed_{\mathrm{Forb}(C_h)}(p)$ for all values of $h > 3$, the proof of which makes up Chapter 3.

2.6 Miscellaneous Results about Edit Distance

The following observations about the edit distance function are useful for the proofs in Chapter 4.

Lemma 37. *[13] Let* H *be a simple graph with complement* $\overline{H}$. *Then* $ed_{\mathrm{Forb}(H)}(p) = ed_{\mathrm{Forb}(\overline{H})}(1-p)$.

Lemma 38. *[4] The edit distance function must be continuous and concave down.*

Lemma 39. *[1] For any hereditary property* $\mathcal{H}$, $ed_{\mathcal{H}}(1/2) = \gamma_{\mathcal{H}}(1/2)$

2.7 Previous Results for other Hereditary Properties

This section summarizes some results known for the edit distance from hereditary properties other than $\mathrm{Forb}(C_h)$.

Balogh and Martin used the concept of colored regularity graphs (CRGs, discussed in depth in Section 2.1) to obtain exact values of the edit distance function for several types of graphs.

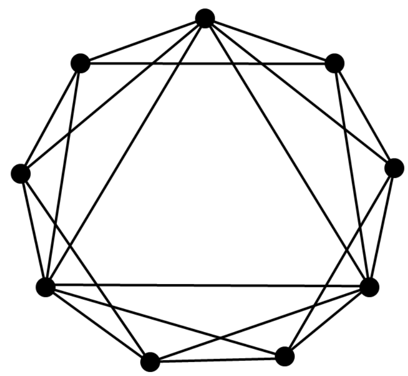

Figure 2.2 The graph H_9.

Theorem 40. *[4] Let $H = K_a + E_b$ be the disjoint union of an a-clique and an independent set of size b with $a \geq 2$, $b \geq 1$. Then for* $\mathrm{Forb}(K_a + E_b)$,

$$(p^*, d^*) = \left(\frac{a-1}{a+b-1}, \frac{1}{a+b-1}\right).$$

The maximum edit distance for this set of graphs is significant because neither the upper nor the lower bounds found in [3] using the binary chromatic number hold.

Theorem 41. *[4] For the hereditary property* $\mathrm{Forb}(K_{3,3})$, $(p^*, d^*) = (\sqrt{2}-1, 3-2\sqrt{2})$.

Theorem 42. *[4] Let H_9 be the graph shown in Figure 2.2. Then for* $\mathrm{Forb}(H_9)$, $d^* \leq \dfrac{3-\sqrt{5}}{4}$.

Theorem 43. *[4] For* $\mathrm{Forb}(\overline{P_3 + K_1})$, $(p^*, d^*) = (2/3, 1/3)$.

In [13], Martin introduced the symmetrization techniques which were discussed in the Section 2.4. Using these, he was able to find bounds on the edit distance function for a hereditary property which forbids complete graphs.

Definition 44. *For a hereditary property $\mathcal{H}$, $\mathcal{F}(\mathcal{H})$ is the set of graphs for which*

$$\mathcal{H} = \bigcap_{H \in \mathcal{F}(\mathcal{H})} \mathrm{Forb}(H)$$

and $\mathcal{F}(\mathcal{H})$ is minimal with respect to vertex deletion.

Definition 45. *The chromatic number $\chi(\mathcal{H})$ of a hereditary property $\mathcal{H}$ is*

$$\chi(\mathcal{H}) = \min\{\chi(H) : H \in \mathcal{F}(\mathcal{H})\}.$$

Theorem 46. *[13] Let $\mathcal{H}$ be a hereditary property such that $\mathcal{F}(\mathcal{H})$ contains a complete graph and let h be the smallest positive integer such that $K_h \in \mathcal{F}(\mathcal{H})$. Let $\chi = \chi(\mathcal{H})$ and let m be the smallest positive integer such that a complete multipartite graph with m parts is a member of $\mathcal{F}(\mathcal{H})$. Then*

$$\min\left\{\frac{p}{\chi-1}, \frac{1-p}{\chi-1}, \frac{2p-1}{m-1}\right\} \leq ed_{\mathcal{H}}(p) \leq \min\left\{\frac{p}{\chi-1}, 1-p+\frac{2p-1}{m-1}\right\},$$

and $ed_{\mathcal{H}}(p) = \dfrac{p}{\chi-1}$ *if* $\mathcal{H} = \mathrm{Forb}(K_h)$.

Using colored regularity graphs and the symmetrization tools developed in [13], Martin and McKay [12] found the exact edit distance function for $\mathrm{Forb}(K_{2,3})$ and $\mathrm{Forb}(K_{2,4})$. They also found the values of the edit distance function for $\mathrm{Forb}(K_{2,t})$ for large values of p and found many bounds on the function.

Theorem 47. *[12] Let $\mathcal{H} = \mathrm{Forb}(K_{2,t})$. Then*

i. Let $t = 3$. Then $ed_{\mathcal{H}}(p) = \min\left\{p(1-p), \dfrac{1-p}{2}\right\}$. *Then* $(p^*, d^*) = (1/2, 1/4)$.

ii. Let $t = 4$. Then $ed_{\mathcal{H}}(p) = \min\left\{p(1-p), \dfrac{7p+1}{15}, \dfrac{1-p}{3}\right\}$. *Then* $(p^*, d^*) = (1/3, 2/9)$.

iii. Let $t \geq 4$. Then if $p \geq \dfrac{2}{t+1}$, $ed_{\mathcal{H}}(p) = \dfrac{1-p}{t-1}$.

Several other bounds on the function $ed_{\mathrm{Forb}(K_{2,t})}(p)$ were obtained using colored regularity graphs, Zarenkiewicz constructions, and strongly regular graph constructions. Better upper bounds were determined for $t = 5, 6, 7, 8$.

In [14], Martin determined the edit distance function for split graphs.

Definition 48. *A split graph is a graph whose vertices can be partitioned into one independent set and one clique.*

Definition 49. *The independence number of a graph G is the size of the largest independent set in G. The clique number of a graph G is the size of the largest clique in G.*

Theorem 50. *[14] Let H be a split graph with clique number ω and independence number α. Let H not be a complete graph or an empty graph. Then*

$$ed_{\mathrm{Forb}(H)}(p) = \min\left\{\frac{p}{\omega-1}, \frac{1-p}{\alpha-1}\right\}.$$

Thus, $(p^*, d^*) = \left(\frac{\omega - 1}{\alpha + \omega - 2}, \frac{1}{\alpha + \omega - 2}\right)$.

Martin also completed the work with the edit distance function for H_9, determining the function completely:

Theorem 51. *[14] For the graph H_9 (Figure 2.2),*

$$ed_{\mathrm{Forb}(H_9)}(p) = \min\left\{\frac{p}{3}, \frac{p}{1+4p}, \frac{1-p}{2}\right\}.$$

Thus, $(p^*, d^*) = \left(\frac{1+\sqrt{17}}{8}, \frac{7-\sqrt{17}}{16}\right)$.

CHAPTER 3. EDIT DISTANCE FROM $\mathrm{Forb}(C_h)$

Before proceeding to prove the main theorem, we will define a type of CRG with characteristics that summarize several key assumptions that come from the previous work by Martin in [13].

Definition 52. *For specified values of $h > 3$ and $0 \le p \le 1$, a CRG K is a candidate CRG if the following are true:*

i. K has all black vertices. This comes from Theorems 33 and 34.

ii. K is a p-core CRG with $p < 1/2$. This also comes from Theorems 33 and 34.

iii. K is not isomorphic to $K(2,0)$, $K(0, \lceil h/2 \rceil - 1)$, or $K(1, \lceil h/3 \rceil - 1)$ (these are the CRGs which generate the function $\gamma_{\mathrm{Forb}(C_h)}$ defined in Example 23).

iv. K has no gray cycles with length in $\{\lceil h/2 \rceil, ..., h\}$. This is the result given in Lemma 35.

v. $g_K(p) < \gamma_{\mathrm{Forb}(C_h)}(p)$.

If no such candidate CRG exists in $\mathcal{K}(\mathrm{Forb}(C_h))$ at a given value of p, then $ed_{\mathrm{Forb}(C_h)}(p) = \gamma_{\mathrm{Forb}(C_h)}(p)$.

The proof proceeds as follows: first, we will establish some characteristics of a candidate CRG K. Specifically, we will focus on the length of the longest gray path in K. Then, in Section 3.2, we will narrow the possibilities for K down to a certain type of graph. Then we will obtain a contradiction to such a graph existing in $\mathcal{K}(\mathrm{Forb}(C_h))$ for our desired range of p values, establishing the upper bound given by $\gamma_{\mathrm{Forb}(C_h)}$ as the value of the edit distance function over the range of p given in Theorem 7. Finally, we will proceed to discuss the maximum point of the edit distance function.

3.1 Preliminary Characteristics of a Candidate CRG

We now proceed to establish several characteristics of a candidate CRG K, culminating with the length of its longest gray path. Since much of the proof concerns the gray subgraph of K, it may be helpful to refer to the terminology in Definition 28.

Proposition 53. *Let $0 \le p \le 1$ for h odd and $\frac{1}{\lceil h/3 \rceil} \le p \le 1$ for h even. Let K be a candidate CRG. For all $v, w \in V(K)$, either vw is a gray edge of K or there is a vertex $u \in V(K)$ such that uv and uw are gray edges of K (i.e., the gray subgraph of K has diameter at most 2 and hence is connected).*

Proof. <u>Case 1</u>: h is odd

Since $p < 1/2$ and $g \le \gamma_{\text{Forb}(C_h)} \le p/2$, we may use Lemma 30 to obtain:

$$\begin{aligned} d_G(v) &> \frac{p - p/2}{p} + \frac{1-2p}{p} x(v) \\ &> 1/2. \end{aligned}$$

<u>Case 2</u>: h is even

In this case, we have that $\dfrac{1}{\lceil h/3 \rceil} \le p \le 1/2$ and $g \le \gamma_{\text{Forb}(C_h)} \le \dfrac{p(1-p)}{1 + p(\lceil h/3 \rceil - 2)}$. We again use Lemma 30 and these bounds on p and g to obtain:

$$\begin{aligned} d_G(v) &> \frac{p - \frac{p(1-p)}{1+p(\lceil h/3 \rceil - 2)}}{p} + \frac{1-2p}{p} x(v) \\ &\ge 1 - \frac{1-p}{1 + p(\lceil h/3 \rceil - 2)} \\ &> 1 - \frac{1 - 1/\lceil h/3 \rceil}{1 + \lceil h/3 \rceil(\lceil h/3 \rceil - 2)} \\ &> 1/2. \end{aligned}$$

Therefore, for any $v, w \in V(K)$, either vw is a gray edge, or else v and w share at least one common gray neighbor. □

Proposition 54. *Let K be a candidate CRG. Then $|V(K)| \ge \lceil h/2 \rceil$.*

Proof. The CRG with the smallest g function on $h/2$ or fewer black vertices is the one with all gray edges, i.e., $K(0, \lceil h/2 \rceil - 1)$. But that CRG is the one that generates the third g function above, i.e., $g_{K(0,\lceil h/2 \rceil - 1)} \ge \gamma_{\text{Forb}(C_h)}$. Thus, K must have at least $\lceil h/2 \rceil$ vertices. □

Definition 55. *Let K be a CRG with vertex set $V(K)$. Then the unweighted gray degree of a vertex v in $V(K)$ is the number of vertices adjacent to v by a gray edge. The unweighted minimum gray degree is denoted by δ.*

It is helpful to note the subtle but important difference between the unweighted gray degree, which counts vertices adjacent by gray edges, and the weighted gray degree $d_G(v)$ (given in Definition 29) which sums the weights of vertices adjacent by gray edges. Both of these definitions are used in the following claim.

Proposition 56. *Let K be a candidate CRG. Let δ be the unweighted minimum gray degree of vertices in K. Then $\delta \geq \lceil h/3 \rceil$.*

Proof. The minimum unweighted gray degree must be at least the minimum weighted gray degree of a vertex in $V(K)$ divided by the largest weight of any vertex in $V(K)$. So by Lemma 30 and Lemma 31, we have that

$$\begin{aligned}
\delta &\geq \left\lceil \frac{\min_v\{d_G(v)\}}{\max_v\{\mathbf{x}(v)\}} \right\rceil \\
&\geq \left\lceil \frac{\frac{p-g}{p} + \frac{1-2p}{p}x(v)}{\frac{g}{1-p}} \right\rceil \\
&> \frac{(p-g)(1-p)}{pg} \\
&= \frac{1-p}{g} - \frac{1}{p} + 1.
\end{aligned}$$

We know that $g \leq \dfrac{p(1-p)}{1+p(\lceil h/3 \rceil - 2)}$, so

$$\begin{aligned}
\delta &> (1-p) \div \frac{p(1-p)}{1+p(\lceil h/3 \rceil - 2)} - \frac{1}{p} + 1 \\
&= \lceil h/3 \rceil - 1.
\end{aligned}$$

Therefore, $\delta \geq \lceil h/3 \rceil$. □

Lemma 57. *Let l be the length of the longest gray path in a candidate CRG K. Then $2\lceil h/3 \rceil - 1 \leq l \leq h - 2$.*

Proof. We will first show that $l \leq h-2$.

Suppose K has a gray path of length $h-1$ composed of vertices $v_1, v_2, ..., v_{h-1}$. The edge $v_1 v_{h-1}$ cannot be gray, otherwise a gray cycle of length $h-1$ would be formed in contradiction to Lemma 35 since $h-1 \geq \lceil h/2 \rceil$. Thus, v_1 and v_{h-1} must have a common gray neighbor by Proposition 53. This common gray neighbor cannot be off the path, because then K would have a gray cycle of length h, a contradiction to Lemma 35.

Thus, there is some integer $i \in \{2, ..., h-2\}$ such that $v_1 v_i$ is gray and $v_i v_{h-1}$ is gray.

Suppose $i \leq \frac{h-1}{2}$. Then the vertices $v_i, v_{i+1}, ..., v_{h-1}$ form a gray cycle of length

$$h-1-(i-1) = h-i \geq h - \frac{h-1}{2} \geq \lceil h/2 \rceil.$$

Suppose $i \geq \frac{h+1}{2}$. Then the vertices $v_1, v_2, ..., v_i$ form a cycle of length

$$i \geq \frac{h+1}{2} \geq \lceil h/2 \rceil.$$

If h is even, suppose $i = h/2$. Then the vertices $v_1, v_2, ..., v_i$ form a cycle of length

$$i = h/2 = \lceil h/2 \rceil.$$

Hence, there is a cycle with length in $\{\lceil h/2 \rceil, ..., h\}$, a contradiction to Lemma 35.

We will now proceed to show that $l \geq 2\lceil h/3 \rceil + 1$.

Let $v_1, ..., v_l$ be a longest gray path in K. Since this is a longest path, the endpoints v_1 and v_l cannot have gray neighbors off the path.

Assume by contradiction that v_1 and v_l are adjacent. Then the path $v_1, ..., v_l$ would form a cycle of length l. If there are vertices off the path, at least one of them must be adjacent to a vertex in the path since the gray subgraph of K is connected by Proposition 53. However, since it is adjacent to a cycle of length l, this would form a path of length $l+1$, a contradiction to $v_1, ..., v_l$ being the longest path. On the other hand, suppose $V(K) = \{v_1, ..., v_l\}$, so $|V(K)| = l$. By Proposition 54, $l \geq \lceil h/2 \rceil$. We know that $l \leq h-2$ from the first half of the proof. Then there is a cycle in $\{\lceil h/2 \rceil, ..., h\}$, a contradiction to Lemma 35. Thus, v_1 and v_l are not adjacent.

Let $N_G(v)$ denote the set of gray neighbors of a vertex v. (Note that $|N_G(v)| \neq d_G(v)$, since $|N_G(v)|$ denotes the unweighted gray degree of v and $d_G(v)$ denotes the weighted gray degree of v.)

Let $A = N_G(v_1)$, the set of gray neighbors of v_1. By Proposition 56, $|A| \geq \lceil h/3 \rceil$.

Let $B = \{v_{i+1} : v_i \in N_G(v_l)\}$, the set of successors along the path of neighbors of v_l. By Proposition 56, $|B| \geq \lceil h/3 \rceil$.

We claim that $A \cap B$ is empty. Suppose it is not empty, and $v_y \in A \cap B$. Then there is a gray cycle of length l on $v_1, v_2, ..., v_{y-1}, v_l, v_{l-1}, ...v_y, v_1$. But this implies that the starting and ending vertices of a longest path in K are adjacent, a contradiction to Proposition 53 discussed above. So $|A \cup B| = |A| + |B|$.

Observe also that $A \cup B$ is a subset of $\{v_2, ..., v_l\}$, so

$$\begin{aligned}
l - 1 &\geq |A \cup B| \\
l - 1 &\geq |A| + |B| \\
l - 1 &\geq 2\lceil h/3 \rceil \\
l &\geq 2\lceil h/3 \rceil + 1.
\end{aligned}$$

□

3.2 Proof of Theorem 7

The proof of the main theorem now proceeds to examine the structure of the gray subgraph of a candidate CRG K.

Let $h > 3$ and let $0 \leq p \leq 1$ for h odd and $\frac{1}{\lceil h/3 \rceil} \leq p \leq 1$ for h even. Assume by contradiction that there is some CRG $K \in \mathcal{K}(\text{Forb}(C_h))$ that is a candidate CRG.

Proposition 58. *Let $v_1, ..., v_l$ be a maximal gray path in K. Then this path is structured such that the starting and ending vertices, v_1 and v_l have exactly one common gray neighbor, v_c. Additionally, all of the other gray neighbors of v_1 are on the path before v_c, and all of the other gray neighbors of v_l are on the path after v_c.*

Proof. Let the path (minus the endpoints v_1 and v_l) be partitioned into sets $A_1, ..., A_{k+1}$ and $G_1, ..., G_k$ as follows (see Figure 3.1): Each set A_i starts with a gray neighbor of v_1, call it v_p. There is a gray neighbor of v_l, call it v_q, with $q \geq p$ but q as small as possible. Then let v_r be the gray neighbor of v_l with the highest index before another gray neighbor of v_1 (note that it

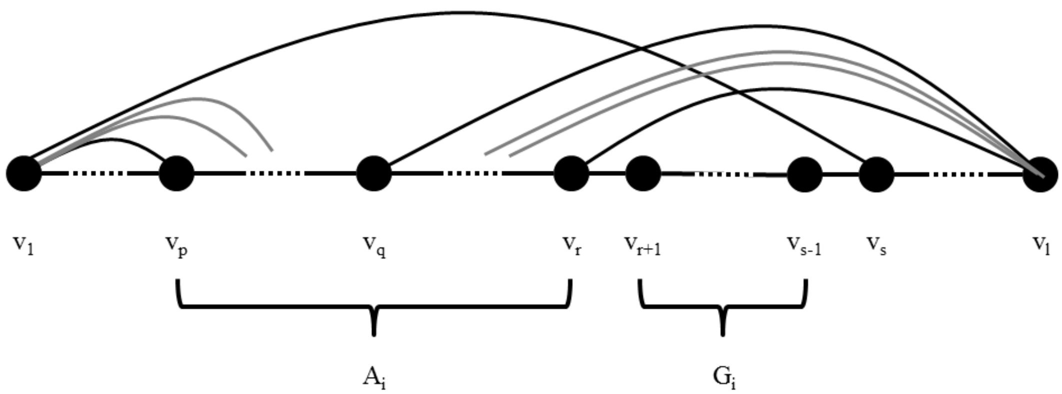

Figure 3.1 The partition of the longest gray path in K into sets A_i and G_i.

is possible that $q = r$). Then $A_i = \{v_p, ..., v_r\}$. Let v_s be the gray neighbor of v_1 with $s > r$ but s as small as possible. Then $G_i = \{v_{r+1}, ..., v_{s-1}\}$, with v_s as the first vertex in A_{i+1} (note that G_i may be empty if $r + 1 = s$). With this set up, v_1 may have other neighbors in $\{v_p, ..., v_q\}$ and v_l may have other neighbors in $\{v_q, ..., v_r\}$, but G_i does not contain any neighbors of v_1 or v_l.

This partition is constructed so that all of the neighbors of v_1 and v_l are contained in the A_i sets. The G_i sets are "gap" sets - no vertex in a G_i is adjacent to either end vertex. We wish to show that there is no gap set G_i (that is, $k = 0$). This result arises because K cannot have a very long cycle as required by Lemma 35. The absence of a gap set is important since it implies that the end vertices v_1 and v_l have exactly one common neighbor.

Since this is a partition of the interior vertices of the path,

$$\sum_{i=1}^{k+1} |A_i| + \sum_{i=1}^{k} |G_i| = l - 2.$$

Let v_r be the last vertex in A_i and v_s be the first vertex in A_{i+1}. Then $v_1v_s, v_rv_l \in EG(K)$

since by definition the first vertex in any A_i is a neighbor of v_1 and the last vertex in any A_i is a neighbor of v_l. Since G_i contains the vertices between A_i and A_{i+1}, $G_i = \{v_{r+1}, ..., v_{s-1}\}$. Thus, $|G_i| = s - r - 1$.

There is a cycle on the vertices $v_1, v_2, ..., v_r, v_l, v_{l-1}, ..., v_s, v_1$. This cycle has length $l - |G_i|$ and must be shorter than $\lceil h/2 \rceil$ by Lemma 35 and Lemma 57. Thus, for any $1 \leq i \leq k$,

$$\begin{aligned} l - |G_i| &\leq \lceil h/2 \rceil - 1 \\ l - \lceil h/2 \rceil + 1 &\leq |G_i|. \end{aligned}$$

So for any k, we have that

$$k(l - \lceil h/2 \rceil + 1) \leq \sum_{i=1}^{k} |G_i|. \tag{3.1}$$

Observe that v_1 and v_l can have at most 1 common neighbor in any given A_i. Thus,

$$\begin{aligned} 2\delta &\leq \sum_{i=1}^{k+1} |A_i| + (k+1) \\ 2\delta - (k+1) &\leq \sum_{i=1}^{k+1} |A_i| \\ 2\delta - (k+1) + k(l - \lceil h/2 \rceil + 1) &\leq \sum_{i=1}^{k+1} |A_i| + \sum_{i=1}^{k} |G_i| = l - 2 \\ 2\delta + k(l - \lceil h/2 \rceil) + 1 &\leq l \\ 2\lceil h/3 \rceil + k(l - \lceil h/2 \rceil) + 1 &\leq l, \end{aligned}$$

where the last inequality is obtained by using the minimum degree condition in Claim 56.

Assume by contradiction that $k \geq 1$. Then since $l \geq 2\lceil h/3 \rceil + 1$ by Claim 57, $l - \lceil h/2 \rceil \geq 0$, so we have

$$\begin{aligned} 2\lceil h/3 \rceil + l - \lceil h/2 \rceil + 1 &\leq l \\ 2\lceil h/3 \rceil &\leq \lceil h/2 \rceil - 1, \end{aligned}$$

which is a contradiction for all h. Therefore, it must be that $k = 0$.

Then because there are no gaps, we know that the furthest gray neighbor of v_1 along the path is v_c, then the furthest gray neighbor along the path from v_l is at furthest v_c. Since v_1 and

v_l must have a common neighbor and it must be on the path to avoid creating a longer path, v_c must be that common neighbor. Note that since $k = 0$, v_c is the only common neighbor of v_1 and v_l. □

Definition 59. *Let $v_1, ..., v_l$ be a gray path. Then if some vertex w in the path is adjacent to v_i by a gray edge, v_{i-1} is a predecessor of a neighbor of w and v_{i+1} is a successor of a neighbor of w. We denote $N_G(w)$ as the gray neighborhood of w, and $N'_G(w)$ as the set of predecessors of neighbors of w (i.e., $N'_G(w) = \{w_i : w_{i+1} \in N_G(w)\}$).*

Proposition 60. *Let $P = \{v_1, ..., v_l\}$ be a longest gray path in K, where v_c is the unique common neighbor of v_1 and v_l specified by Proposition 58. Then the gray subgraph of the sub-CRG of K induced by P is composed of two gray cliques $\{v_1, ..., v_c\}$ and $\{v_c, ..., v_l\}$ with a common vertex v_c.*

Proof. Let $\{v_1, ..., v_l\}$ be a longest gray path in K. By Proposition 58, v_1 and v_l have exactly one common neighbor, call it v_c. This common neighbor is the last (largest index) neighbor of v_1 on the path and the first (smallest index) neighbor of v_l on the path.

But let v_p be a predecessor of a neighbor of v_1. Then v_p is an equivalent endpoint to the longest path containing the same l vertices. But by Proposition 58, v_p must have a common neighbor with v_l, and this common neighbor still must be the first neighbor of v_l on the path, that is, v_c. By the same argument, all of the successors of neighbors of v_l must also have v_c as their common neighbor with v_1.

But what about the other vertices? Because of the requirement in Lemma 35 that any cycle be strictly shorter than $\lceil h/2 \rceil$, there are at most $\lceil h/2 \rceil - 2$ vertices before v_c, and there are at most $\lceil h/2 \rceil - 2$ vertices after v_c. So there may be vertices that are not predecessors of neighbors of v_1 or successors of neighbors of v_l - that is, vertices which cannot be an endpoint of a longest path composed of the vertices $v_1, ..., v_l$.

Let $v_x \in \{v_1, ..., v_c\}$ but $v_x \notin N_G(v_1)$. But then $v_{x-1} \notin N'_G(v_1)$, i.e., v_{x-1} is one of the vertices that is not a predecessor of a neighbor of v_1.

Let $v_n \in N_G(v_1)$, with $n \leq x - 1$, so $v_{n-1} \in N'_G(v_1)$. Then v_{n-1} cannot be adjacent to v_x. If it were, then there would be a longest path starting with v_{x-1}:

$v_{x-1}, v_{x-2}, ..., v_n, v_1, v_2, ..., v_{n-1}, v_x, v_{x+1}, ..., v_c, ..., v_l$.

But this implies that $N_G(v_x) \cap N'_G(v_1)$ is empty. Thus, v_x does not have enough neighbors in the first part of the path to satisfy its minimum degree requirement. Therefore, it cannot be a starting vertex of a longest path, i.e., $v_x \notin N'_G(v_1)$. But that means that $v_{x+1} \notin N_G(v_1)$. So every vertex following a nonneighbor of v_1 must also be a nonneighbor of v_1.

Let $v_n \in N_G(v_1)$, with $n \geq x+1$, so $v_{n-1} \in N'_G(v_1)$. Then v_{n-1} cannot be adjacent to v_{x-2}. If it were, then there would be a longest path starting with v_{x-1}:

$v_{x-1}, v_x, ..., v_{n-1}, v_{x-2}, v_{x-3}, ..., v_1, v_n, v_{n+1}, ..., v_c, ..., v_l$.

But this implies that $N_G(v_{x-2}) \cap N'_G(v_1)$ is empty. Thus, v_{x-2} does not have enough neighbors in the first part of the path to satisfy its minimum degree requirement. Therefore, it cannot be a starting vertex of a longest path, i.e., $v_{x-2} \notin N'_G(v_1)$. But that means that $v_{x-1} \notin N_G(v_1)$. By induction, every vertex preceding a nonneighbor of v_1 must also be a nonneighbor of v_1.

Clearly these two parts exclude all of the possible neighbors of v_1 from its neighborhood, a contradition. Thus, $N'_G(v_1) = \{v_1, ..., v_{c-1}\}$. A parallel argument shows that $N'_G(v_l) = \{v_{c+1}, ..., v_l\}$. Thus, every vertex except v_c in the longest path is a starting vertex of the longest path, and therefore cannot be adjacent to anything off that path. From above, v_c is the only common vertex between the vertices in the first part of the path and the vertices in the second part of the path.

Additionally, since $N'_G(v_1) = \{v_1, ..., v_{c-1}\}$, we have that $N_G(v_1) = \{v_2, ..., v_c\}$. Since every vertex in $\{v_1, ..., v_l\}$ except v_c could equivalently be considered the first vertex in the path, the neighborhood of every vertex in the longest path is composed of all of the vertices between it and v_c, including v_c. □

Definition 61. *Let $v_1, ..., v_l$ be a longest gray path in a candidate CRG K. Then the central vertex, v_c, is the unique common neighbor of v_1 and v_1.*

Proposition 62. *Let K be a candidate CRG with a longest gray path $v_1, ..., v_l$ with central vertex v_c. Let w be a vertex in $V(K)$ that is not in $\{v_1, ...v_l\}$. Let $w_1, ..., w_m$ be the longest gray path containing w. Let $N_G(w)$ be the gray neighborhood of w. Then* *i.* $v_c \in N_G(w)$

ii. $v_i \notin N_G(w)$ *for* $i \neq c$

iii. xy *is a gray edge in* K *for all* $x, y \in N_G(w)$

iv. xu *is a white edge in* K *for all* $x \in N_G(w)$ *and* $u \notin N_G(w) \cup w$

Proof. Let K be a candidate CRG with a longest gray path $v_1, ..., v_l$ with central vertex v_c. Let w be a vertex in $V(K)$ that is not in $\{v_1, ...v_l\}$. Let $w_1, ..., w_m$ be the longest gray path containing w. Let $N_G(w)$ be the gray neighborhood of w.

By Proposition 53, w must have a common gray neighbor with every vertex in $\{v_1, ..., v_l\}$. But since every vertex in the longest path except v_c is an equivalent start to the longest path by Proposition 60, w cannot be adjacent to any vertex in the longest path except v_c (otherwise, $v_1, ..., v_l, w$ would be a gray path of length $l+1$). Thus we have that $v_c \in N_G(w)$ and $v_i \notin N_G(w)$ for $i \neq c$, which proves i and ii.

Apply the proof of Proposition 60 to the path $w_1, ..., w_m$, the longest path containing w. The same arguments used in that proof hold to show that $N_G(w) \cup w$ is a gray clique that contains v_c. This proves iii and iv. □

So K is composed of gray cliques which are otherwise disjoint but share a common central vertex. The cliques must be of size at least $\lceil h/3 \rceil$ because of the minimum degree requirement in Proposition 56. The cliques can be at most size $\lceil h/2 \rceil - 1$ because of the maximum path length given in Lemma 57. We will refer to these gray cliques as "petals". The two largest petals of K form the longest path in K.

Definition 63. *Let* K *be a candidate CRG with central vertex* v_c *of its longest path. Let* $v \neq v_c$ *be a vertex in* $V(K)$ *with gray neighborhood* $N_G(v)$*. Then* $N_G(v)$ *is a petal of* K*.*

Proposition 64. *Let* K *be a candidate CRG with a longest path* $v_1, ..., v_l$*. Then* $V(K) = \{v_1, ..., v_l\}$*. (That is,* K *is composed of exactly two petals, and they form the longest path in* K*).*

Proof. First, observe that since $l \geq 2\lceil h/3 \rceil + 1$, K cannot have only one petal. If it did, then K would be a clique with $2\lceil h/3 \rceil + 1 > \lceil h/2 \rceil$ vertices, which is a contradiction to Lemma 35.

Now assume that there are at least 3 petals in K. Let x_c be the weight of the central vertex, and let X_1, X_2, X_3 be the total weights (not counting the weight of the central vertex) of all the vertices in the three petals of largest weight. Then if K has at least 3 petals,

$$X_1 + X_2 + X_3 + x_c \leq 1 \tag{3.2}$$

We may also observe from Proposition 53 that since $d_G(v) > 1/2$ for each vertex in K, for the i^{th} petal,

$$X_i + x_c > 1/2 \tag{3.3}$$

This is due to the fact that all of the neighbors of any member of the i^{th} petal must be the central vertex plus other vertices from the same petal.

If we combine inequalities in 3.3 generated for each petal, we have that

$$X_1 + X_2 + X_3 + 3x_c > 1.5. \tag{3.4}$$

By combining 3.4 with the inequality 3.2, we have:

$$\begin{aligned} 2x_c &> 0.5 \\ x_c &> 0.25. \end{aligned}$$

But, we also know that the weight of any single vertex in K is bounded above by $g_K(p)/(1-p)$, so we have:

$$\begin{aligned} \frac{g_K(p)}{1-p} &\geq x_c > 0.25 \\ \frac{g_K(p)}{1-p} &> 0.25. \end{aligned}$$

From Theorem 33 and Theorem 34, we have that $g_K(p) \leq \dfrac{1-p}{\lceil h/2 \rceil - 1}$. Therefore,

$$\begin{aligned} \frac{1}{\lceil h/2 \rceil - 1} &> 0.25 \\ \lceil h/2 \rceil &< 5 \\ h &\leq 8. \end{aligned}$$

Thus we have a contradiction for $h \geq 9$. (Since we already know the edit distance function for small h from Theorems 33 and 34, this is sufficient). □

Proposition 65. *For* $p \geq \frac{1}{\lceil h/2\rceil - 1}$, $ed_{\mathcal{H}}(p) = \gamma_{\mathrm{Forb}(C_h)}(p)$.

Proof. We know from Proposition 64 that K must be composed of two gray cliques that share exactly one vertex, v_c. Let X_1 and X_2 be the weights of each of the cliques (excluding the central vertex), and let x_c be the weight of v_c. Then

$$\begin{aligned} d_G(v_c) + x_c &= 1 \\ \frac{p-g}{p} + \frac{1-p}{p}x_c &= 1 \\ x_c &= \frac{g}{1-p}. \end{aligned}$$

Let v_1 be a vertex in the first clique with weight x_1. Then since there are at most $\lceil h/2\rceil - 2$ vertices in the first clique (excluding v_c),

$$x_1 \geq \frac{X_1}{\lceil h/2\rceil - 2}.$$

We also observe as we did for the central vertex that

$$d_G(v_1) + x_1 = X_1 + x_c.$$

We then make use of Lemma 30 and the value of x_c determined above to obtain

$$\begin{aligned} \frac{p-g}{p} + \frac{1-p}{p}x_1 &= X_1 + x_c \\ \frac{p-g}{p} + \frac{1-p}{p}\frac{X_1}{\lceil h/2\rceil - 2} &\leq X_1 + \frac{g}{1-p} \\ X_1\left[\frac{1-p}{p}\frac{1}{\lceil h/2\rceil - 2} - 1\right] &\leq \frac{g}{1-p} - 1 + \frac{g}{p} \\ X_1\left[\frac{1-p}{p}\frac{1}{\lceil h/2\rceil - 2} - 1\right] &\leq \frac{g}{p(1-p)} - 1. \end{aligned}$$

We obtain an analogous inequality for the second clique by replacing X_1 with X_2 in the final inequality. By combining these two inequalities, we have

$$(X_1 + X_2)\left[\frac{1-p}{p}\frac{1}{\lceil h/2\rceil - 2} - 1\right] \leq 2\frac{g}{p(1-p)} - 2.$$

Observe that $X_1 + X_2 = 1 - x_c = 1 - \dfrac{g}{1-p}$. Then the above inequality becomes

$$\begin{aligned}
\left(1 - \frac{g}{1-p}\right)\left[\frac{1-p}{p}\frac{1}{\lceil h/2\rceil - 2} - 1\right] &\leq 2\frac{g}{p(1-p)} - 2 \\
\frac{1-p}{p}\frac{1}{\lceil h/2\rceil - 2} - 1 - \frac{g}{p}\frac{1}{\lceil h/2\rceil - 2} + \frac{g}{1-p} &\leq 2\frac{g}{p(1-p)} - 2 \\
\frac{1-p}{p}\frac{1}{\lceil h/2\rceil - 2} + 1 &\leq g\left(\frac{2}{p(1-p)} - \frac{1}{1-p} + \frac{1}{p(\lceil h/2\rceil - 2)}\right) \\
(1-p)^2 + p(1-p)(\lceil h/2\rceil - 2) &\leq g[2(\lceil h/2\rceil - 2) - p(\lceil h/2\rceil - 2) + 1 - p].
\end{aligned}$$

We know from Theorems 33 and 34 that $g \leq \dfrac{1-p}{\lceil h/2\rceil - 1}$. Thus,

$$\begin{aligned}
(1-p)^2 + p(1-p)(\lceil h/2\rceil - 2) &\leq \frac{1-p}{\lceil h/2\rceil - 1}2(\lceil h/2\rceil - 2) - p(\lceil h/2\rceil - 2) + 1 - p] \\
1 + p(\lceil h/2\rceil - 3) &\leq \frac{2\lceil h/2\rceil - 3}{\lceil h/2\rceil - 1} - p \\
1 + p(\lceil h/2\rceil - 2) &\leq \frac{2\lceil h/2\rceil - 3}{\lceil h/2\rceil - 1} \\
p(\lceil h/2\rceil - 2) &\leq \frac{2\lceil h/2\rceil - 3 - \lceil h/2\rceil + 1}{\lceil h/2\rceil - 1} \\
p(\lceil h/2\rceil - 2) &\leq \frac{\lceil h/2\rceil - 2}{\lceil h/2\rceil - 1} \\
p &\leq \frac{1}{\lceil h/2\rceil - 1}.
\end{aligned}$$

In other words, there is a contradiction for $p > \dfrac{1}{\lceil h/2\rceil - 1}$. Thus, for $\dfrac{1}{\lceil h/2\rceil - 1} \leq p \leq 1$, the edit distance function is equal to the upper bound given in Theorems 33 and 34. □

The rest of the proof of the main theorem may be broken down into odd and even cases.

<u>Case 1</u>: h is even

Since

$$\frac{1}{\lceil h/2\rceil - 1} \leq \frac{1}{\lceil h/3\rceil}$$

for $h > 4$, we have that $ed_{\mathcal{H}}(p) = \gamma_{\mathrm{Forb}(C_h)}(p)$ over the range $1/\lceil h/3\rceil \leq p \leq 1$ from Proposition 65. (For $h = 4$, this result was already confirmed in Theorem 34).

<u>Case 2</u>: h is odd

From Theorem 33, we know that $ed_{\mathrm{Forb}(C_h)} \leq \gamma_{\mathrm{Forb}(C_h)}(p)$. It is easily verified that for the range $0 \leq p \leq \lceil h/3\rceil$ and for $h > 3$, $\gamma_{\mathrm{Forb}(C_h)}(p) = p/2$. But since $p/2$ is also a lower bound based on concavity, we know that the edit distance function is $p/2$ for $0 \le p \le 1/\lceil h/3 \rceil$. But,

$$\frac{1}{\lceil h/2 \rceil - 1} \le \frac{1}{\lceil h/3 \rceil}$$

for $h > 4$, so we have that $ed_{\mathrm{Forb}(C_h)}(p) = \gamma_{\mathrm{Forb}(C_h)}(p)$ for $1/\lceil h/3 \rceil \le p \le 1$ by Proposition 65. Therefore, $ed_{\mathrm{Forb}(C_h)}(p) = \gamma_{\mathrm{Forb}(C_h)}(p)$ for $0 \le p \le 1$.

Plots of the function $ed_{\mathrm{Forb}(C_h)}(p)$ for $4 \le h \le 18$ are shown in Figures A.1 through A.15 in Appendix A. Upper and lower bounds for $h \ge 10$ even are shown using dotted lines.

3.3 Proof of Corollary 8

We will now discuss the maximum point of the edit distance function of C_h for $h \ge 4$.

Consider the function $\gamma_{\mathrm{Forb}(C_h)}(p)$ over the interval $\dfrac{1}{\lceil h/3 \rceil} \le p \le 1$. Since $p/2 \ge \gamma_{\mathrm{Forb}(C_h)}(p)$ over this range, we are only concerned with the last two functions that make up $\gamma_{\mathrm{Forb}(C_h)}(p)$.

The value of p at which the function changes from the middle to the last function is

$$p = \frac{1}{\lceil h/2 \rceil - \lceil h/3 \rceil + 1}.$$

The local maximum of $\dfrac{p(1-p)}{1 + p(\lceil h/3 \rceil - 2)}$ occurs at

$$p = \frac{1}{1 + \sqrt{\lceil h/3 \rceil - 1}}.$$

Thus, $\gamma_{\mathrm{Forb}(C_h)}(p)$ achieves its maximum at

$$(p^*, d^*) = \left(\frac{1}{\lceil h/2 \rceil - \lceil h/3 \rceil + 1}, \frac{\lceil h/2 \rceil - \lceil h/3 \rceil}{(\lceil h/2 \rceil - 1)(\lceil h/2 \rceil - \lceil h/3 \rceil + 1)} \right)$$

unless

$$\frac{1}{1 + \sqrt{\lceil h/3 \rceil - 1}} < \frac{1}{\lceil h/2 \rceil - \lceil h/3 \rceil + 1},$$

that is, unless the local maximum of the parabola occurs before it intersects with the line.

Proposition 66. *If* $\dfrac{1}{1 + \sqrt{\lceil h/3 \rceil - 1}} < \dfrac{1}{\lceil h/2 \rceil - \lceil h/3 \rceil + 1}$, *then* $h \in \{4, 7, 8, 10, 16\}$.

Proof. Beginning with the inequality in the statement of the proposition,

$$\begin{aligned} \frac{1}{1 + \sqrt{\lceil h/3 \rceil - 1}} &< \frac{1}{\lceil h/2 \rceil - \lceil h/3 \rceil + 1} \\ \lceil h/2 \rceil - \lceil h/3 \rceil &< \sqrt{\lceil h/3 \rceil - 1} \\ \lceil h/2 \rceil &< \lceil h/3 \rceil + \sqrt{\lceil h/3 \rceil - 1}. \end{aligned}$$

But this last inequality can only occur if

$$\begin{aligned}
\frac{h}{2} &< \frac{h+2}{3} + \sqrt{\frac{h+2}{3} - 1} \\
3h &< 2h + 4 + 6\sqrt{\frac{h-1}{3}} \\
h^2 - 20h + 28 &< 0 \\
h &< 10 + \sqrt{72} \\
h &< 18.
\end{aligned}$$

We can easily check using values of h between 4 and 17 to verify that the strict inequality holds only when $h \in \{4, 7, 8, 10, 16\}$. □

The values of p where $\gamma_{\mathrm{Forb}(C_h)}(p)$ achieves its maximum can be substituted into the function to verify the values of d^* given in the corollary.

It is interesting to note that for the values of h in $\{4, 7, 8, 10, 16\}$, the maximum value of the function is irrational. So it is possible for the edit distance function to achieve a maximum at an irrational value. In all other cases, the maximum value achieved by the function is rational.

CHAPTER 4. EDIT DISTANCE FROM $\mathrm{Forb}(C_h^2)$

This chapter is concerned with the edit distance from $\mathrm{Forb}(C_h^2)$. We will begin by recalling some relevant background in Section 4.1. Then we will proceed to deduce the clique spectrum for $\mathrm{Forb}(C_h^2)$, which will supply us with the upper bound $\gamma_{\mathrm{Forb}(C_h^2)}(p)$. Finally, we will prove some requirements for CRGs in $\mathcal{K}(\mathcal{H})$ which could have a $g_K(p)$ function smaller than this lower bound for some value of p.

4.1 Background

We begin by recalling some relevant definitions and previous results that were discussed in Chapter 2.

First, we will make use of the clique spectrum of $\mathcal{H}$ and the corresponding function $\gamma_{\mathcal{H}}(p)$.

Definition 67. *[13] Let $\mathcal{H} = \mathrm{Forb}(H)$ be a hereditary property for some graph H. The clique spectrum of $\mathcal{H}$ is the set*

$$\Gamma(\mathcal{H}) = \{(r, s) : H \not\mapsto K(r, s)\}.$$

The point (r, s) is an extreme point in the clique spectrum if $(r, s) \in \Gamma$ but $(r + 1, s) \notin \Gamma$ and $(r, s + 1) \notin \Gamma$. The set of extreme points is denoted Γ^*.

Definition 68. *[13] Let Γ be the clique spectrum of $\mathcal{H}$. Then the function $\gamma_{\mathcal{H}}(p)$ is defined as*

$$\gamma = \min\{g_{K(r,s)}(p) : (r, s) \in \Gamma\} = \min\{g_{K(r,s)}(p) : (r, s) \in \Gamma^*\}$$

The following result will help us in the case of squared cycles with length between 8 and 12, where certain small gray cycles are forbidden for CRGs in $\mathcal{K}(\mathcal{H})$.

Lemma 69. *[13] Let $0 < p < 1/2$. Let K be a p-core CRG that has all black vertices and white or gray edges. Then* *i. If there is no gray 3-cycle in K, then $g_K(p) > p/2$.*

ii. If there is no gray 4-cycle in K, then for $0 < p < 1/3$, $g_K(p) > p(1-p)$.

iii. If K contains a gray 3-cycle but does not contain 4 vertices that induce 5 gray edges, then $g_K(p) > \min\{2p/3, (1-p)/3\}$.

We recall some observations which will help us deduce the edit distance functions in some cases.

Lemma 70. *[4] The edit distance function must be continuous and concave down.*

Lemma 71. *[1] For any hereditary property $\mathcal{H}$, $ed_{\mathcal{H}}(1/2) = \gamma_{\mathcal{H}}(1/2)$*

Lemma 72. *[13] Let H be a simple graph with complement $\overline{H}$. Then $ed_{\mathrm{Forb}(H)}(p) = ed_{\mathrm{Forb}(\overline{H})}(1-p)$.*

4.2 Preliminary results

We begin by formally defining a squared cycle.

Definition 73. *Let h be an integer. The square of a cycle of length h, C_h^2, is the graph formed by taking a cycle of length h and adding edges between vertices which have only one vertex between them on the cycle.*

For small values of h, the edit distance from $\mathrm{Forb}(C_h^2)$ can be derived from other known edit distances using Lemma 72. For $h = 4$ and $h = 5$, C_h^2 is isomorphic to K_h. For $h = 6$, C_h^2 is isomorphic to the complement of $3P_2$, a perfect matching on 6 vertices. For $h = 7$, C_7^2 is isomorphic to the complement of C_7. Thus, we are interested in finding the edit distance from $\mathrm{Forb}(C_h^2)$ where $h \geq 8$.

We will begin by finding the clique spectrum of $\mathrm{Forb}(C_h^2)$. This will give us the function $\gamma_{\mathrm{Forb}(C_h^2)}$, which will serve as an upper bound for the edit distance function.

Proposition 74. *Let $h \geq 8$ and $\mathcal{H} = \mathrm{Forb}(C_h^2)$. Let C_h^2 have chromatic number χ, where $\chi = 3$ if h is divisible by 3 and $\chi = 4$ otherwise. Then $\Gamma^*(\mathcal{H}) \subseteq \{(\chi - 1, 0), (0, \lceil h/3 \rceil - 1), (1, \lceil h/4 \rceil - 1), (2, \lceil h/5 \rceil - 1)\}$.*

Proof. If $(r, s) \notin \Gamma$, that means that $K(r, s) \notin \mathcal{K}(\mathcal{H})$. Since all CRGs of the form $K(r, s)$ have all gray edges, to prove that $(r, s) \notin \Gamma$, we need to show that there is a way to partition the vertices of C_h^2 into r independent sets and s cliques. Conversely, if $(r, s) \in \Gamma$, $K(r, s) \in \mathcal{K}(\mathcal{H})$. To prove this, we need to show that there is no way to partition the vertices of C_h^2 into r independent sets and s cliques.

Using this method, we will prove whether or not several points are in the clique spectrum of $\mathrm{Forb}(C_h^2)$ for a general h. These containments will generate the extreme points of our clique spectrum and give us the γ function.

- $(\chi, 0) \notin \Gamma$

  Since χ is the chromatic number of C_h^2, there is a way to partition the vertices of C_h^2 into χ independent sets.

- $(\chi - 1, 0) \in \Gamma$

  By the definition of the chromatic number, there is no way to partition the vertices of C_h^2 into $\chi - 1$ independent sets.

- $(0, \lceil h/3 \rceil) \notin \Gamma$

  Since every vertex in C_h^2 is adjacent to the next two vertices on the cycle, these three consecutive vertices form a clique of size three. Thus, the vertices of C_h^2 may be partitioned into cliques of size three. If h is not divisible by 3, the leftover vertex or two vertices also form a clique. Thus, it is possible to partition the vertices of C_h^2 into $\lceil h/3 \rceil$ cliques.

- $(0, \lceil h/3 \rceil - 1) \in \Gamma$

  In order for the vertices of C_h^2 to be partitioned into $\lceil h/3 \rceil - 1$ cliques, one of these cliques would have to be of size at least four. However, there are no cliques of size larger than three in C_h^2, so there is no way to partition the vertices of C_h^2 in this way.

- $(3, 1) \notin \Gamma$

  If h is divisible by 3, then $\chi = 3$, so the vertices of C_h^2 may be partitioned into 3 independent sets and a clique of size zero.

If h is not divisible by 3, then form three independent sets by choosing every third vertex around the cycle. The remaining one or two vertices form a clique. Thus the vertices of C_h^2 may be partitioned into three independent sets and one clique.

- $(2, 1) \in \Gamma$

  Choose any three consecutive vertices around the cycle as the one clique of maximum size. Then since $h > 5$, there is at least one more K_3 remaining. So the remaining vertices cannot be partitioned into three independent sets.

- $(1, \lceil h/4 \rceil) \notin \Gamma$

  Let every fourth vertex around the cycle form an independent set. Then the three consecutive vertices that were skipped form $\lfloor h/4 \rfloor$ cliques of size three. If h is not divsible by 4, the leftover vertex or two form the last clique. Thus the vertices of C_h^2 may be partitioned into one independent set and $\lceil h/4 \rceil$ cliques.

- $(1, \lceil h/4 \rceil - 1) \in \Gamma$

  Assume by contradiction that the vertices of C_h^2 may be partitioned into one independent set and $\lceil h/4 \rceil - 1$ cliques. Let x and y be the number of 2- and 3- cliques in the partition, respectively. Then the maximum size of the independent set is $x + y$, since it can have at most one vertex between two of the consecutive cliques around the circle (if there were two, they would be adjacent and thus not both in the independent set). Then we have the following equation and inequality:

$$(x + y) + 2x + 3y \geq h$$

$$\lceil h/4 \rceil - 1 = x + y$$

Where the equation counts the total number of vertices in C_h^2 and the inequality counts the number of cliques.

Combining these two to solve for x gives the result that

$$x \leq 4\lceil h/4 \rceil - (h + 4) < 0,$$

which is a contradiction. Therefore, the vertices of C_h^2 cannot be partitioned into one independent set and $\lceil h/4 \rceil - 1$ cliques.

- $(2, \lceil h/5 \rceil) \notin \Gamma$

  The vertices of C_h^2 may be partitioned into 2 independent sets and $\lceil h/5 \rceil$ cliques as follows: choose $\lfloor h/5 \rfloor$ cliques of size three by skipping two vertices between each clique. The two skipped vertices will belong to opposite independent sets. If h is not divisible by 5, the leftover one or two vertices form the last clique.

- $(2, \lceil h/5 \rceil - 1) \in \Gamma$

  Assume by contradiction that the vertices of C_h^2 may be partitioned into two independent sets and $\lceil h/5 \rceil - 1$ cliques. Let x and y be the number of 2- and 3- cliques in the partition, respectively. Then the maximum size of the independent set is $2(x+y)$, since it can have at most two vertices between two of the consecutive cliques around the circle (if there were three, they would form a 3-clique that could not be partitioned into the two independents sets). Then we have the following equation and inequality:

$$2(x + y) + 2x + 3y \geq h$$

$$\lceil h/5 \rceil - 1 = x + y$$

  Where the inequality counts the total number of vertices in C_h^2 and the equation counts the number of cliques.

  Combining these two to solve for x gives the result that

$$x \leq 5\lceil h/5 \rceil - (h + 5) < 0,$$

  which is a contradiction. Therefore, the vertices of C_h^2 cannot be partitioned into two independent sets and $\lceil h/5 \rceil - 1$ cliques.

Because of the definition of an extreme point of the clique spectrum (Definition 17), we know that the extreme points of $\Gamma(\text{Forb}(C_h^2))$ must be in the set $\{(\chi - 1, 0), (0, \lceil h/3 \rceil - 1), (1, \lceil h/4 \rceil - 1), (2, \lceil h/5 \rceil - 1)\}$.

□

Using only this result, we may already prove what the edit distance function is in the case that $h = 10$.

Corollary 75.

$$ed_{\mathrm{Forb}(C_{10}^2)}(p) = \min\left\{\frac{1-p}{3}, \frac{p}{3}\right\}$$

Proof. From Proposition 74, we have that

$$ed_{\mathrm{Forb}(C_{10}^2)}(p) \leq \gamma_{\mathrm{Forb}(C_{10}^2)}(p) = \min\left\{\frac{1-p}{3}, \frac{p}{3}, \frac{p(1-p)}{1+p}, \frac{p(1-p)}{2-p}\right\} = \min\left\{\frac{1-p}{3}, \frac{p}{3}\right\}.$$

By the requirement that the edit distance function must be concave down in Lemma 70, $ed_{\mathrm{Forb}(C_{10}^2)}(p) = \gamma_{\mathrm{Forb}(C_{10}^2)}(p)$ for $p \in (0,1)$. □

A plot of the function $ed_{\mathrm{Forb}(C_{10})}(p)$ can be found in Figure B.3 in Appendix B.

We now are interested in finding out what a CRG K must look like in order to have a $g_K(p)$ function that beats the upper bound of $\gamma_{\mathrm{Forb}(C_h^2)}$. For large values of h we may eliminate many types of CRGs.

Lemma 76. *Let $h \geq 13$. If there is a p-core CRG $K \in \mathcal{K}(\mathrm{Forb}(C_h^2))$ such that $g_K(p) < \gamma_{\mathrm{Forb}(C_h^2)}(p)$ for some value of $p \in (0,1)$, then $p < 1/2$.*

Proof. Since we know the extreme points of the clique spectrum for $\mathrm{Forb}(C_h^2)$ from Proposition 74, we may define the gamma function for this hereditary property using Corollary 20:

$$\gamma_{\mathrm{Forb}(C_h^2)}(p) = \min\left\{\frac{p}{\chi - 1}, \frac{1-p}{\lceil h/3\rceil - 1}, \frac{p(1-p)}{1 + p(\lceil h/4\rceil - 2)}, \frac{p(1-p)}{2 + p(\lceil h/5\rceil - 3)}\right\}. \tag{4.1}$$

A simple computation verifies that $\gamma_{C_h^2}(1/2) = \dfrac{1}{2(\lceil h/3\rceil - 1)}$ for $h \geq 13$. We know from Lemma 71 that $\gamma_{\mathcal{H}}(1/2) = ed_{\mathcal{H}}(1/2)$ for all hereditary properties. Thus, $ed_{\mathrm{Forb}(C_h^2)}(1/2) = \dfrac{1}{2(\lceil h/3\rceil - 1)}$. The edit distance function must be concave down from Lemma 70, so we have that $ed_{\mathrm{Forb}(C_h^2)}(p) = \dfrac{1-p}{\lceil h/3\rceil - 1}$ for all $p \geq 1/2$. □

Lemma 77. *Let $6 \leq h \leq 23$ and let $p < 1/2$. If there is a p-core CRG $K \in \mathcal{K}(\mathrm{Forb}(C_h^2))$ such that $g_K(p) < \gamma_{\mathrm{Forb}(C_h^2)}(p)$ for some value of p, K has at most one white vertex.*

Proof. Since the chromatic number χ of C_h^2 is at most 4, K can have at most 3 white vertices. If h is divisible by 3, then K cannot have 3 white vertices since $\chi = 3$ in that case.

Suppose that h is not divisible by 3 and K has 3 white vertices. But then there can be no black vertices, because C_h^2 would embed in K since it can be partitioned into 3 independent sets and one or two leftover vertices that form a clique. But then $K = K(3,0)$, so $g_K(p) \geq \gamma_{\text{Forb}(C_h^2)}(p)$. Thus, K can have at most 2 white vertices.

Suppose K has 2 white vertices. Then if K has at least $\lfloor h/3 \rfloor$ black vertices, C_h^2 embeds into K as follows: The two independent sets are filled by taking two adjacent vertices and skipping one around the cycle. There are $\lfloor h/3 \rfloor$ skipped vertices with at most 2 leftover vertices after the last skipped vertex. These skipped vertices form $\lfloor h/3 \rfloor$ cliques, none of which have any edges between them. Thus, K can have at most $\lfloor h/3 \rfloor - 1$ black vertices. The CRG $K = K(2, \lfloor h/3 \rfloor - 1)$ has the function $g_K(p) = \dfrac{p(1-p)}{2 + p(\lfloor h/3 \rfloor - 3)}$ from Corollary 20. But $\dfrac{p(1-p)}{2 + p(\lfloor h/3 \rfloor - 3)} \geq \dfrac{p(1-p)}{2 + p(\lceil h/4 \rceil - 2)}$ for $h \leq 23$, so $g_K(p) \geq \gamma_{\text{Forb}(C_h^2)}(p)$. Thus, K can have at most one white vertex.

□

Lemma 78. *Let $h \geq 6$ and $p < 1/2$. If there is a p-core CRG $K \in \mathcal{K}(\text{Forb}(C_h^2))$ such that $g_K(p) < \gamma_{\text{Forb}(C_h^2)}(p)$ for some value of p with exactly one white vertex, then K cannot have a gray cycle of black vertices with length in $\{\lceil h/4 \rceil, ..., 2\lfloor h/3 \rfloor\}$.*

Proof. Consider a partition of the vertices of C_h^2 where an independent set is composed of every third vertex around the cycle. Then the two consecutive vertices between vertices in the independent set can make up either one or two cliques which have edges only to the nearest clique on each side. If $h = 1 \mod 3$ then there are 3 vertices in the last set that is skipped, which can still be partitioned into one or two cliques as desired. If $h = 2 \mod 3$, then the last vertex in the independent set can be shifted forward one to leave two sets of skipped vertices of size 3. Since there are $\lfloor h/3 \rfloor$ sets of skipped vertices that can be partitioned into one or two cliques which have edges only to the nearest clique on each side, using this partition C_h^2 can embed into K if K has a gray cycle of black vertices of length in $\{\lfloor h/3 \rfloor, ..., 2\lfloor h/3 \rfloor\}$.

Consider another partition of the vertices of C_h^2 where an independent set is now composed

h	Forbidden cycle lengths
6	$2, 3, 4$
7	$2, 3, 4$
8	$2, 3, 4$
9	$3, 4, 5, 6$
10	$3, 4, 5, 6$
11	$3, 4, 5, 6$
12	$3, 4, 5, 6, 7, 8$
13	$4, 5, 6, 7, 8$
14	$4, 5, 6, 7, 8$

Figure 4.1 Forbidden cycle lengths.

of every fourth vertex around the cycle. Then then three consecutive vertices between vertices in the independent set can be partitioned into one or two cliques which have edges only to the nearest clique on either side. If h is not divisible by 4, the last set of skipped vertices must be partitioned into two cliques. Thus, C_h^2 can embed into K using this partition if K has a gray cycle of black vertices of length in $\{\lceil h/4 \rceil, ..., 2\lfloor h/4 \rfloor\}$.

Since $2\lfloor h/4 \rfloor \geq \lfloor h/3 \rfloor$, we have that K cannot have any gray cycles on black vertices of length in $\{\lceil h/4 \rceil, ..., 2\lfloor h/3 \rfloor\}$. □

4.3 The edit distance function for $\mathrm{Forb}(C_h^2)$ for $h = 8, 9, 10, 11, 12$

We have already found the edit distance function from $\mathrm{Forb}(C_h^2)$ for $h = 10$ and have discussed how to derive the edit distance function for $h < 8$. If we could prove that a CRG K as discussed above has only black vertices, we know the edit distance function for $h = 8$ and $h = 9$ and we know a portion of the function for $h = 11$ and $h = 12$.

Lemma 79. *Let $h \in \{6, ..., 14\}$ and $p < 1/2$. If K is a p-core CRG with exactly one white vertex such that $C_h^2 \not\mapsto K$, then $g_K(p) \geq \gamma_{\mathrm{Forb}(C_h^2)}(p)$.*

Proof. Let K' be the sub-CRG of K that is induced by the black vertices. We may obtain forbidden lengths of gray cycles in K' using the result from Lemma 78. These are listed in Table 4.3 below.

By the results from [13] in Theorem 34, Theorem 36, and Lemma 35, we know that, for $p < 1/2$, $h_0 \leq 8$ and h_0 is even, a black-vertex p-core CRG, K', that does not admit a gray cycle of length in $\{h_0/2, \ldots, h_0\}$ has

$$g_{K'}(p) \geq \min\left\{\frac{p(1-p)}{1+p(\lceil h_0/3\rceil - 2)}, \frac{1-p}{\lceil h_0/2\rceil - 1}\right\}.$$

Using Proposition 19, we can compute the value of $g_K(p)$ from $g_{K'}(p)$. Namely,

$$\begin{aligned}
(g_K(p))^{-1} &= p^{-1} + (g_{K'}(p))^{-1} \\
&\leq p^{-1} + \max\left\{\frac{1+p(\lceil h_0/3\rceil - 2)}{p(1-p)}, \frac{\lceil h_0/2\rceil - 1}{1-p}\right\} \\
&= \max\left\{\frac{2+p(\lceil h_0/3\rceil - 3)}{p(1-p)}, \frac{1+p(\lceil h_0/2\rceil - 2)}{p(1-p)}\right\} \\
g_K(p) &\geq \min\left\{\frac{p(1-p)}{2+p(\lceil h_0/3\rceil - 3)}, \frac{p(1-p)}{1+p(\lceil h_0/2\rceil - 2)}\right\}.
\end{aligned}$$

If $h \in \{6, 7, 8\}$, then $h_0 = 4$ and $g_K(p) \geq \min\left\{\frac{p(1-p)}{2-p}, p(1-p)\right\} = \frac{p(1-p)}{2-p}$.
If $h \in \{9, 10, 11\}$, then $h_0 = 6$ and $g_K(p) \geq \min\left\{\frac{p(1-p)}{2-p}, \frac{p(1-p)}{1+p}\right\} = \frac{p(1-p)}{2-p}$.
If $h \in \{12, 13, 14\}$, then $h_0 = 8$ and $g_K(p) \geq \min\left\{\frac{p(1-p)}{2}, \frac{p(1-p)}{1+2p}\right\} = \frac{p(1-p)}{2}$.
Recall that, for $\chi = \chi(C_h^2)$, we have $\gamma_{\mathrm{Forb}(C_h^2)}(p) = \min\left\{\frac{p}{\chi-1}, \frac{p(1-p)}{2+p(\lceil h/5\rceil-3)}, \frac{p(1-p)}{1+p(\lceil h/4\rceil-2)}, \frac{1-p}{\lceil h/3\rceil-1}\right\}$.
For $6 \leq h \leq 10$,

$$g_K(p) \geq \frac{p(1-p)}{2-p} = \frac{p(1-p)}{2+p(\lceil h/5\rceil - 3)} \geq \gamma_{\mathrm{Forb}(C_h^2)}(p).$$

For $12 \leq h \leq 14$,

$$g_K(p) \geq \frac{p(1-p)}{2} = \frac{p(1-p)}{2+p(\lceil h/5\rceil - 3)} \geq \gamma_{\mathrm{Forb}(C_h^2)}(p).$$

For $h = 11$, we know that $\chi(C_{11}^2) = 4$ and so for $p < 1/2$,

$$g_K(p) \geq \frac{p(1-p)}{2-p} > \frac{p}{3} \geq \gamma_{\mathrm{Forb}(C_h^2)}(p).$$

□

Lemma 80. *Let $h \geq 6$ and let $\mathcal{H} = \mathrm{Forb}(C_h^2)$. Let $p < 1/2$ and let K be a p-core CRG in $\mathcal{K}(\mathcal{H})$ with only black vertices. Then there are no gray cycles of length in $\{\lceil h/3\rceil, ..., \lfloor h/2\rfloor\}$ in K.*

Proof. Assume by contradiction that $K \in \mathcal{K}(\mathcal{H})$. Partition the vertices of C_h^2 into 2- and 3-cliques around the cycle. Then there are between $\lceil h/3\rceil$ and $\lfloor h/2\rfloor$ cliques in the partition.

Since there are at least 2 vertices in each clique, a clique has edges only with the cliques on either side of it on the cycle. Thus, if K has a gray cycle of length in $\{\lceil h/3\rceil, ..., \lfloor h/2\rfloor\}$, C_h^2 will embed in K using this partition. Thus, $K \notin \mathcal{K}(\mathcal{H})$. □

We will now proceed to examine the edit distance function for the specific cases of $h = 8, 9, 11, 12$.

Corollary 81.

$$ed_{\mathrm{Forb}(C_8^2)}(p) = \min\left\{\frac{p}{3}, \frac{p(1-p)}{2-p}, \frac{1-p}{2}\right\}$$

Proof. From Proposition 74, we have that $ed_{\mathrm{Forb}(C_8^2)}(p) \leq \gamma_{\mathrm{Forb}(C_8^2)}(p) = \min\left\{\frac{p}{3}, \frac{p(1-p)}{2-p}, p(1-p), \frac{1-p}{2}\right\} = \min\left\{\frac{p}{3}, \frac{p(1-p)}{2-p}, \frac{1-p}{2}\right\}$.

Let $K \in \mathcal{K}(\mathrm{Forb}(C_8^2))$ and let $p \in (0, 1/2)$. We know from Lemma 79 that K has all black vertices if $g_K(p) < \gamma_{\mathrm{Forb}(C_8^2)}(p)$. By Proposition 80, K has no gray 3-cycle. But according to Lemma 27, this means that $g_K(p) > p/2 \geq \gamma_{\mathrm{Forb}(C_8^2)}(p)$. Therefore, $ed_{\mathrm{Forb}(C_8^2)}(p) = \gamma_{\mathrm{Forb}(C_8^2)}(p)$ for $p \in (0, 1/2)$.

Let $\overline{C_8^2}$ denote the complement of C_8^2. Observe that if $\overline{C_8^2} \not\mapsto K$ for some CRG K, then K has no gray 4-cycle. But this means that either K has no gray 3-cycle, or it has a gray 3-cycle but no 4 vertices that induce 5 gray edges. But then either $g_K(p) > p/2$ or $g_K(p) > \min\{2p/3, (1-p)/3\}$ by Lemma 27. But using the relationship between edit distance functions of complements in Lemma 72, we have that $ed_{\mathrm{Forb}(C_8^2)}(p) \leq \gamma_{\mathrm{Forb}(C_8^2)}(p)$ for $p \in (1/2, 1)$ and thus the entire range of $p \in (0, 1)$. □

A plot of the function $ed_{\mathrm{Forb}(C_8)}(p)$ can be found in Figure B.1 in Appendix B.

Corollary 82.

$$ed_{\mathrm{Forb}(C_9^2)}(p) = \min\left\{\frac{p(1-p)}{2-p}, \frac{p(1-p)}{1+p}\right\}$$

Proof. From Proposition 74, we have that $ed_{\mathrm{Forb}(C_9^2)}(p) \leq \gamma_{\mathrm{Forb}(C_9^2)}(p) = \min\left\{\frac{p(1-p)}{2-p}, \frac{p(1-p)}{1+p}\right\}$.

Let $K \in \mathcal{K}(\mathrm{Forb}(C_9^2))$ and let $p \in (0, 1/2)$. We know from Lemma 79 that K has all black vertices if $g_K(p) < \gamma_{\mathrm{Forb}(C_9^2)}(p)$.By Proposition 80, K has no gray 3-cycle. But according to

Lemma 27, this means that $g_K(p) > p/2 \geq \gamma_{\mathrm{Forb}(C_9^2)}(p)$. Therefore, $ed_{\mathrm{Forb}(C_9^2)}(p) = \gamma_{\mathrm{Forb}(C_9^2)}(p)$ for $p \in (0, 1/2)$.

Let $\overline{C_9^2}$ denote the complement of C_9^2. Observe that if $\overline{C_9^2} \not\mapsto K$ for some CRG K, then K has no gray 3-cycle on black vertices. But then $g_K(p) > p/2$ by Lemma 27. But using the relationship between edit distance functions of complements in Lemma 72, we have that $ed_{\mathrm{Forb}(C_9^2)}(p) \leq \gamma_{\mathrm{Forb}(C_9^2)}(p)$ for $p \in (1/2, 1)$ and thus the entire range of $p \in (0, 1)$. □

A plot of the function $ed_{\mathrm{Forb}(C_9)}(p)$ can be found in Figure B.2 in Appendix B.

Corollary 83. *Let $p < 1/2$. Then*

$$ed_{\mathrm{Forb}(C_{11}^2)}(p) = \min\left\{\frac{1-p}{3}, \frac{p}{3}, \frac{p(1-p)}{2}\right\}.$$

Further, $ed_{\mathrm{Forb}(C_{11}^2)}(p)$ achieves its maximum at $(p^*, d^*) = (1/2, 1/8)$.

Proof. From Proposition 74, we have that $ed_{\mathrm{Forb}(C_{11}^2)}(p) \leq \gamma_{\mathrm{Forb}(C_{10}^2)}(p) = \min\left\{\frac{1-p}{3}, \frac{p}{3}, \frac{p(1-p)}{2}\right\}$.

Let $K \in \mathcal{K}(\mathrm{Forb}(C_{11}^2))$ and let $p \in (0, 1/2)$. We know from Lemma 79 that K has all black vertices if $g_K(p) < \gamma_{\mathrm{Forb}(C_{11}^2)}(p)$. By Proposition 80, K has no gray 4-cycle and thus has no gray C_4^+ (that is, 4 vertices that induce gray edges). If K has a gray 3-cycle, then by Lemma 27 $g_K(p) \geq \min\{2p/3, (1-p)/3\} \geq \gamma_{\mathrm{Forb}(C_{10}^2)}(p)$. On the other hand, if K has no gray 3-cycle, then by Lemma 27 $g_K(p) > p/2 \geq \gamma_{\mathrm{Forb}(C_{10}^2)}(p)$. Thus, $ed_{\mathrm{Forb}(C_{11}^2)}(p) = \gamma_{\mathrm{Forb}(C_{10}^2)}(p)$ for $p \in (0, 1/2)$.

Since the edit distance function must be concave down by Lemma 70, we have that the maximum point is at $(p^*, d^*) = (1/2, 1/8)$. □

A plot of the function $ed_{\mathrm{Forb}(C_{11})}(p)$ can be found in Figure B.4 in Appendix B. The known upper and lower bounds for $p > 1/2$ are shown using dotted lines.

Corollary 84. *Let $p < 1/2$. Then*

$$ed_{\mathrm{Forb}(C_{12}^2)}(p) = \min\left\{\frac{p(1-p)}{2}, \frac{1-p}{3}\right\}.$$

Further, $ed_{\mathrm{Forb}(C_{12}^2)}(p)$ achieves its maximum at $(p^*, d^*) = (1/2, 1/8)$.

Proof. The proof of this Corollary follows the same line of argument as the proof of Corollary 83.

□

A plot of the function $ed_{\mathrm{Forb}(C_{12})}(p)$ can be found in Figure B.5 in Appendix B. The known upper and lower bounds for $p > 1/2$ are shown using dotted lines.

4.4 The case that K has only black vertices

We now return to examine the CRG $K \in \mathcal{K}(\mathrm{Forb}(C_h^2))$ for the general case where $h \geq 6$. Although we have not yet eliminated the case that K has white vertices, here we examine the case that K is composed only of black vertices (and white or gray edges). We may follow the lines of argument used in Chapter 3 for the edit distance from $\mathrm{Forb}(C_h)$ to obtain two lemmas analogous to those for cycles about the gray degree of vertices in $V(K)$.

Lemma 85. *Let $p < 1/2$ and let K be a CRG in $\mathcal{K}(C_h^2)$ that has only black vertices. Then for all $v, w \in V(K)$, either vw is a gray edge of K or there is a vertex $u \in V(K)$ such that uv and uw are gray edges of K.*

In fact, if h is not divisible by 3, then $d_G(v) > 2/3$ for all $v \in V(K)$.

Proof. If h is divisible by 3, then $\chi = 3$ and $g \leq p/2$. If h is not divisible by 3, then $\chi = 4$ and $g \leq p/3 \leq p/2$. Since $p < 1/2$ and $g \leq p/2$ in both cases, we may follow the odd case in the proof of Proposition 53 to obtain the result that $d_G(v) > 1/2$. Therefore, for any $v, w \in V(K)$, either vw is a gray edge or else v and w share at least one common gray neighbor.

If h is not divisible by 3, then $g \leq p/3$. Since $p < 1/2$ and $g \leq \gamma_{\mathrm{Forb}(C_h^2)} \leq p/3$, we may use Lemma 30 to obtain:

$$\begin{aligned} d_G(v) &> \frac{p - p/3}{p} + \frac{1-2p}{p} x(v) \\ &> 2/3. \end{aligned}$$

□

Lemma 86. *Let $p < 1/2$ and let K be a CRG in $\mathcal{K}(C_h^2)$ that has only black vertices. Let δ be the unweighted minimum degree of vertices in K. Then $\delta \geq \lceil h/4 \rceil$.*

Proof. From the proof of Proposition 56, we have that

$$\delta > \frac{1-p}{g} - \frac{1}{p} + 1.$$

Since $g \leq \gamma_{C_h^2}(p) \leq \dfrac{p(1-p)}{1+p(\lceil h/4 \rceil - 2)}$, we have that

$$\begin{aligned} \delta &> (1-p) \div \frac{p(1-p)}{1+p(\lceil h/4 \rceil - 2)} - \frac{1}{p} + 1 \\ &= \lceil h/4 \rceil - 1. \end{aligned}$$

Therefore, $\delta \geq \lceil h/4 \rceil$. □

Although we may replicate the proofs used to find $ed_{\text{Forb}(C_h)}(p)$ in Chapter 3 for these two lemmas, we are unable to replicate the proof of Lemma 57. This is due to the fact that the set of forbidden cycle lengths is smaller in the squared cycle case.

CHAPTER 5. CONCLUSIONS AND OPEN QUESTIONS

We now know the maximum edit distance from the hereditary property $\mathrm{Forb}(C_h)$ for all values of $h \geq 3$. The entire function is known for h odd. The edit distance function is known for even h except over a range of p which is very small, especially for very large cycles.

In all cases where the edit distance function is known, it corresponds to a value from the extreme points in the clique spectrum. The maximum value of the edit distance function for most values of h corresponds to the intersection of two of the functions that make up $\gamma_{\mathrm{Forb}(C_h)}(p)$. This implies that there are two equally optimal editing "recipes". However, in some cases the edit distance function corresponds to the local maximum of the function $\dfrac{p(1-p)}{1+p(\lceil h/3\rceil - 2)}$. It is unknown why this occurs for certain values of h. It is interesting to note that in these cases, the maximum edit distance is an irrational value.

We also know the edit distance from $\mathrm{Forb}(C_h^2)$ for $h = 8, 9, 10$ and half of the function including the maximum value for $h = 11, 12$. In all of these cases except $h = 8$, the maximum value of the function occurs at $p = 1/2$, whether or not that maximum occurs at a point where the function changes between optimal CRG "recipes".

There is also a great deal of information known about a CRG $K \in \mathcal{K}(\mathrm{Forb}(C_h^2))$ for general values of h. Specifically, if $g_K(p) < \gamma_{\mathrm{Forb}(C_h^2)}(p)$ for some value of p, K can have no more than one white vertex and has certain cycle lengths which are forbidden in the sub-CRG induced by its black vertices. However, while the forbidden cycle lengths were the key to unlocking the edit distance function from $\mathrm{Forb}(C_h)$, the same methods cannot be used for squared cycles because there are not enough forbidden cycle lengths to forbid long paths.

Several questions still remain. The edit distance from $\mathrm{Forb}(C_h)$ for even h with $p < 1/\lceil h/3\rceil$ is still unknown. It would also be interesting to further examine the edit distance from $\mathrm{Forb}(C_h^2)$ for larger values of h to see if the trends we have observed continue. Finally, it would be interesting to study the edit distance from higher powers of cycles: that is, C_h^k, where $k > 2$, the graph where the vertices are adjacent to the nearest k vertices on either side around a cycle of length h.

APPENDIX A. PLOTS OF THE EDIT DISTANCE FUNCTION FOR Forb(C_h) for $4 \leq h \leq 18$

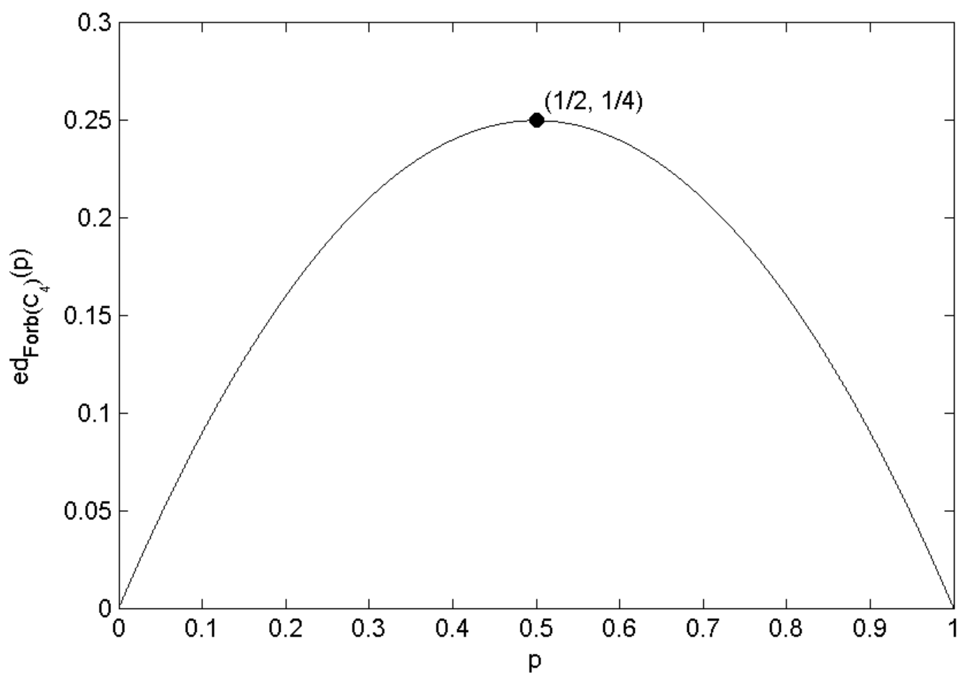

Figure A.1 Plot of the edit distance function for Forb(C_4).

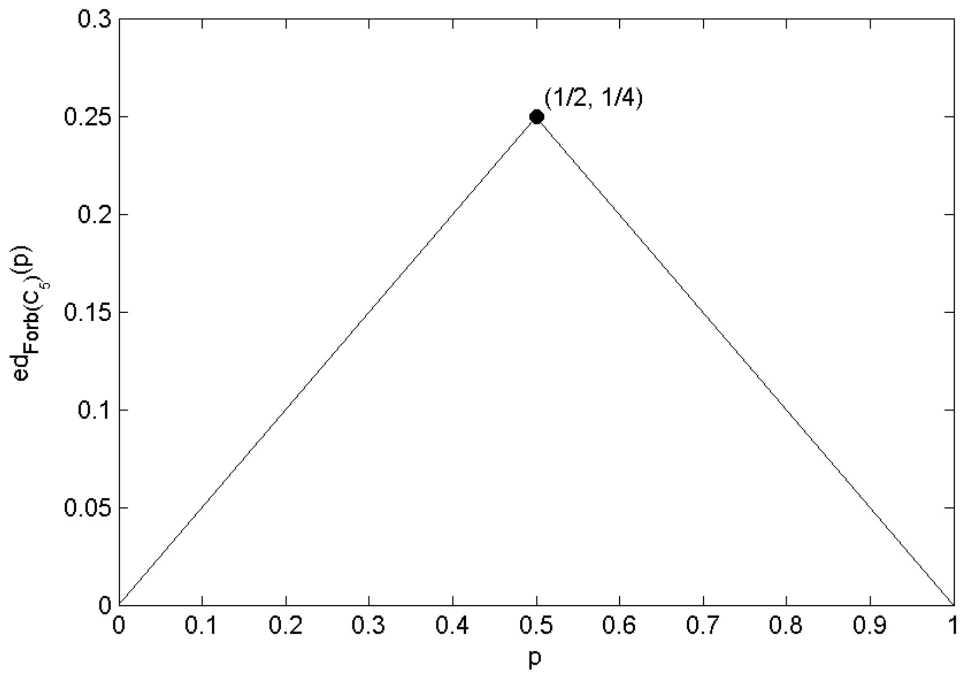

Figure A.2 Plot of the edit distance function for $\mathrm{Forb}(C_5)$.

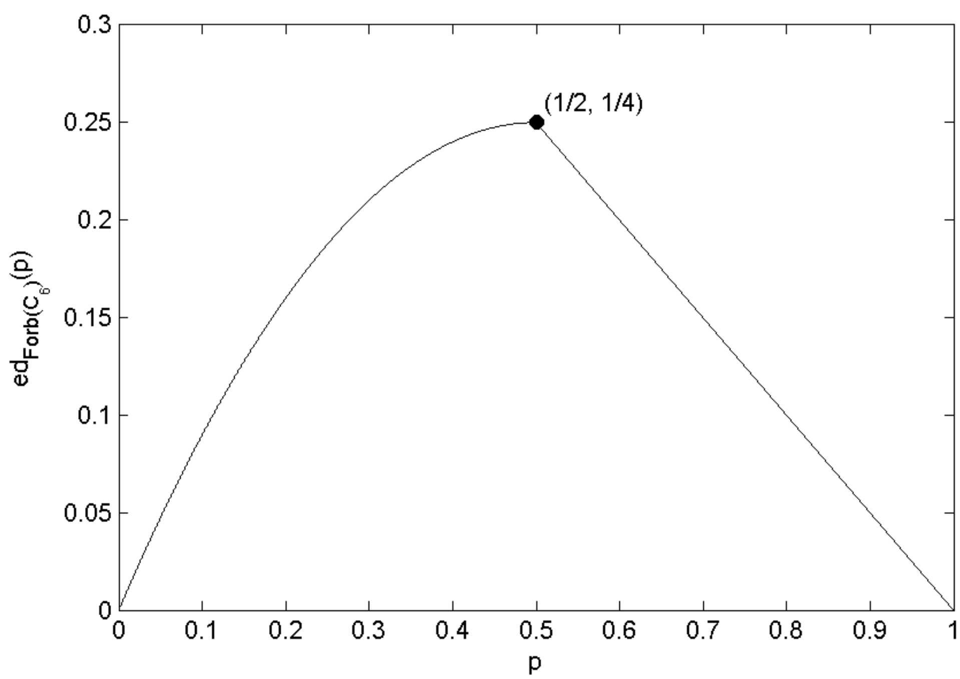

Figure A.3 Plot of the edit distance function for $\mathrm{Forb}(C_6)$.

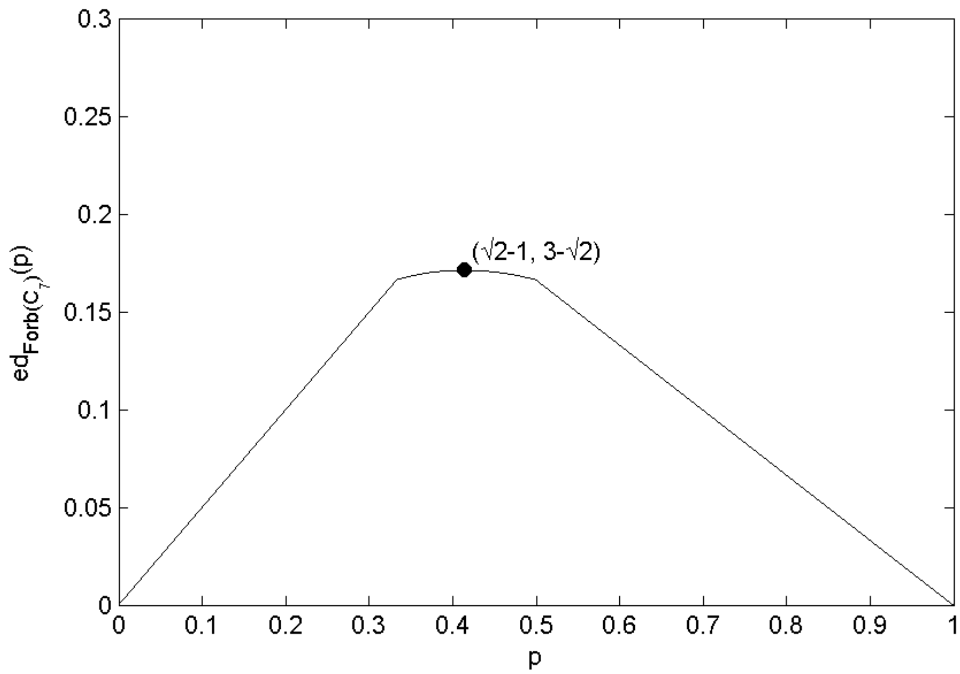

Figure A.4 Plot of the edit distance function for Forb(C_7).

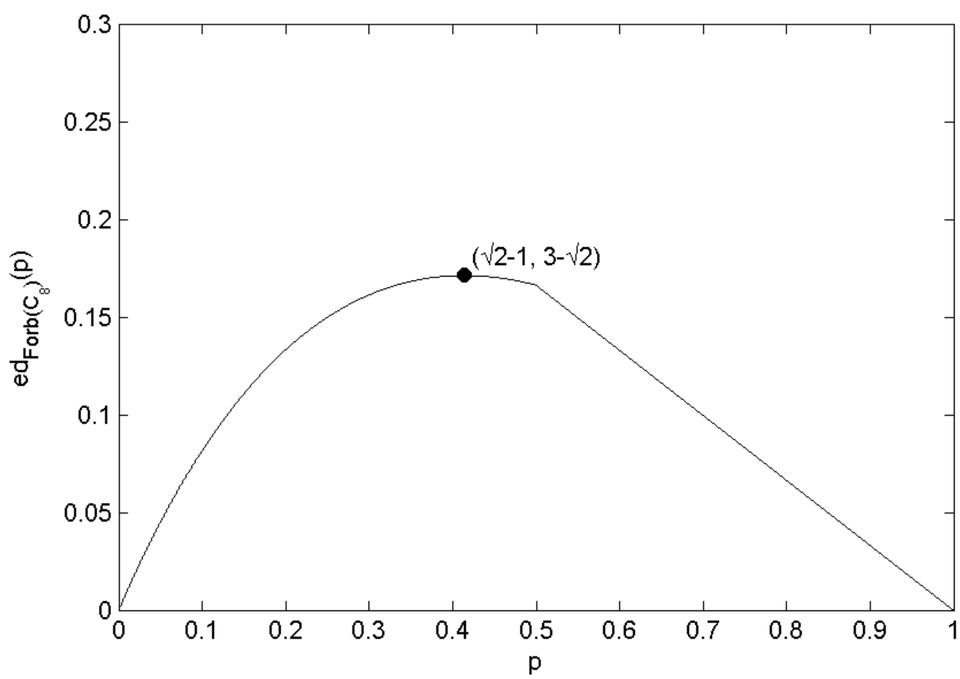

Figure A.5 Plot of the edit distance function for Forb(C_8).

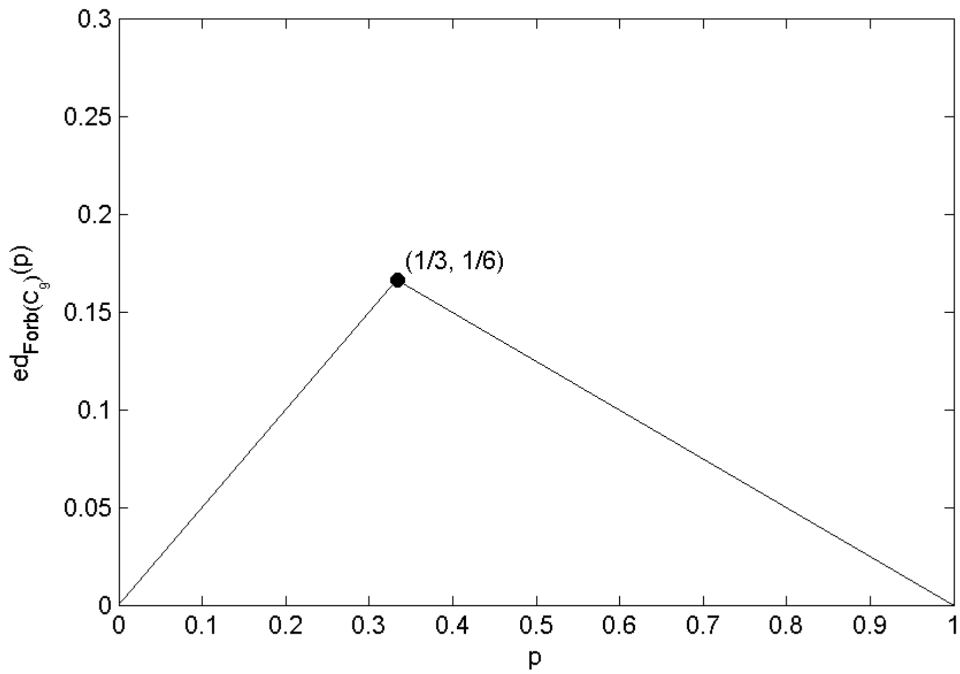

Figure A.6 Plot of the edit distance function for Forb(C_9).

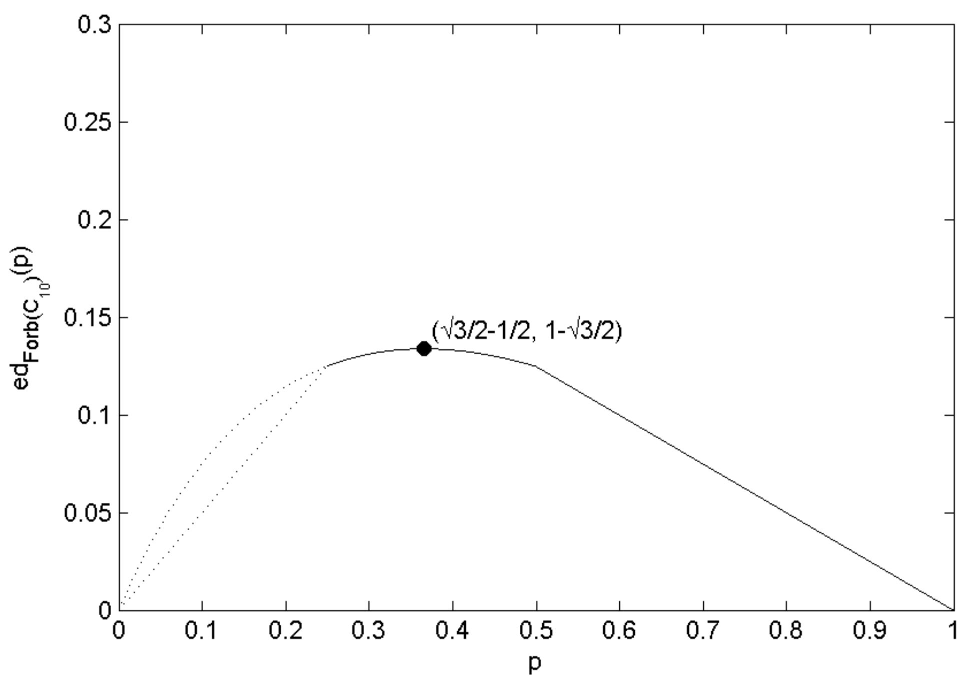

Figure A.7 Plot of the edit distance function for Forb(C_{10}).

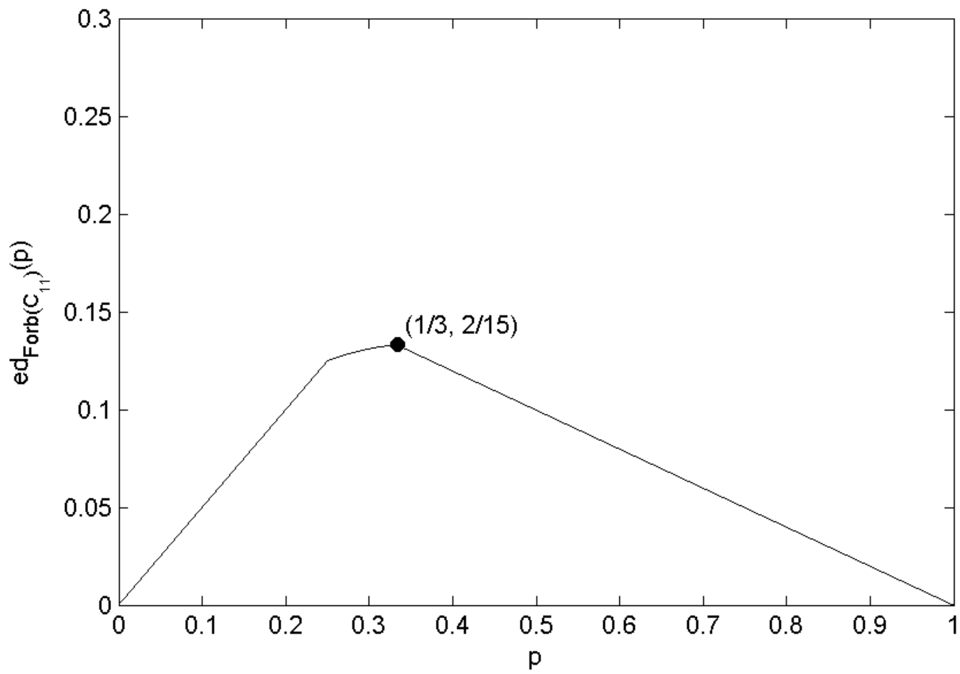

Figure A.8 Plot of the edit distance function for Forb(C_{11}).

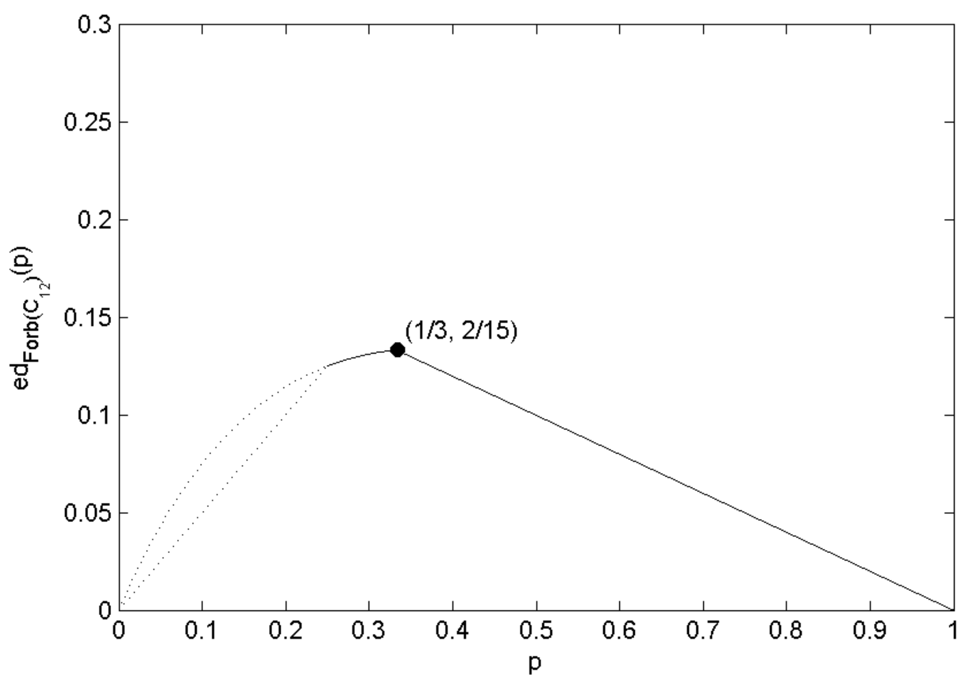

Figure A.9 Plot of the edit distance function for Forb(C_{12}).

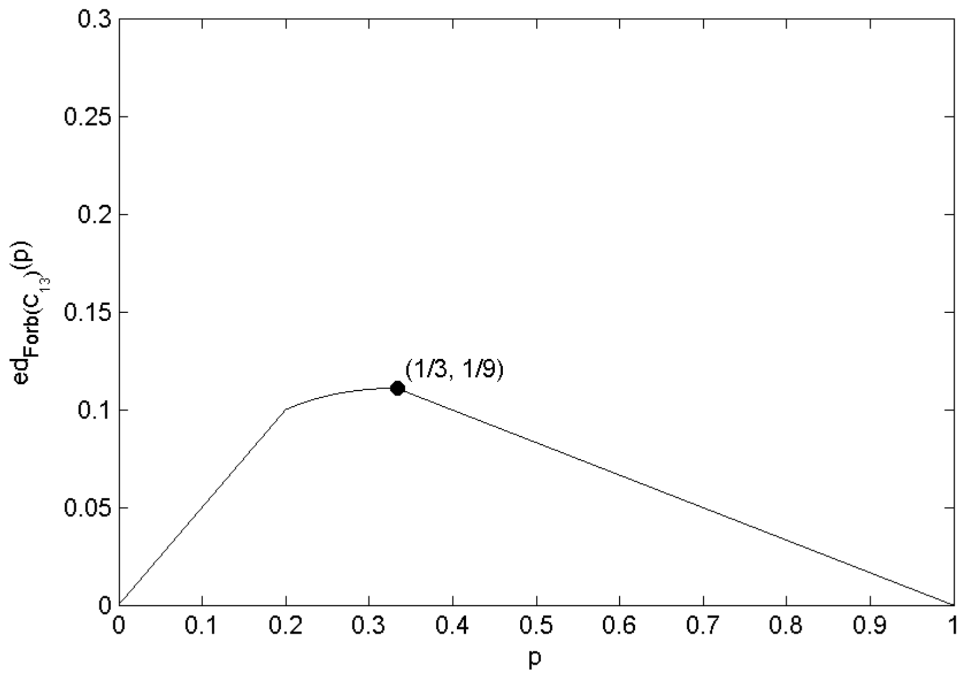

Figure A.10 Plot of the edit distance function for Forb(C_{13}).

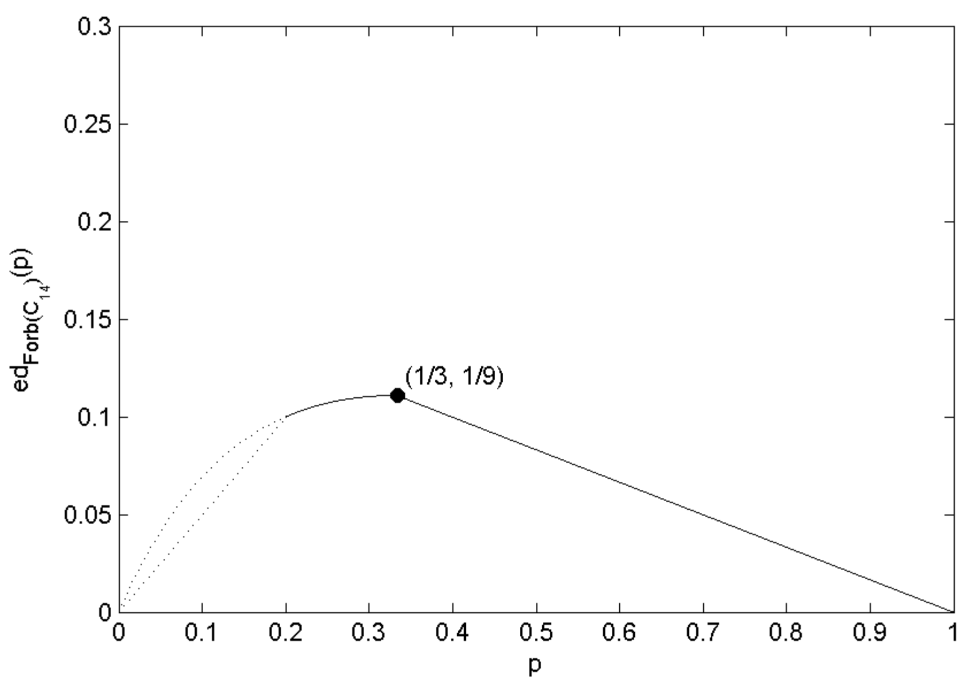

Figure A.11 Plot of the edit distance function for Forb(C_{14}).

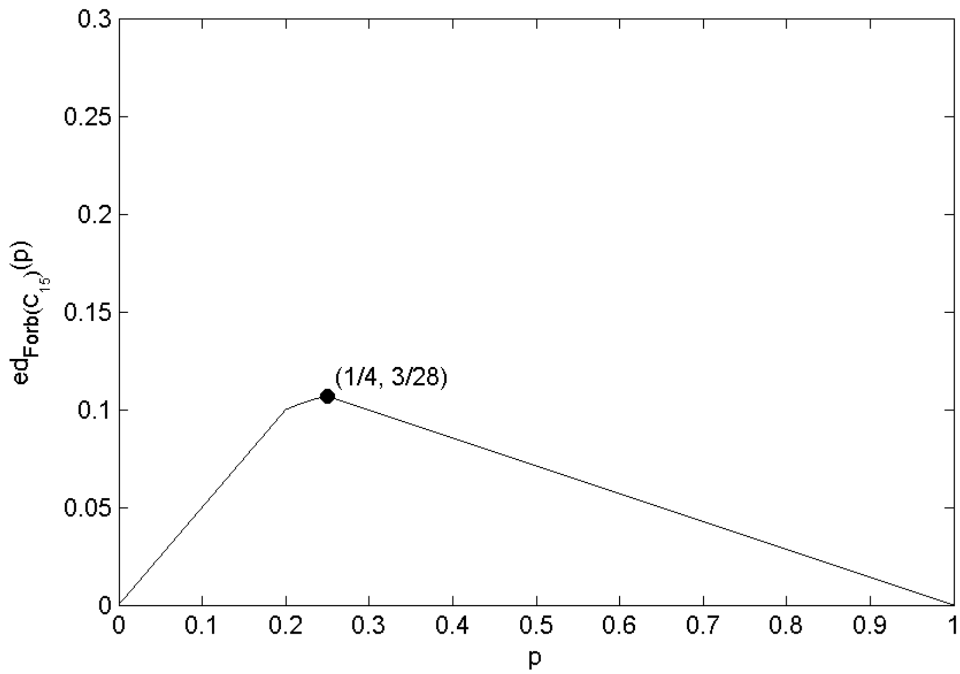

Figure A.12 Plot of the edit distance function for Forb(C_{15}).

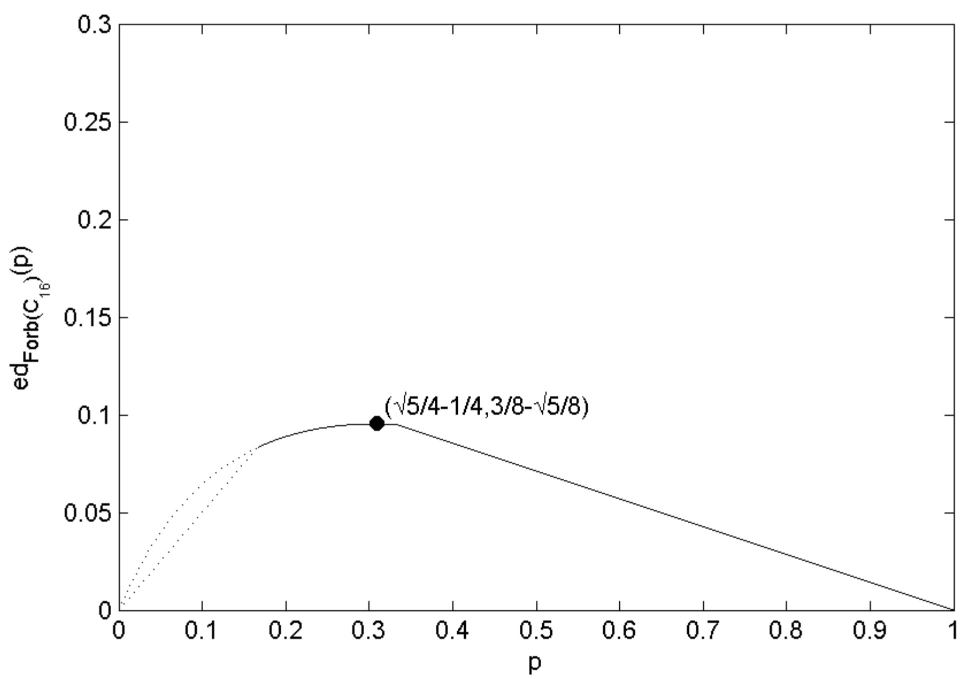

Figure A.13 Plot of the edit distance function for Forb(C_{16}).

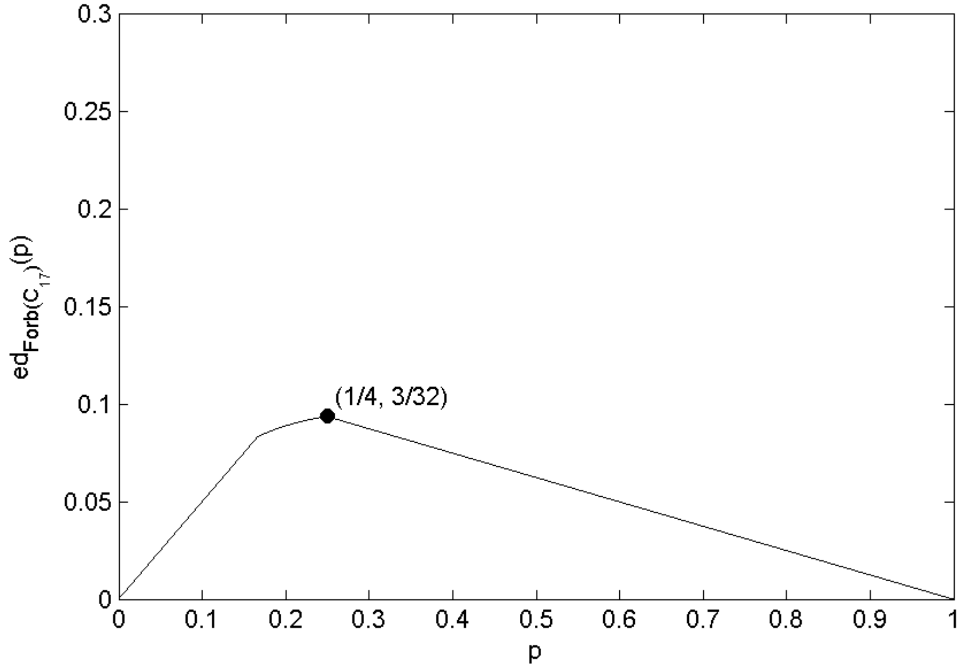

Figure A.14 Plot of the edit distance function for Forb(C_{17}).

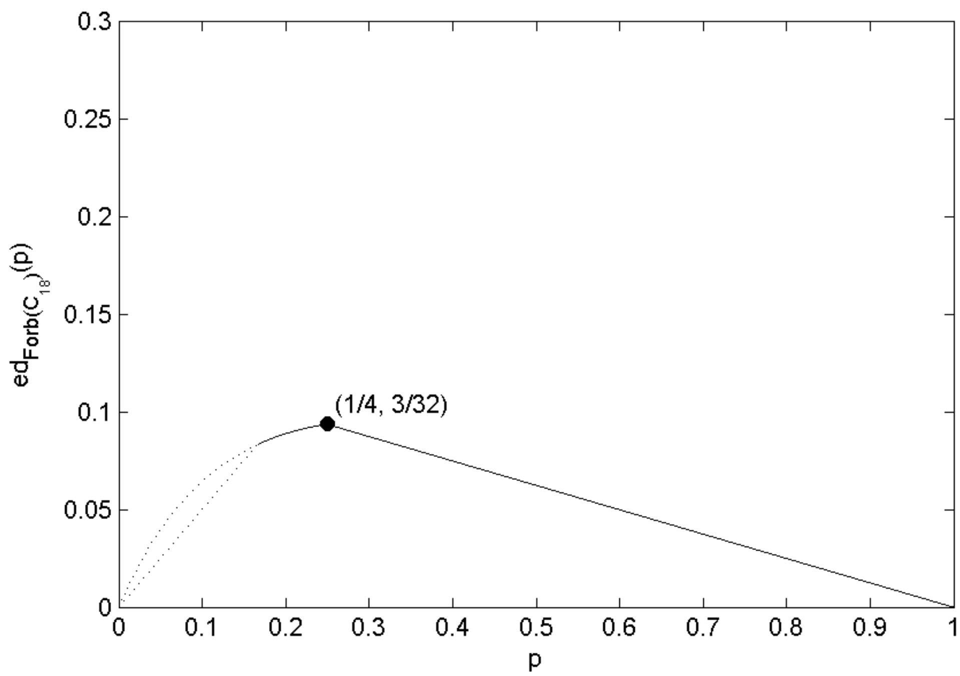

Figure A.15 Plot of the edit distance function for Forb(C_{18}).

APPENDIX B. PLOTS OF THE EDIT DISTANCE FUNCTION FOR $\mathrm{Forb}(C_h^2)$ for $8 \leq h \leq 12$

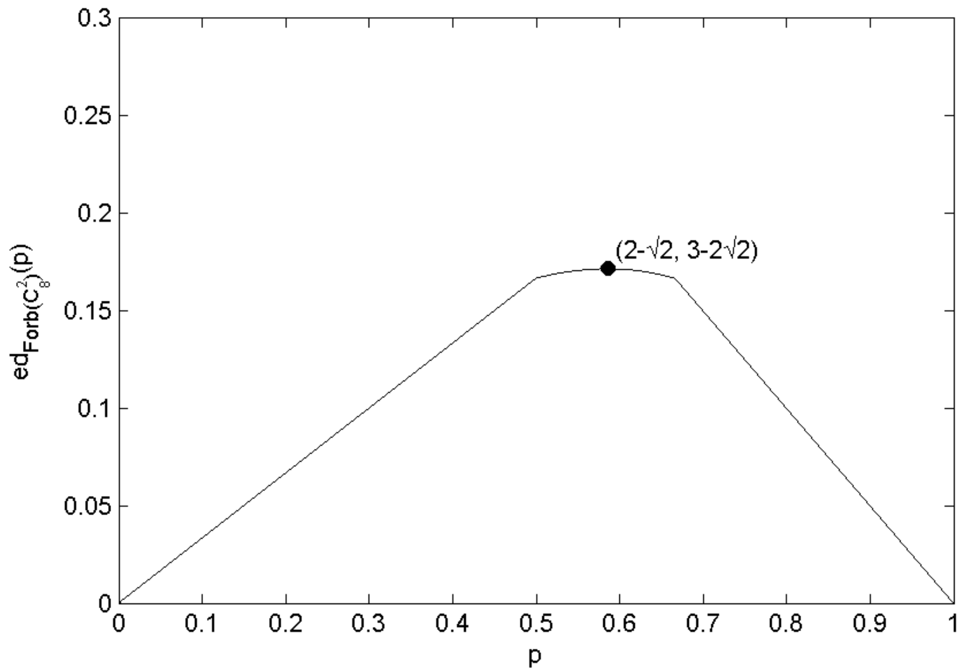

Figure B.1 Plot of the edit distance function for $\mathrm{Forb}(C_8^2)$.

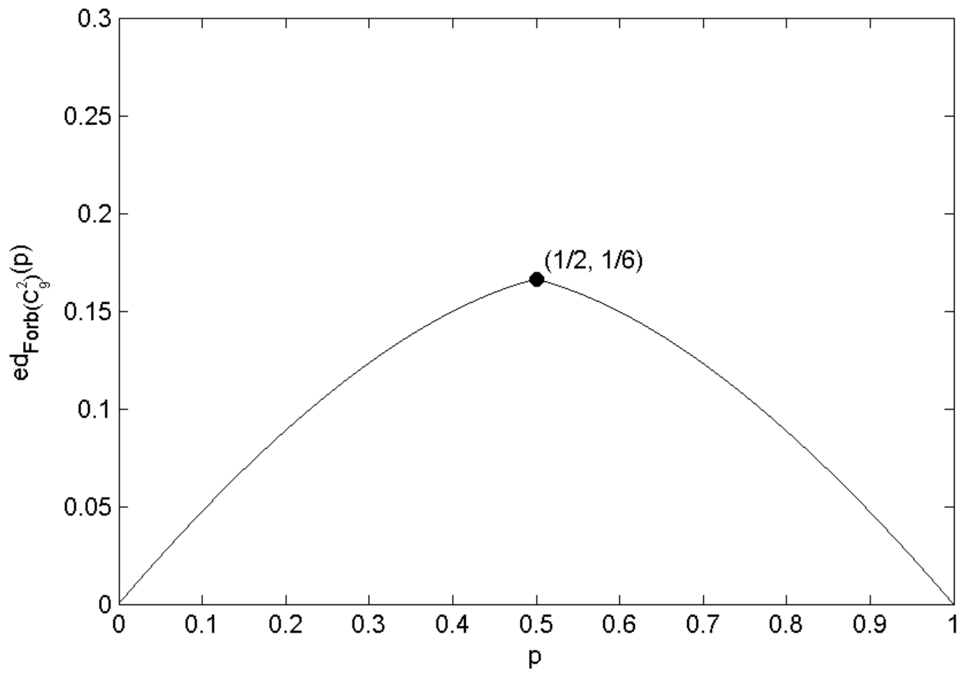

Figure B.2 Plot of the edit distance function for Forb(C_9^2).

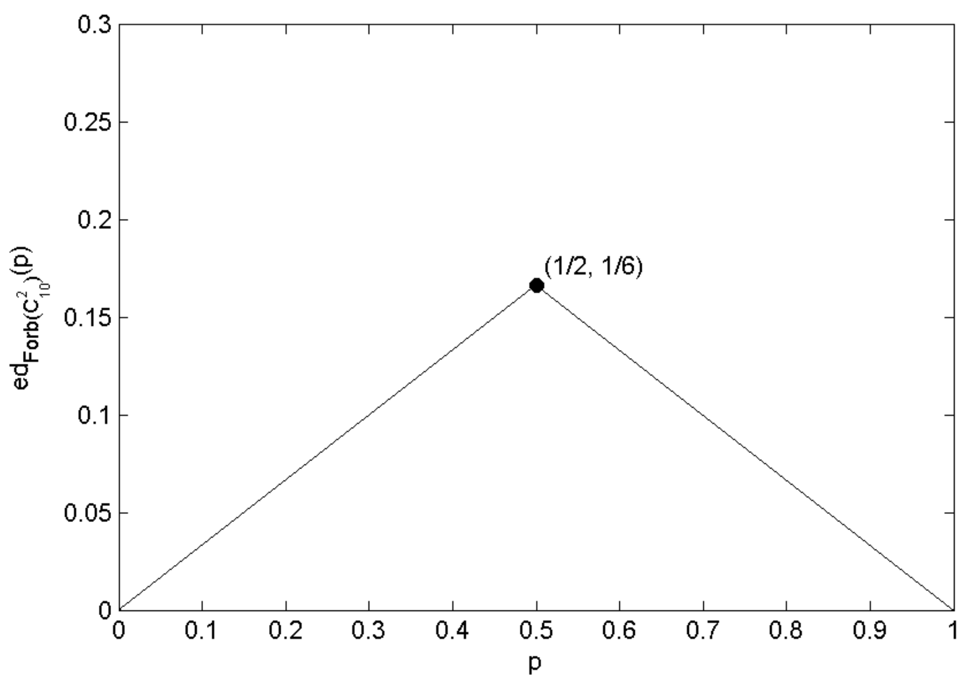

Figure B.3 Plot of the edit distance function for Forb(C_{10}^2).

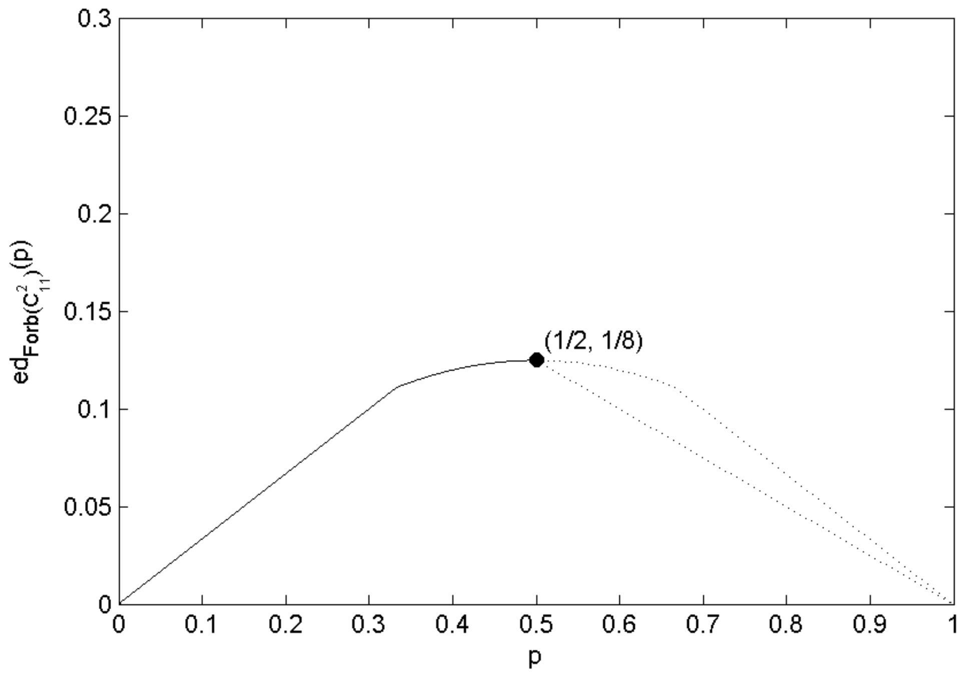

Figure B.4 Plot of the edit distance function for $\mathrm{Forb}(C_{11}^{2})$.

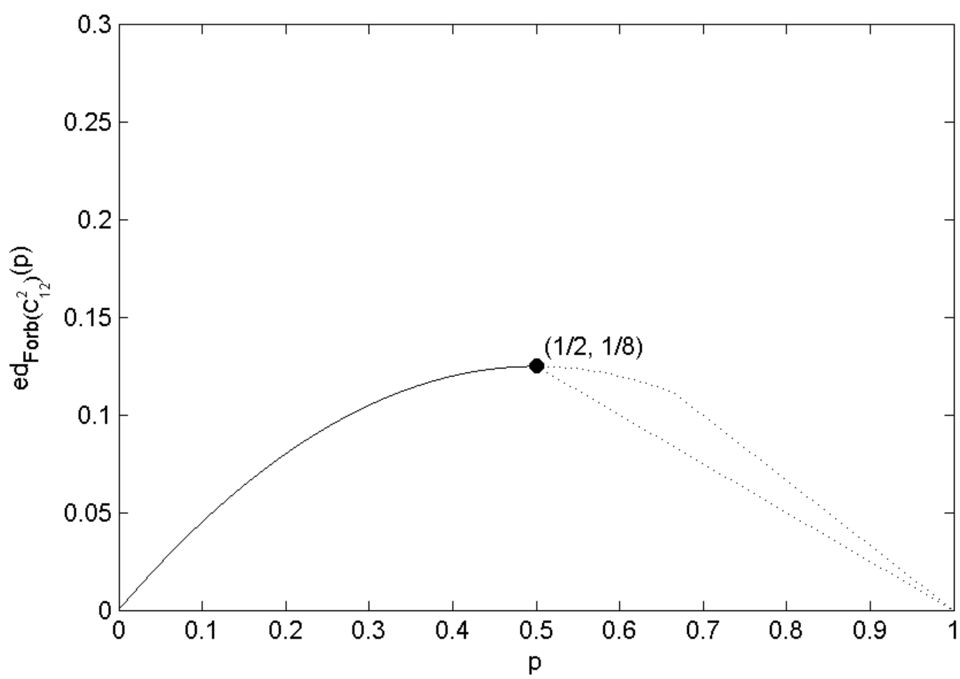

Figure B.5 Plot of the edit distance function for $\mathrm{Forb}(C_{12}^{2})$.

APPENDIX C. BACKGROUND OF SZEMEREDI'S REGULARITY LEMMA

The tool that is the foundation for the development of the edit distance function is Szemerédi's Regularity Lemma. This technical lemma is explained intuitively by Komlós and Simonovits in [9]. The basic idea of this result is to approximate graphs of a given edge density using a random graph with the same edge density. This approximation is useful because random graphs are relatively easy to work with and sometimes the structure of a graph in a given problem is not apparent.

The Regularity Lemma uses the concept of a regular pair. A regular pair can be thought of as a uniform bipartite graph. It is uniform in the sense that the density of its subgraphs is close to the density of the original graph. In this sense, a regular pair is approximately a random bipartite graph. Each independent set in a regular pair is called a cluster.

The Regularity Lemma says that any graph may be partitioned into clusters that form regular pairs plus a small number of leftover edges.

The Regularity Lemma is useful because it can be used to find a small graph H in a large graph G. In order to see this, a reduced graph R is constructed. The vertices of R correspond to the clusters of the partition of G from the regularity lemma. The vertices of R are adjacent if the corresponding clusters in G are regular with a high enough density. The Regularity Lemma can be used to show that if the reduced graph R has an induced copy of H, then G will have an induced copy of H with high probability.

The Regularity Lemma was first applied to the edit distance problem by Axenovich, Kézdy, and Martin [3]. They defined a parameter called the binary chromatic number and used this parameter along with the regularity lemma to obtain general bounds on the edit distance.

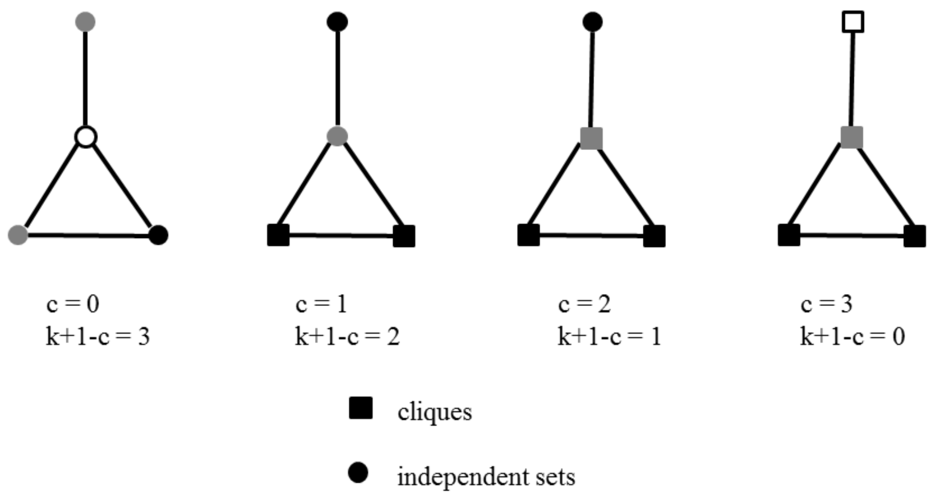

Figure C.1 The binary chromatic number of the graph H is 3.

Definition 87. *[3] Let G be a simple graph. The binary chromatic number of G, $\chi_B(G)$, is the least integer $k+1$ such that for every $c \in \{0, ..., k+1\}$, there exists a partition of the vertices of G into c cliques and $k+1-c$ independent sets.*

For example, let H be the graph on four vertices that is a triangle with a pendant edge (see Figure **).

For $k+1 = 3$, we have that for every $c \in \{0, 1, 2, 3\}$ there is a way to partition H into c cliques and $k+1-c$ independent sets (these partitions are shown in Figure C.1). However, if $k+1 = 2$ and $c = 0$, there is no way to partition H into 2 independent sets. This is due to the fact that H has a universal vertex which must be one of the independent sets, and the rest of the vertices do not form an independent set. Thus, $\chi_B(H) = 3$.

Some properties of the binary chromatic number were proven in [3]:

Theorem 88. *[3]*

i. For any graph G, let $\chi(G)$ be the chromatic number of G. Then $\chi(G), \chi(\overline{G}) \leq \chi_B(G) =$

$\chi_B(\overline{G}) \leq \chi(G) + \chi(\overline{G}) - 1$. If G is a complete p-partite graph such that each set is of size q, then the last inequality is an equality.

ii. *Let C_n be the cycle on n vertices. Then if $n \geq 5$, $\chi_B(C_n) = \lceil n/2 \rceil$.*

iii. *Let P_n be the path on n vertices. Then if $n \geq 3$, $\chi_B(P_n) = \lceil n/2 \rceil$.*

iv. *If G is a complete p-partite graph such that each set is of size q, then $\chi_B(G) = p + q - 1$.*

v. *For $n \geq 1$, $\sqrt{n} \leq \min_{|V(G)|=n} \chi_B(G) \leq \sqrt{n} + (1 + o(1))n^{0.2625}$.*

Alon and Stav [1] did further work with hereditary properties. For a given n and a hereditary property $\mathcal{H}$, they sought to find the graph on n vertices with the largest edit distance from $\mathcal{H}$.

Theorem 89. *[1] Let $\mathcal{H}$ be a hereditary graph property. Then there exists $p = p(\mathcal{H}) \in [0, 1]$ such that with high probability*

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

where $G(n, p)$ is the Erdős-Rényi random graph on n vertices with edge density p.

This theorem states that the edit distance is achieved by a random graph - that is, the furthest graph from a hereditary property in terms of edit distance is some random graph.

Let G^* be the furthest graph from $\mathcal{H}$ and has edge density p. Let G' be the closest graph in $\mathcal{H}$ to $G(n, p)$, so that $\mathrm{dist}(G(n, p), \mathcal{H}) = \mathrm{dist}(G(n, p), G')$. Then a strengthened version of the regularity lemma is applied to G' to form a reduced graph, which is used to show that the edit distance from G' to $G(n, p)$ is about the edit distance from G^* to $\mathcal{H}$. This proves the upper bound for $\mathrm{dist}(n, \mathcal{H})$ which depends on n. The proof also uses colored regularity graphs (CRGs), a tool originated by Bollobás and Thomason [5] in relation to their work with hereditary properties.

Note that this theorem only specifies the existence of a p where the maximum edit distance is achieved - it gives no indication as to how p might be determined for a specific hereditary property.

Alon and Stav expanded their work in [2] to calculate d^* and p^* (the maximum edit distance and the edge density for the random graph which achieves it, respectively) for several specific hereditary properties.

BIBLIOGRAPHY

[1] Noga Alon and Uri Stav. What is the furthest distance from a hereditary property? *Random Structures Algorithms*, 33(1):87-104, 2008.

[2] Noga Alon and Uri Stav. The maximum edit distance from hereditary graph properties. *Journal of Combinatorial Theory, Series B*, 98:672-697, 2008.

[3] Maria Axenovich, André Kézdy , and Ryan Martin. On the edit distance of graphs. *Journal of Graph Theory.* 2008. 58: 123-138.

[4] József Balogh and Ryan Martin. Edit distance and its computation. *Electronic Journal of Combinatorics,* 15(1):Research Paper 20, 27, 2008.

[5] Béla Bollobás and Andrew Thomason. The structure of hereditary properties and colourings of random graphs. *Combinatorica* 20(2), 2000.

[6] D. Chen, O. Eulenstein, D. Fernández-Baca, and M. Sanderson. Supertrees by Flipping. *Computing and Combinatorics.* Lecture Notes in Computer Science, 2387:128-137, 2002.

[7] Martin Charles Golumbic. *Algorithmic Graph Theory and Perfect Graphs, 2nd ed.* Annals of Discrete Mathematics, 57, 2004. Elsevier: Amsterdam.

[8] R.G. Downey, M.R. Fellows, R. Niedermeier, and P. Rossmanith. *Parameterized Complexity.* Springer: New York, 1999.

[9] Janós Komlós and Miklós Simonovits. Szeméredi's Regularity Lemma and its applications in graph theory. Combinatorics, Paul Erdös is eighty, Vol. 2 (Keszthely, 1993), Bolyai Soc Math Stud Vol. 2, János Bolyai Math Soc, Budapest, 1996, pp. 295-352.

[10] Edward Marchant and Andrew Thomason. Extremal graphs and multigraphs with two weighted colours. *Fete of Combinatorics and Computer Science*, Bolyai Society Mathematical Studies, 20:239-286, 2010.

[11] Terry McKee & F.R. McMorris. *Topics in Intersection Graph Theory*. Monographs on Discrete Mathematics and Applications, 1999. Society for Industrial and Applied Mathematics: Philadelphia.

[12] Ryan Martin and Tracy McKay. On the edit distance from $K_{2,t}$-free graphs, submitted, arXiv:1012.0800.

[13] Ryan Martin. The edit distance function and symmetrization, submitted, arXiv:1007.1897.

[14] Ryan Martin. On the computation of the edit distance function, submitted, arXiv:1012.3716.